Andris Pelšs

STATISTISKĀS METODES EKONOMIKĀ

Andris Pelšs

STATISTISKĀS METODES EKONOMIKĀ

Andris PELŠS. 2015. *Statistiskās metodes ekonomikā*. Rēzekne: Rēzeknes Tehnoloģiju akadēmija. 240 lpp.

Recenzenti:

- Dr. oec. **Zaiga MATULE** (Rēzeknes Tehnoloģiju akadēmija)
- Mg.math., Dr.paed. **Ilmārs KANGRO** (Rēzeknes Tehnoloģiju akadēmija)

Mācību līdzeklis sagatavots un izdots ar Rēzeknes Tehnoloģiju akadēmijas finansiālo atbalstu.

RĒZEKNES TEHNOLOĢIJU AKADĒMIJA

Publicēšanai rekomendējusi Rēzeknes Tehnoloģiju akadēmijas Studiju padome 2014.gada 26. augustā, protokola Nr. 1.

Mācību līdzeklī tiek apskatītas aprakstošās statistikas metodes: grupēšana, sadalījumu grafiskā attēlošana, raksturotāju aprēķināšana, secinošās statistikas metodes: hipotēžu pārbaude, dispersijas analīze, korelācijas un regresijas analīze.
Darbā ir iekļauta nodaļa par varbūtības teorijas pamatiem, teorētiskajiem sadalījumiem un izlases metodi tādā apjomā, lai varētu izprast statistisko secinājumu veikšanas būtību.
Mācību līdzeklis ir paredzēts RTA studentiem studiju kursu "Statistika 1" un "Statistika ekonomikā" apguvei, bet to var izmantot arī citos studiju kursos saistīto tēmu apguvei un statistiskās pētniecības metožu izzināšanai.

Redaktore: **Vita Ansone**

Vāka autore: **Jana Kozule**

ISBN 978-9984-44-241-9

SATURA RĀDĪTĀJS

IEVADS

Mācību grāmata paredzēta Rēzeknes Tehnoloģiju akadēmijas studentiem studiju kursa “Statistika 1” un “Statistika ekonomikā” apgūšanai, bet to var izmantot arī citos studiju kursos saistīto tēmu apguvei un statistiskās pētniecības metožu izzināšanai. Tajā ir apskatīti statistisko datu iegūšanas un analīzes pamati: statistiskā pētījuma (novērošanas) metodes, statistisko rādītāju veidi, varbūtību teorijas pamatnostādnes, gadījuma lielumu veidi un funkcijas, izlases novērošanas metodika, statistiskās hipotēzes, dispersijas, regresijas un korelācijas analīze.

Mācību līdzekļa satura vieglākai uztveršanai teksts ir noformēts vairākos stilos:

- jēdzieni ir izcelti treknrakstā; jēdzienu un statistiskās analīzes metožu būtības skaidrojumi noformēti *Cambria* burtrakstā;
- papildu teorija par pētāmo jautājumu, informācija par praktisku datu apstrādes pielietojumu, kas nav tik aktuāli mūsdienās, noformēti *Calibri* burtrakstā;
- lielāku piemēru nosacījumi ir numurēti un ierāmēti;
- piemēru manuālie risinājumi ir doti slīprakstā;
- piemēru risinājumu ar *Excel* programmu skaidrojumi noformēti *Arial Narrow* burtrakstā.

1. STATISTISKAS PRIEKŠMETS, DATU IEGŪŠANA UN ATTĒLOŠANA

Pēc nodaļas apgūšanas studentiem:

- *jāzina, kas ir statistika, kur izmanto statistiskās pētniecības metodes;*
- *jāzina, kas ir statistiskās pētniecības objekts, datu avoti un iegūšanas veidi;*
- *jāzina, kas ir datu grupēšana, kā datus apkopot tabulās un attēlot grafikos;*
- *jāsaprot dažādas datu mērījumu skalas.*

1.1. Statistikas priekšmets

Jēdzienus var izskaidrot ar to tulkojumiem. Termins "statistika" ir cēlies no latīņu vārda *status*, kas nozīmē "noteikts lietu stāvoklis". Tāpat no šī vārda ir radies arī itāļu vārds *stato* – valsts. Tātad statistika skaidro lietu stāvokli, apraksta valsti.[1]

Statistiskās informācijas vākšana bija pazīstama jau sirmā senatnē. Tās pirmsākumi ir cieši saistīti ar grāmatvedības uzskaites rašanos. Par pirmajiem statistiskās uzskaites gadījumiem var uzskatīt senvalstu inventarizācijas gan tautas skaitīšanā, gan īpašumu inventarizācijā.

Statistika kā zinātne sāka attīstīties 18. gadsimta vidū Vācijā. 1746. gadā Marburgas universitātē vācu zinātnieks G. Ahenvals pirmoreiz vēsturē sāka lasīt lekciju kursu ar nosaukumu "Statistika". Šajā laikā statistika vienkārši aprakstīja valsti – teritoriju, valsts iekārtu, iedzīvotājus, reliģiju, politiku. Statistika saturēja arī kvalitatīvus aprakstus, bet netika analizētas sabiedrisko procesu un parādību likumsakarības. šo virzienu sauca par valsts pārzināšanas skolu.

Apmēram 100 gadus vēlāk Anglijā (V. Petijs) radās statistikas matemātiskais virziens, kur par galveno uzdevumu uzskatīja parādīt ekonomisko parādību savstarpējās atkarības, likumsakarības. Secinājumi tika balstīti uz skaitliskiem datiem.

Statistiski matemātiskais virziens (A. Ketlē) radās 19. gadsimta pirmajā pusē. Tas pētīja statistiskās likumsakarības masveida parādībām un secinājumos balstījās uz varbūtību teoriju.[2, 3]

Varbūtību teoriju un matemātisko statistiku plaši lieto visdažādāko zinātņu nozaru pētījumos. Varbūtību teorija ir pamats pētījumu un analīzes metodikām dažādās zinātnēs. Atkarībā no zinātnes nozares statistiskās pētījumu metodes iegūst atbilstošus nosaukumus.

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 7. lpp.
[2] Turpat, 9. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 5. lpp.

Šo metožu praktisko pielietojumu bioloģijā (arī lauksaimniecībā un mežsaimniecībā) sauc par **biometriju**, kā atsevišķs virziens tas parādījās 1889. gadā (F. Galtons).

Jēdziens **psihometrija** parādās 1904. gadā. To gan vairāk lieto attiecībā uz mērījumiem psiholoģijas pētījumos.[1]

1926. gadā Ragnārs Frišs ievieš jēdzienu – ekonometrija.

Ekonometrija ir robežzinātne starp ekonomikas teoriju, statistiku un matemātiku. Kā robežzinātnei tās pētāmo jautājumu loks nav strikti noteikts, un dažādi autori to vai nu paplašina, vai sašaurina. Ekonometrija atšķiras no citas uz matemātikas balstītās zinātnes – matemātiskās ekonomikas – ar to, ka ekonometrijā secinājumi tiek izdarīti par empīriskiem (apkārtējā pasaulē novērotiem) datiem, bet matemātiskajā ekonomikā ar matemātiskām metodēm apraksta ekonomikas likumus.[2]

Gan biometrija, gan psihometrija, gan ekonometrija izmanto vienādu matemātisko aparātu – matemātiskās datu apstrādes metodes, atšķiras tikai pētījuma objekts un aprēķiniem nepieciešamo datu ieguves metodes. Neskatoties uz iepriekš minēto metožu līdzību, svarīgi ir lietotās metodes ilustrēt ar piemēriem no studenta apgūstamās profesijas puses, tādējādi padarot matemātisko metožu pielietojumu saprotamāku un ciešāk saistītu ar izvēlēto profesiju.

Visi cilvēki saskaras ar statistisko informāciju. Masu mediji ziņo par iekšzemes kopprodukta pieauguma tempiem, inflāciju, bezdarba līmeni u.tml. Daudzi cilvēki interesējas par sporta statistiku, ar interesi izlasa informāciju par kriminālnoziegumu vai negadījumu statistiku. Neskatoties uz to, ka statistika caurvij mūsu dzīvi, nav daudz tādu cilvēku, kuri prot un mērķtiecīgi iegūst statistisko informāciju, bet vēl mazāk ir cilvēku, kuru profesija ir statistiķis.

Statistiskie pētījumi kā pamatnodarbošanās ir valsts statistikas pārvaldē strādājošajiem, kā arī socioloģisko pētījumu firmu darbiniekiem. Taču arī citos uzņēmumos un iestādēs ir cilvēki, kas pilnībā vai daļēji ir nodarbināti ar statistiskajiem pētījumiem.

Statistiskos pētījumus plaši lieto kvalitātes kontrolē, tirgus pētījumos, jo uzņēmuma vadību interesē klientu apmierinātība ar piedāvātajām precēm, pakalpojumiem, darba laiku utt. Statistiskie pētījumi kaut kādā apjomā būtu jāveic katrā uzņēmumā.

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 10. lpp

[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 10. lpp.

Statistikai ir ļoti daudz dažādu definīciju, bet, tās apkopojot, statistikas termina lietojumu var atklāt šādos veidos:

- statistika ir zinātne, kura izstrādā teoriju, ko vēlāk var pielietot pētījumu praksē;
- statistika ir praktiskā datu iegūšana, apstrāde un analīze;
- statistika ir dati, kuri tiek uzrādīti uzņēmumu pārskatos, publicēti statistisko datu krājumos un presē (kaut arī tos varētu uzskatīt par statistiskā darba rezultātu);
- statistika ir statistiskās metodes, kuras lieto datu vākšanas, apkopošanas, attēlošanas, analīzes un **interpretācijas** ("*interpretatio*" latīniski nozīmē „satura atklāšana") procesā.[1]

Kurss „Statistika I" balstās uz šo pēdējo statiskas nozīmes skaidrojumu, tajā tiek apgūtas statistisko pētījumu metodes.

Statistikā nodarbojas ar pētījumam vajadzīgo datu savākšanu, apstrādi un rezultātu izvērtēšanu. Tā ir teorijas un metožu kopums, ko lieto, apstrādājot skaitliskus notikumus, kad pieņem lēmumus riska un nenoteiktības apstākļos.

1.2. Dati statistiskajā analīzē

Statistikas pētījuma objekts vienmēr ir masveida objekti un parādības, turklāt šo objektu vienības nav pilnīgi vienādas. Likumsakarības statistiskajos pētījumos izpaužas tikai kopumā.

Datu vākšanas procesu sauc par **statistisko novērošanu**.[2, 3]

Datus var iegūt, tieši novērojot – izpētes objekts nezina, ka par to tiek ievākta informācija, piemēram, satiksmes plūsmu novēro ar mērķi labāk organizēt satiksmi. Autovadītāji nezina, ka viņu pārvietošanos attiecīgajā satiksmes mezglā reģistrē.

Ekonomiskajos pētījumos ļoti bieži informāciju iegūst aptaujājot, piemēram, tirgvedības pētījumos anketē potenciālos vai faktiskos pircējus, uzdodot jautājumus, kas ļauj novērtēt klientu pirkšanas motivāciju, paradumus, segmentēt pircēju kopu.

Eksperimentu ceļā iegūst informāciju par produkcijas kvalitāti, piemēram, ķieģeļu izturību (spiedes, lieces, sasalšanas – atkušanas ciklu) nosaka, ķieģeļus ievietojot speciālās ierīcēs un izmērot, cik lielu slodzi tie spēj izturēt līdz sabrukšanai.

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 7. lpp.
[2] Turpat, 22. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 15. lpp

Visus iepriekš apskatītos informācijas iegūšanas veidus kopumā sauc par statistisko novērošanu.

Ja statistiskos datus iegūst tieši – novērojumu vai eksperimentu ceļā, tad tādu informāciju sauc par **primāro statistisko informāciju.**

Ja statistiskos datus iegūst no iepriekš iegūtas un publicētas vai nepublicētas informācijas, tad tādu informāciju sauc par **sekundāro statistisko informāciju.**

Statistiskā kopa ir liels skaits vienību, kurām piemīt:

- masveidīgums;
- vienveidīgums, atsevišķu vienību savstarpēja saistība;
- zināmas vērtības;
- variācija.[1, 2]

Pētāmās kopas sastāvu nosaka pētījuma mērķis. Piemēram, kopa būs visi valsts iedzīvotāji, ja vēlas raksturot valsti pēc demogrāfiskajiem rādītājiem. Savukārt, ja interesē studentu sasniegumi statistikas priekšmetā, tad pētāmā kopa būs studenti, kuri apguvuši šo kursu, bet, iegūstot informāciju par produktu kvalitāti, pētāmā kopa būs saražotās produkcijas vienības u.tml.

Tālāk tiks sīkāk apskatītas kopas īpašības.

Masveidīgums. Statistiskās izpētes objekts nekad nav atsevišķa vienība. Ar statistiskajām metodēm nepēta atsevišķus, unikālus objektus, kā arī par masveida objektiem neizdara secinājumus pēc vienas kopas vienības apsekošanas. Piemēram, ja pētnieks nolēmis izpētīt Latvijas lauksaimniecības stāvokli, tad viņš neņems informāciju tikai par vienu zemnieku saimniecību un pēc tās neizdarīs secinājumus par kopējo situāciju Latvijas laukos. Zemnieku saimniecības ir ļoti atšķirīgas pēc lieluma, specializācijas, mehanizācijas pakāpes, ražīguma un citiem rādītājiem. Kopējo ieskatu par lauksaimniecības stāvokli var iegūt tikai tad, ja apsekojums ir masveida (var iegūt informāciju par visām saimniecībām vai par reprezentablu – visu saimniecību kopu labi pārstāvošu saimniecību daļu).

Vienveidīgums. Lai identificētu pētāmās kopas vienības, tad tām ir jābūt kādai vienojošai pazīmei, pēc kuras tiek noteikta vienības piederība pētāmajai kopai. Piemēram, ja pētījuma mērķis ir iegūt informāciju par studentu zināšanām statistikas priekšmetā, tad pētījuma objekts būs studenti, kuri ir mācījušies minēto priekšmetu, bet citu fakultāšu studenti, kuri nemācās šādu kursu, vai jaunieši, kas vispār nemācās augstskolā, nepieder šai pētāmajai kopai un netiks anketēti.

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi".11. lpp.
[2] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 14. lpp

Bez pazīmēm, kas identificē pētāmās kopas vienības, tiek vākta informācija par pētāmo vai parasti vairākām pētāmajām pazīmēm. Piemēram, ja pēta studentu zināšanas statistikas priekšmetā, tad identifikācijas pazīme būs – vai studentam šis priekšmets ir bijis mācību programmā, bet pētāmās pazīmes – vērtējums eksāmenā, nopelnīto punktu skaits kādā speciālā pārbaudes testā, kā papildus informācija analīzei varētu būt vērtējumi kontroldarbos, nodarbību apmeklētības līmenis u.tml. Ja tiek veikta produkcijas kvalitātes pārbaude, tad identifikācijas pazīme būs produkta veids vai ražojumu partija, un pētāmās pazīmes ir atkarīgas no kontrolējamās produkcijas veida – tilpums, svars, izmēri, nostrādāto stundu skaits, trokšņa līmenis u. c.).

Zināmas vērtības. Pētāmajām pazīmēm jāvar noteikt vērtības. Tās var būt skaitliskas (kvantitatīvas), piemēram, fiziskas mērvienības (garums, tilpums, svars u. c.), laika vienības (stundas, minūtes u. tml.), finanšu vienības (eiro, dolāri u. c.). Vērtības var būt arī atributīvas (kvalitatīvas) – pazīmi raksturo ar jēdzieniem (šķirne, tautība, krāsa u. tml.). Pazīmēm izšķir vairākas mērījumu skalas:[1,2,3]

- **nominālā** skala, kad variāciju apraksta ar jēdzieniem, piemēram, tautība reliģiskā piederība, augu šķirne u. c.;
- **ordinārā (kārtas)** skala, kad pazīmes vērtības ir saranžējamas, bet intervāli starp atsevišķām variantēm nav vienādi, ordinārās skalas piemērs ir izglītības līmeņi (pamatizglītība, vidējā izglītība, augstākās izglītības 1. līmenis utt.);
- **intervālu** skala, kad pazīmes vērtību var izmērīt un raksturot ar skaitli, vienādi intervāli parāda vienādas pazīmes izmaiņas, bet nulle nenozīmē, ka pazīme vispār nav konstatēta. Intervālu skalas piemērs ir temperatūras skalas – enerģijas daudzums, kas jātērē, lai paaugstinātu kāda ķermeņa temperatūru par 10 ^{0}C, būs vienāds, ja ķermeni sasilda no 0 līdz 10 ^{0}C un no 20 līdz 30 ^{0}C, turklāt nulle grādi nenozīmē, ka ķermenim temperatūras nav;
- **attiecību (proporcionālā)** skala – novērojumus var izmērīt ar skaitli. Atšķirībā no intervālu skalas nulle nozīmē absolūtu pazīmes trūkumu. Ja svars ir nulle, tad tas nozīmē, ka objekta nav vispār, ja alga ir 0,0 EUR, tad arī algas nav.

Variācijas pastāvēšana. Statistikā interesējas tikai par pazīmēm, kuras variē, tas ir, tādām pazīmēm, kurām var būt dažādas vērtības. Pēc vienveidīgām pazīmēm tiek identificēta kopa, bet analizētas un pētītas tiek pazīmes, kuras variē. Pētot studentu panākumus statistikas mācību priekšmetā, pētāmā pazīme

---

[1] Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 31. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.] : The Dryden Press. p. 8.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 11. lpp.

nebūs studentu bioloģiskā sugas piederība, jo visi studentu pieder sugai *Homo sapiens*. Taču pētīt var studentu panākumus statistikā atkarībā no tautības (šī pazīme variē) vai svara un censties pierādīt hipotēzi, ka kādas tautības studenti labāk apgūst statistiku, vai arī meklēt saistību starp eksāmena atzīmi statistikā un studentu ķermeņa masu. Pārbaudāmās hipotēzes izvirza tajā jomā (zinātnes nozarē), kurā izmanto statistiskās pētījumu metodes. Statistika ir instruments pētnieka rokā. Datu savākšanas procesu sauc par statistisko novērošanu neatkarīgi no informācijas iegūšanas veida.

Statistiskās novērošanas gaitā parasti savāc datus par vairākām, nereti daudzām **pazīmēm**.

Var novērot visas interesējošās kopas vienības. Šādu kopu sauc par **ģenerālkopu** jeb **ģenerālo kopu** (*population*).

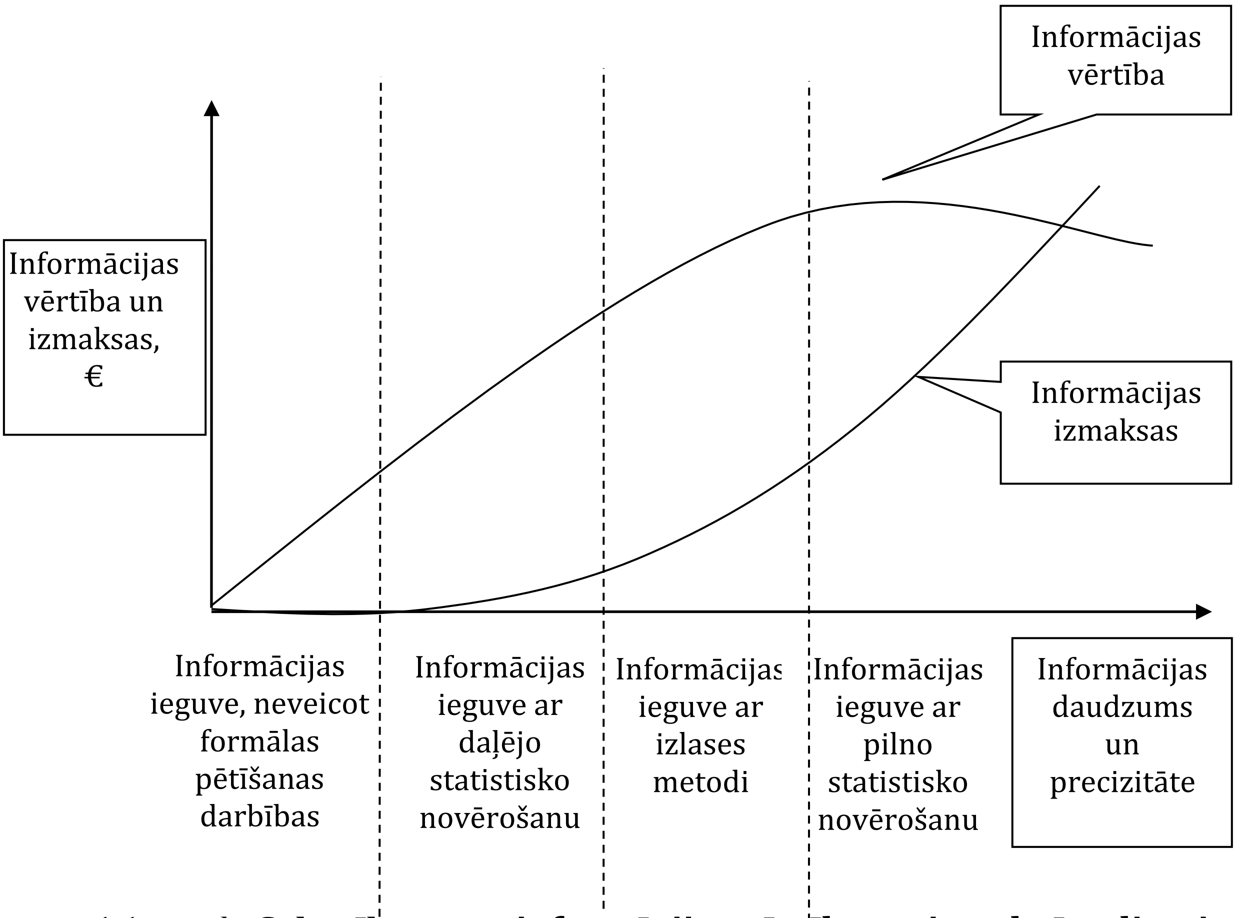
Ģenerālkopa var būt **galīga** un reāli eksistēt laikā, piemēram, valsts iedzīvotāji kādā noteiktā laika momentā, kādas augstskolas vai konkrētas specialitātes studenti. Kopa var būt arī **hipotētiska** (neierobežota), piemēram, saražotās produkcijas apjoms, kas nepārtraukti papildinās. Ja ir iespējams iegūt informāciju par visām galīgās kopas vienībām, piemēram, aptaujāt visus studentus par mācību kvalitāti, tad nav iespējams iegūt informāciju par visu saražotās produkcijas vienību ilgmūžību vai izturību ekstremālos apstākļos, ja eksperimenta rezultātā izstrādājums tiek sabojāts. Nav ekonomiski pamatoti ražot produkciju, lai to uzreiz salauztu. Priekšstatu par pētāmās kopas īpašībām var iegūt, arī neizpētot visas kopas vienības, bet tikai daļu. Šādu pētīšanu sauc par **nepilno statistisko novērošanu.**

Hipotētisko kopu pētīšanai, kā iepriekš tika noskaidrots, pilnā novērošana nemaz nav iespējama. Taču arī tur, kur ir iespējams iegūt informāciju par visu ģenerālo kopu, tas parasti nav izdevīgi, jo ir pārāk dārgi. Informācijas vērtības un iegūšanas izmaksu sakarība atkarībā no izvēlētās metodes ir parādīta 1.1.attēlā.

Nepilnā statistiskā novērošana tiek iedalīta **daļējā statistiskajā novērošanā** *(determinate sampling)* un **izlases metodē** *(random sampling)*.[1,2,3] Daļējā statistiskā novērošana nav zinātniska izpētes metode. Informācija tiek iegūta par daļu no kopas vienībām, turklāt tās nav reprezentabli atlasītas no visas kopas, bet tādas, par kurām ir pieejama informācija. Šo metodi plaši lieto uzņēmumu vadīšanā, jo tās izmantošana ir lētāka un neprasa zinātniskas rezultātu izvērtēšanas metodes. Ja ir jāizvēlas starp informācijas trūkumu un nezinātnisko daļējo statistisko novērošanu, tad noteikti pēdējā ir labāka. Ja pētniekam (uzņēmuma vadītājam), izanalizējot esošo informāciju, ir radies

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 116 lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 9.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 14. lpp.

priekšstats par pētāmo problēmu un viņš apzinās iespējamās kļūdīšanās vietas, tad ļoti iespējams, ka tiks pieņemts pareizs lēmums, turklāt iztērēti salīdzinoši mazi līdzekļi, kaut arī pētījums nav īsti zinātnisks.



1.1. attēls. **Sakarība starp informācijas vērtību un izmaksām, lietojot dažādas statistiskās novērošanas metodes**

Var novērot tikai kādu ģenerālkopas daļu, kuru atlasa tā, lai tā labi pārstāvētu (reprezentētu) visu ģenerālkopu. Šādi atlasītu kopu sauc par **izlasi (paraugkopu)**.[1,2,3,4]

Izlases metodes būtību var labi raksturot ar sadzīviska piemēra palīdzību. Lai noteiktu, vai zupai ir pietiekami sāls un garšvielu, nav jāizēd viss katls, bet ir rūpīgi jāsamaisa un jāpagaršo dažas karotes zupas. Daļējā statistiskā novērošana no izlases atšķiras tieši ar "zupas samaisīšanu". Daļējā statistiskā novērošana "garšo" to, kas pagadās priekšā, bet izlase "rūpīgi samaisa un nogaršo" tā, lai tiešām rastos pārliecība, ka visa zupa ir tieši tikpat sāļa un piparota kā

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 116. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis*. Rīga: Datorzinību Centrs. 14. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 9.
[4] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 9. lpp

pagaršotās karotes zupas (izpētītās vienības labi pārstāv – reprezentē visu kopu).

Datu ieguves un apstrādes posmi ir parādīti 1.2. attēlā.

1. Laba izpratne par pētāmo objektu

2. Sākotnējā, virspusējā novērošana un problēmas konstatācija

3. Mērķa izvirzīšana

4. Statistiskās novērošanas uzdevuma formulēšana

5. Aptaujas anketas izstrādāšana, novērojuma, eksperimenta formulēšana

6. Datu iegūšana (anketēšana, novērošana, eksperimenta veikšana)

7. Datu derīguma izvērtēšana

8. Datu apkopošana un apstrādāšana

9. Sakarību modeļu izveide, piemērotības un ticamības izvērtēšana

10. Rezultātu interpretācija un pielietošana

1.2. attēls. **Statistisko datu ieguves un apstrādes posmi**

Tālāk šī shēma tiks analizēta sīkāk.

Lai veiktu statistisko novērošanu, pētniekam ir jābūt ***labai izpratnei par objektu***, ko tas grasās izpētīt. Ja pētījumus veic uzņēmuma vai iestādes ietvaros, tad šis jautājums atrisinās pats no sevis, jo pētnieks ir uzņēmuma vai iestādes darbinieks, kurš ar pētāmajiem jautājumiem nodarbojas ikdienā, tādēļ viņam ir skaidra problēma un atliek tikai noformulēt uzdevumu. Ja pētnieks ir cilvēks "no malas" – statistisko pētījumu firmas darbinieks, students, kurš vāc informāciju diplomdarbam u. tml., tad viņam noteikti vispirms ir jāiepazīstas ar uzņēmuma (iestādes) darba tehnoloģiju visplašākajā šī jēdziena nozīmē, ir jāveic ***esošās***

situācijas analīze un problēmu konstatācija, kuras vēlāk var izpētīt ar statistiskajām metodēm.

Svarīgs posms ir ***mērķa izvirzīšana***. Statistiskās informācijas iegūšana nav pašmērķis. Iegūto informāciju izmanto, lai uzlabotu uzņēmuma vai iestādes darbu. Piemēram, produkcijas un pakalpojumu sortimentu maksimāli tuvinātu klientu vēlmēm u. tml.

Statistiskās novērošanas uzdevuma formulēšana atšķiras no pētījuma mērķa. Ja mērķis ir uzlabot klientu apkalpošanu, tad statistiskās novērošanas uzdevumā tiks noformulēts, cik uzņēmuma klienti ir jāaptaujā, kādi jautājumi ir jānoskaidro. Jo pētījums ir komplicētāks un apjomīgāks, jo nopietnāks ir sagatavošanās darbs.

Nākamajā posmā tiek ***izstrādātā aptaujas anketa***. Bieži ekonomisko informāciju iegūst anketējot (tirgzinības pētījumi). Produkcijas kvalitātes pētījumi ir ***eksperiments***, kad eksperimentālos apstākļos pārbauda izstrādājumu izturību ekstremālos darba apstākļos, ilgmūžību u.tml. Eksperimentus kādreiz organizē arī tirgus vai socioloģiskajos pētījumos. Eksperimentā atšķirībā no novērojuma apstākļi (nosacījumi) tiek mākslīgi radīti. Ir ļoti svarīgi pareizi formulēt jautājumus anketējot vai izvēlēties piemērotus mērinstrumentus un mērīšanas tehniku, ja kontrolē produkcijas kvalitāti. No šī posma ir atkarīgs, cik vēlāk iegūtie rezultāti patiesi atspoguļos realitāti un atbilstoši būs derīgi lēmumu pieņemšanai. Kā pareizi formulēt jautājumus, piedāvāt atbilžu variantus, jūs apgūsiet turpmākajos mācību kursos, kā tirgzinības pētījumi, socioloģija, psiholoģija. Galvenais nosacījums – novērojumam (eksperimentam) ir jābūt neitrālam. Piemēram, aptaujā pētnieks nedrīkst virzīt respondentu atbildes uz pētniekaprāt pareizajiem variantiem.

Anketēšana (eksperimenta informācijas iegūšana)**.** Pēc anketas izstrādāšanas (eksperimenta uzdevuma formulēšanas) ir jāiegūst informācija – jāveic anketēšana, novērošana vai eksperiments. Jebkuru fizisku vai juridisku (uzņēmums, iestāde) personu, kuru aptaujā, sauc par respondentu. Informāciju iegūstot anketējot, ir jārēķinās, ka respondenti savas atbildes variē atkarībā no tā, kāds ir viņu priekšstats par pētāmo problēmu. Ja respondents uzskatīs, ka pētījuma rezultāti var ietekmēt viņa finansiālo stāvokli, piemēram, paaugstinātu nodokļu veidā, tad viņš apzināti dos nepareizas atbildes, un pētījuma rezultāti nebūs derīgi lēmumu pieņemšanai. Anketējot ir jācenšas panākt, lai respondenti sniegtu pēc iespējas patiesākas atbildes, neizskaistinot vai nenonievājot pašreizējo stāvokli. Veicot darbības ar nedzīviem priekšmetiem vai arī augiem un dzīvniekiem, piemēram, pētot ražošanas tehnoloģijas vai saražotās produkcijas kvalitāti, jācenšas, lai mērījumi būtu precīzi, veikti pēc vienotas metodikas ar atbilstošas precizitātes (ja nepieciešams, tad sertificētām) mērierīcēm, tādējādi novēršot sistemātiskās pētījuma kļūdas.

Iegūtajai **informācijai novērtē ticamību.** Piemēram, anketās parasti ir kontroljautājumi, kuri parāda, vai respondenti ir devuši patiesas atbildes, tāpat iegūto informāciju salīdzina ar iepriekš zināmo, hipotēzēm. Ja tiek konstatētas būtiskas nepilnības informācijas ieguves procesā, tad šī informācija nav derīga tālākai apstrādei un rezultāti nav izmantojami lēmumu pieņemšanai.

Metožu ***informācijas apkopošanai un apstrādāšanai***, kā arī ***sakarības modeļu izveidei, to piemērotības un ticamības izvērtēšanas*** apguvei ir veltīts viss šis kurss, tādēļ šeit šie posmi sīkāk netiks analizēti.

Rezultātu interpretācija un lietošana ir svarīgs noslēdzošais posms. Statistiskie pētījumi nav pašmērķis, tas ir instruments citām darbības sfērām. Balstoties uz iegūtajiem rezultātiem, pieņem lēmumus, kas uzlabo uzņēmuma darbību, piemēram, klientu apkalpošanas ātrumu un kvalitāti, produkcijas kvalitāti utt.

1.3. Datu grupēšana

Sākotnēji savākto informāciju (atbildes uz anketas jautājumiem, mērījumu rezultātus) ieraksta darba tabulā. Ja aptauju veic un informāciju reģistrē pats pētnieks, tad respondentu atbildes var piereģistrēt uzreiz darba tabulā. Ja anketas izdala respondentiem (tieši, nosūta pa pastu vai publicē presē), tad pētnieks datus no šādām anketām apkopo darba tabulā. Katram respondentam ir viena rinda, bet katram jautājumam – viena aile. Mūsdienās darba tabulas parasti veido datorā – programmas *Microsoft Excel* darba burtnīcā. Eksperimenta rezultātus parasti uzreiz ieraksta darba tabulā, ko var ievadīt datorā. Modernās mērierīces ir savienotas ar datoru un dati automātiski tiek reģistrēti datorā.

Sākotnējās darba tabulas ir ļoti lielas (nereti vairāki simti ierakstu – informācijas rindiņas par atsevišķām kopas vienībām). Lai izdarītu secinājumu par masveida objektiem un parādībām (statistikas pētījumu objektiem), sākotnējā informācija ir jāapkopo, jāvispārina. Informācijas apkopošanu veic, sākotnējos datus grupējot.

Grupēšana[1] ir viena no statistisko datu apstrādes pamatmetodēm. Grupējot sākotnējie dati tiek sistematizēti – izveido atsevišķas statistiskās kopas. Piemēram, izvērtējot studentu zināšanas statistikas priekšmetā, visus studentus var sagrupēt pēc eksāmena atzīmes vai arī veidot lielākas pamatotas grupas – nesekmīgie studenti (1–3 balles eksāmenā), ar viduvējām zināšanām (4–5 balles), labām zināšanām (6–8 balles) un izcilām zināšanām (9–10 balles). Šāds grupējums ļaus labāk spriest par studentu grupas zināšanu līmeni nekā vienkāršs studentu saraksts ar eksāmena atzīmēm. Grupēšana ir analīzes

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 41. lpp.

metode. **Analīze** (no grieķu valodas tulkojot, nozīmē sadalīšana) – veselais abstrakti (domās) tiek sadalīts sastāvdaļās. Analīzē meklē cēloņus, nosaka sastāvu.

Grupēšanas uzdevumi:[1]

- tipu veidošana;
- struktūras un strukturālo izmaiņu izpēte;
- sakarību atklāšana.

Metodoloģiskās problēmas, kas jārisina, veicot datu grupēšanu, ir šādas:[2]

- grupēšanas pazīmes vai pazīmju kombinācijas izvēle;
- grupu skaita un grupu intervāla garuma noteikšana;
- rādītāju, kuriem jāraksturo izdalītās grupas, noteikšana katram konkrētam gadījumam;
- tabulu maketu izstrāde.

Pazīmi, pēc kuras notiek kopas vienību sadalīšana grupās, sauc par **grupēšanas pazīmi** vai **grupēšanas pamatu**.

Statistiskās kopas vienības var sagrupēt pēc kvantitatīvas (skaitliskas) pazīmes, kā tas bija iepriekš minēts piemērā ar studentu grupas eksāmena atzīmēm. Grupēt var arī pēc atributīvas (kvalitatīvas) pazīmes, piemēram, sagrupēt eksāmenu rezultātus pēc studentu tautības vai dzimuma. Uzrādot tabulā vienību skaitu katrā izdalītajā grupā, iegūst **atributīvu sadalījuma rindu**.

Plaši lietots ir atributīvās variācijas speciāls gadījums – alternatīvā variācija. Pazīmei ir **alternatīva variācija**, ja pazīme var iegūt tikai divas nozīmes, respektīvi, ir iespējami tikai divi varianti, piemēram, vīrietis vai sieviete; strādājošais ar augstāko izglītību vai bez tās utt. Kā var redzēt pēdējā piemērā, alternatīvo variāciju var izveidot arī tur, kur pazīmes variācija ir plašāka nekā tikai divas iespējamās nozīmes. Šādos gadījumos izdala interesējošo grupu un visas pārējās (pamatizglītība, vidējā, vidējā speciālā izglītība u. tml.) apkopo vienā grupā.

Pēc nozīmes pētāmo parādību savstarpējo saistību noteikšanai pazīmes iedala:

- **faktoriālajās pazīmēs**, kuras iedarbojas uz citām pazīmēm;
- **rezultatīvajās pazīmēs**, kuru vērtības ietekmē citu pazīmju vērtības.[3]

Atkarībā no grupēšanā izmantoto pazīmju skaita izšķir:

- vienkāršo grupēšanu, kad grupē pēc vienas pazīmes (piemēram, visu studentu grupu analizē pēc eksāmena rezultātiem);

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 43. lpp.
[2] Turpat. 44. lpp.
[3] Turpat. 47. lpp.

- salikto grupēšanu, kad grupē pēc 2 vai vairākām pazīmēm. Salikto grupēšanu parasti veic kā kombināciju. Grupas, kas sadalītas pēc vienas pazīmes, sadala sīkāk pēc otras pazīmes.[1] (Piemēram, studentus sagrupē pēc tautības un sīkāk katras grupas rezultātus analizē pēc iegūtajām eksāmena atzīmēm).

Atkarībā no izmantoto datu rakstura izšķir:
- primāro (sākotnējo) grupēšanu – izmanto tieši novērošanā iegūtos datus;
- sekundāro (otrreizējo) grupēšanu, kad sākotnējos grupējumus pārgrupē.

Iepriekš tika minēts, ka statistikā interesējas tikai par variējošām pazīmēm. Tagad vēlreiz sīkāk tiks apskatīti variācijas jēdzieni.

Novēroto un reģistrēto datu dažādību sauc par **variāciju**.[2]

Atsevišķu novēroto un reģistrēto vērtību sauc par **varianti**. Variācija var būt vienīgi iespējamo variantu ietvaros (eksāmena atzīmes, bērnu skaits ģimenē u. tml.) vai arī nosacītu variantu ietvaros (piemēram, algas var variēt ļoti plašās robežās, un atsevišķu algas lielumu var būt ļoti, ļoti daudz, tādēļ darbiniekus sagrupē atkarībā no saņemtās algas nosacītos variantos – algas līdz 300 €, 300–500 € u. tml.). **Varianti** ir iespējamās pazīmes vērtības. Piemēram, studentu eksāmena vērtējumu varianti ir no 1 līdz 10 ballēm. Variants „8” var būt reģistrēts 5 studentiem (piecas variantes būs ar vērtību „8”), bet varianti „1” un „2” var nebūt reģistrēti nevienā gadījumā – šādas variantes nav.

Ja novērojamajai pazīmei piemīt šī dažādība (variācija), pazīmi sauc par **variējošu pazīmi.** Praktiski ir nozīme savākt un apstrādāt datus tikai par variējošām pazīmēm.

Pazīmes variācija var būt kvantitatīva un atributīva.

Variācija ir **kvantitatīva**, ja variantus var izteikt ar skaitļiem, piemēram, alga, ražība, pašizmaksa, atzīme eksāmenā, svars u. c.

Variācija ir **atributīva**, ja variantus var raksturot tikai ar jēdzieniem[3], respektīvi, vārdiem, piemēram, tautība, dzimums, šķirne, krāsa utt. (nominālā mērījumu skala).

Kvantitatīva variācija var būt diskrēta un nepārtraukta.

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA “Izglītības soļi". 47. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 12. lpp.
[3] Turpat.

Variācija ir **diskrēta**, ja varianti ir nošķirti, parasti veseli skaitļi[1] (eksāmena atzīme, cilvēku skaits ģimenē, istabu skaits dzīvoklī, traktoru skaits saimniecībā).

Variācija ir **nepārtraukta**, ja viens variants no otra var atšķirties ar lielumu, kas ir pēc patikas mazs[2], piemēram svars, augums.

Grupu skaits ir atkarīgs no grupēšanas pazīmes. Ja datus grupē pēc atributīvas pazīmes, tad grupu skaitu noteiks nozīmju skaits. Ja atsevišķās grupās ir maz novērojumu, tad šādas grupas apvieno, piemēram, tautas skaitīšanā konstatē Latvijā dzīvojošās tautības, bet izdala un iedzīvotāju skaitu norāda tikai lielākajām, mazākās tiek apvienotas un uzrādītas ar nosaukumu "citas tautības". Galējs variants šādai apkopošanai ir alternatīvā variācija, kad iepriekš minētajā piemērā tiek izdalīti latvieši, bet pārējās visas tautības tiek apkopotas alternatīvajā grupā "cittautieši".

Ja grupējuma pazīme ir diskrēta un nozīmju nav daudz (bērnu skaits ģimenē u. tml.), tad grupu skaitu nosaka nozīmju skaits, turklāt arī šādos gadījumos mazākās grupas apvieno. Analizējot piemēru par bērnu skaitu ģimenēs, nav svarīgi izdalīt ģimeņu grupas, kurās aug 6, 7, 8 utt. bērni. Šādu ģimeņu būs ļoti maz un būs arī tukšas grupas, kurās nebūs neviena novērojuma.

Pazīme var būt diskrēta, bet nozīmju skaits ļoti liels, piemēram, ģimenes ieņēmumi eiro un centos. Pēc definīcijas nepārtraukti variējošajai pazīmei atšķirība starp variantiem var būt pēc patikas maza (svara atšķirības noteiks lietojamā mēraparāta precizitāte), bet šajā gadījumā ieņēmumu sīkākā dalījuma robeža ir cents un sadalījums ir diskrēts. Taču tādēļ, ka nozīmju skaits ir tik ļoti liels, šādu sadalījumu parasti uzskata par nepārtraukti variējošo sadalījumu un grupējumus vienmēr veido kā nepārtrauktam sadalījumam.

Sastādot variācijas rindu, sākotnējos variantus var apvienot **intervālos**.

Variācijas rindas, kurās uzrādīti nevis atsevišķi varianti, bet intervāli, sauc par **intervālu variācijas rindām**.[3,4]

Apstrādājot intervālu variācijas rindas, jālieto šādi lielumi:

- $x_{\min}$ – pazīmes vismazākā reģistrētā vērtība, bet, ja tā nav zināma, pirmā intervāla apakšējā robeža;
- $x_{\max}$ – pazīmes lielākā reģistrētā vērtība vai pēdējā intervāla augšējā robeža;

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 436 lpp.
[2] Turpat.
[3] Turpat, 13. lpp.
[4] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 49. lpp

- $x_{max} - x_{min}$ – **variācijas amplitūda** jeb **apjoms**;
- x_i *(z)* – norādītā intervāla zemākā (apakšējā) robeža;
- x_i *(a)* – norādītā intervāla augstākā (augšējā) robeža;
- Δ_i – intervāla garums jeb grupas lielums, $\Delta_i = x_i(a) - x_i(z)$.

Matemātiskajā statistikā parasti lieto **vienāda garuma intervālus**, bet ekonomiskajā statistikā lieto arī **nevienāda garuma intervālus,** visbiežāk tad, ja tie saistīti ar kaut kādiem ekonomiskajiem tipiem, kā arī tad, ja vienāda garuma intervālu rindas malējos intervālos biežumi ir mazi vai parādās tukši intervāli.[1,2]

Ja nav nekādu citu apsvērumu par vēlamo intervālu garumu un novērojumu skaits ir mazāks par 100, to var aprēķināt, izmantojot **Sterdžesa formulu**:[3,4]

$$\Delta = \frac{x_{max} - x_{min}}{1 + 3{,}2 \lg n} \qquad (1.1.)$$

vai **Breksa formulu,** ja novērojumu skaits ir lielāks par 100[5]:

$$\Delta = \frac{x_{max} - x_{min}}{5 \lg n}, \qquad (1.2.)$$

kur n – visu novērojumu skaits.

Ja intervāla garumu aprēķina ar kādu no minētajām formulām, tiek iegūts precīzs skaitlis, kas nav parocīgs grupēšanai. Aprēķināto intervāla garumu noapaļo līdz skaitlim, kas beidzas ar 5 vai arī vienu vai vairākām nullēm. Apaļošana ir atkarīga no skaitļu mēroga. Tāpat arī apakšējā intervāla zemāko robežu nosaka "apaļos" skaitļos tā, lai šai robežai tuvu būtu zemākā pazīmes vērtība. Intervālu robežas nevajadzētu noteikt vairāk kā ar diviem zīmīgiem cipariem (nulles nav zīmīgie cipari). Šādas "apaļas" intervālu robežas ļauj datus sagrupēt ātrāk, samazina kļūdīšanās iespējas grupējot, padara saprotamākus grupējuma rezultātus.

Nevienāda garuma intervāli var būt:

- progresīvi pieaugoši intervāli (aritmētiskajā vai ģeometriskā progresijā);
- progresīvi dilstoši intervāli;
- patvaļīgi intervāli;
- specializēti intervāli.

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 49. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 13. lpp.
[3] Turpat, 14. lpp.
[4] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 23. lpp.
[5] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 50. lpp.

Nevienāda garuma intervālus nosaka pētījuma loģika – sadalījumi ir asimetriski, ar vairākiem novērojumu biežuma pīķiem u. tml.[1]

Statistikā izmanto arī kumulatīvās jeb uzkrāto biežumu variācijas rindas. **Kumulatīvās variācijas rindas** tabulā uzrāda variantus, respektīvi, intervālus un uzkrātos biežumus. Tajos ieskaita visas tās kopas vienības, kurām reģistrētā pētāmās pazīmes vērtība ir mazāka vai vienāda ar attiecīgā intervāla augšējo robežu.[2]

Variācijas rinda izveidojas datu grupēšanas rezultātā.

1.4. Statistisko datu attēlošana

Statistiskos datus varētu parādīt kā tekstu, bet šāda veida informācija būs ļoti grūti pārskatāma un uztverama. Statistisko informāciju parasti atspoguļo:

- analītiskajās tabulās;
- attēlos (grafiskajos un citos).

Pēc tabulām un grafikiem gandrīz vienmēr seko analīzes teksts, kur galvenais ir rādītāju kvalitatīvais apskats – kas noticis, kādēļ, kas no šīs informācijas izriet. Tekstā parādās tikai atsevišķi skaitļi, kuri ir nepieciešami teksta informācijas ilustrācijai.

Vienotas shēmas tabulu izveidei nav un nevar būt. Tabulu formātu nosaka attēlojamā informācija. Tomēr tabulu veidošanā ir jāievēro daži vispārīgi principi. Katrai tabulai ir virsraksts, ir posteņi (rādītāji), kas izvietoti rindās, un pazīmes, kuras tiek parādītas ailēs. Rindas un ailes sadala tabulu šūnās, kurās tiek ierakstīta skaitliskā (dažreiz arī kvalitatīvā) informācija, kura raksturo to pazīmi, kas ir norādītās rindas un ailes krustpunktā. Ja tabulas rindās vai (un) ailēs ir informācija, kas ir summējama, tad tabulā ievieto rindu (aili) "kopā". Tabulu ietvaros var veikt arī citas aritmētiskas darbības. Ja tabulai ir gan kopsummas rinda, gan kopsummas aile, tad kopsummas ailes kopsummai ir jāsakrīt ar kopsummas rindas kopsummu. Tādā veidā pārbauda aritmētisko darbību pareizību. Ja kopsummas nesakrīt, tad ir jāmeklē kļūda aprēķinos.

Tabulā noteikti ir jābūt norādītām skaitlisko rādītāju mērvienībām. Ja visas tabulas skaitliskās vērtības ir vienādā mērvienībā, tad to norāda virsrakstā. Ja tabulas skaitļi ir dažādās mērvienībās, tad mērvienības norāda rindu vai aiļu nosaukumos atkarībā, kādā virzienā mainās mērvienība.

Analītiskajām tabulām nevajadzētu būt lielākām par vienu lapu. "Liekā" informācija ir jāapkopo, izceļot un detalizējot to informāciju, kas interesē

---

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 52. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 21. lpp.

pētnieku. Grupas, kurās skaitļi ir relatīvi nelieli, parasti apkopo, izņemot gadījumus, ja pētniekam ir specifiska interese par kādu no mazajiem posteņiem. Piemēram, analizējot uzņēmumam piederošos līdzekļus, pētnieks īpašu uzmanību pievērš nemateriālajiem aktīviem (uzņēmumam piederošās vērtības, kurām nav materiāla forma, piemēram, licence, kas ļauj nodarboties ar noteikta veida uzņēmējdarbību u.tml.). Šajā gadījumā, neskatoties uz to, ka nemateriālie aktīvi veido niecīgu uzņēmuma līdzekļu daļu, tomēr tos izdala atsevišķā rindā.

Tabulās var būt vairāku pakāpju dalījums – rindās ir minēta galvenā pazīme, ailēs – viena grupējuma pazīme, kura tiek sadalīta apakšailēs ar sīkāku iedalījumu. Tomēr nav ieteicams pārāk aizrauties ar sarežģītu tabulu veidošanu, jo tad tās kļūst nepārskatāmas, bet tabulu galvenais uzdevums ir sniegt saprotamu informāciju.

1.3. attēlā ievietotais tabulas paraugs ir iegūts no Centrālās statistikas pārvaldes datu bāzes, un tajā ir daži trūkumi. Tabulas virsrakstā nav informācijas par vietu, bet konkrētā pētījumā tas varētu būt arī nenorādīts, ja visa informācija attiecas uz Latviju. Sējumu kopplatība ir lielāka nekā izdalīto kultūru sējumu platības. Ja vēlas tabulu pilnveidot, tad pētniekam tā ir jāpārveido pašam.

Tabulas virsraksts ar mērvienību

GALVENO LAUKSAIMNIECĪBAS KULTŪRU SĒJUMU PLATĪBA (tūkst. ha)

	1938	**1985**	**1990**	**1993**	**1998**	**2002**	**2005**	**2010**	**2013**	**2016**
Graudaugi	1005,2	690,80	675,4	693,6	466,0	415,0	468,9	541,5	583,9	716,0
Rapsis	...	...	1,90	1,70	1,20	18,40	71,4	110,6	128,2	101,1
Cukurbietes	13,2	13,5	14,7	12,10	16,3	15,9	13,5	...	...	...
Kartupeļi	133,6	94,3	80,3	87,70	58,8	53,6	45,10	30,1	27,3	23,3
Dārzeņi	12,9	12,1	10,8	18,6	11,6	12,5	12,9	8,1	8,5	8,1
Garšķiedras Lini	63,5	14,3	11,9	0,6	2,2	2,1	2,2	0,0	0,2	0,1
Sējumu kopplatība	1877,4	1652,7	1627,0	1425,6	983,4	877,7	999,60	1102,7	1146,5	1233,9

Avots: LR Centrālā statistikas pārvalde

Skaitlis, kas parāda, cik liela platība 2005. gadā bija sastādīta ar kartupeļiem – tabulas šūna atrodas „2005. gada" ailes un "kartupeļu" rindas krustpunktā

1.3. attēls. **Tabulas paraugs**

Tabulas palīdz labāk saprast, izvērtēt, salīdzinot skaitlisko informāciju, kas raksturo dažādas parādības un procesus, bet tas prasa rūpīgu iedziļināšanos tabulā ietvertajos skaitļos.

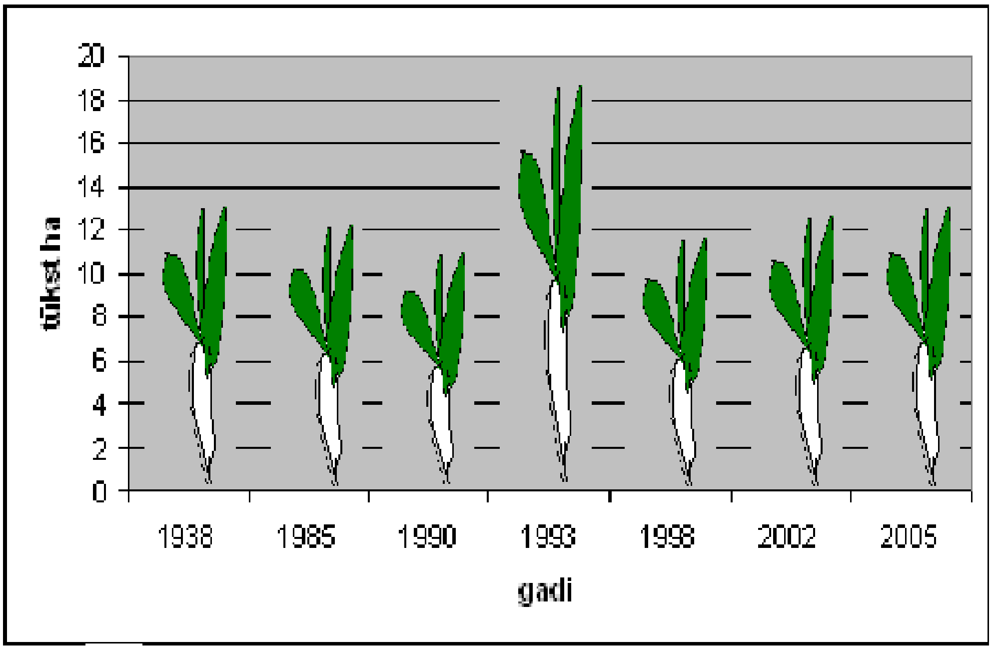
Bieži statistiskajā analīzē svarīgāk ir redzēt galvenās tendences, un šeit palīdz grafiskie attēli. Attēli vizualizē skaitlisko materiālu. Piemēram, ja kāds rādītājs ir pieaudzis 5 reizes, tad skaitļu formā šī atšķirība īpaši neizceļas, toties stabiņu grafikā stabiņš, kas attēlos rādītāju pēc pieauguma, būs piecas reizes augstāks par stabiņu, kurš attēlo to pašu rādītāju līdz pieaugumam, un tas ir ļoti uzskatāmi.

Mūsdienās grafikus parasti manuāli (ar roku) neviens neveido. Grafiku zīmēšanai izmanto programmu *Excel*.

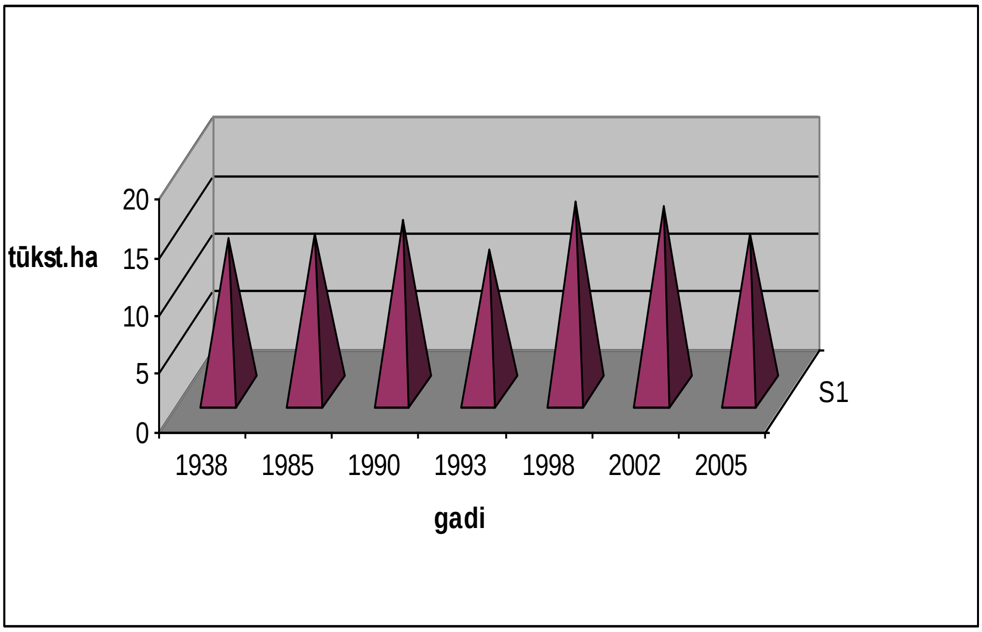
Grafiku veidošana tiek apgūta informātikas kursā, tādēļ par to šeit netiek minēts. Tālāk ir doti daži principi, kuri būtu jāievēro, zīmējot grafikus, lai tie būtu uzskatāmi, palīdzētu uztvert informāciju.

Ja primārais ir skaitliskā informācijas attēlošana (grafikam nav demagoģiski mērķi – sagrozīt informāciju), tad jāizvairās no sarežģītu formu lietošanas grafiskajos attēlos. Nav jālieto piramīdas, cilindri vai dažādu figūru attēli. Labāk veidot attēlus plaknē bez telpiskām formām, jo ir svarīgi atklāt pētāmā objekta skaitliskās attiecības, nevis dizainiski skaistus stabiņus vai kādus „ķiņķēziņus”.

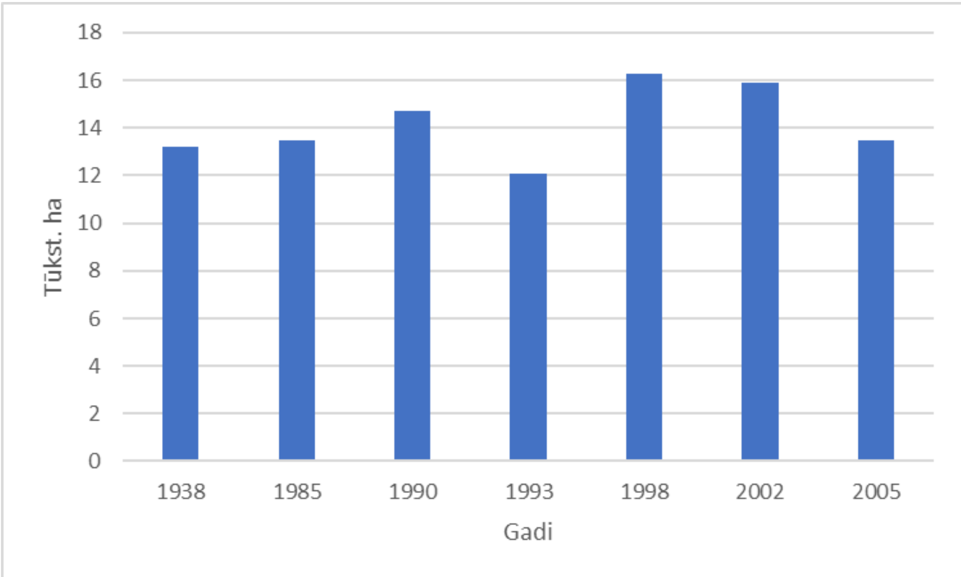
1.4. attēlā ir parādīta cukurbiešu sējumu platību dinamika. A attēlā sējplatības ir parādītas ar stilizētu cukurbiešu attēlu, B – telpiskām piramīdām, bet C – taisnstūriem plaknē.



A



B

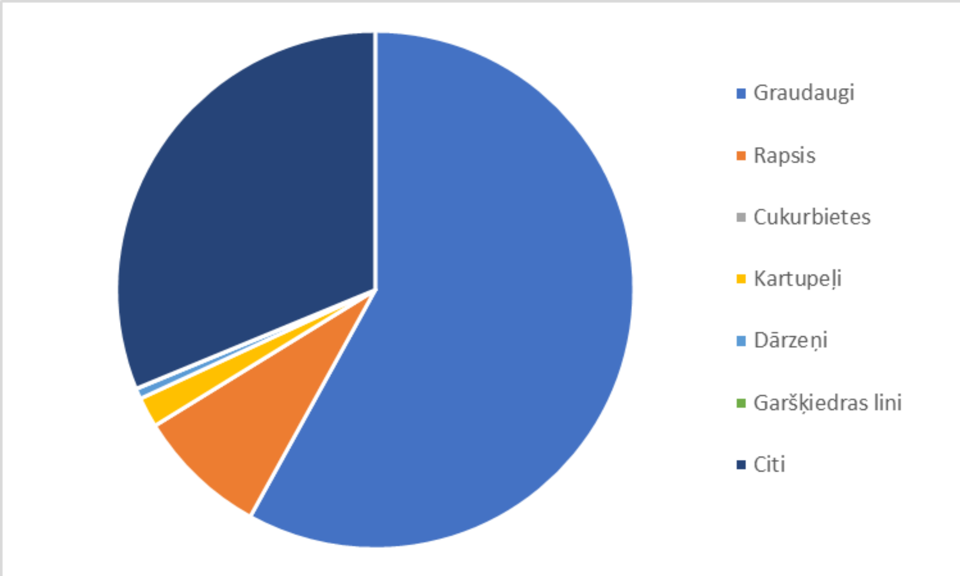


C

1.4. attēls. **Cukurbiešu sējumu platības dinamika, attēlota ar dažādiem grafisko attēlu veidiem**

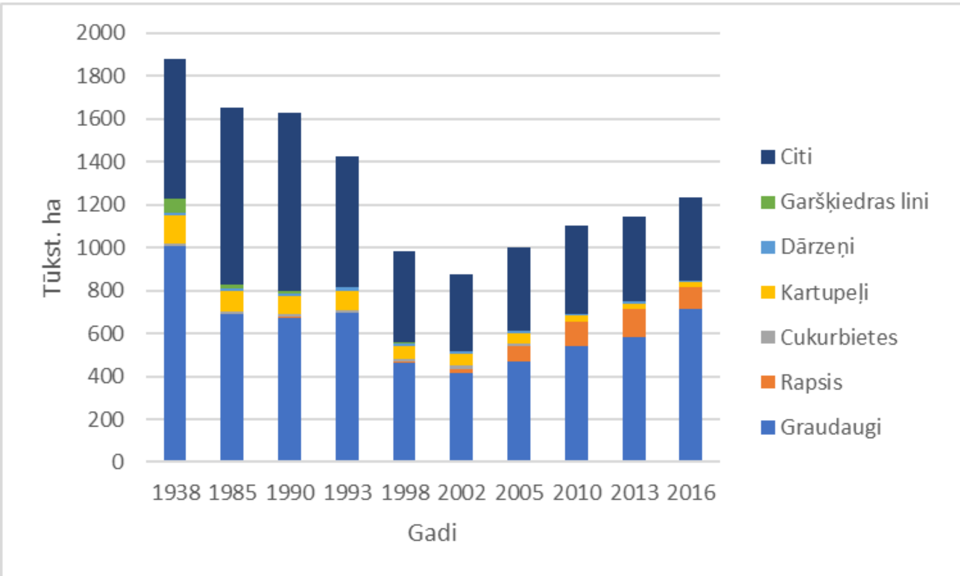
Vienkāršāka forma ļauj labāk uztvert attēloto informāciju.

Ja jāataino viena objekta struktūra (veselā sastāvdaļas), tad izmanto apļa diagrammu. Ja jāsalīdzina vairāku objektu struktūra vai arī viena objekta struktūras izmaiņas laikā, tad izmanto t. s. „sakrauto klucīšu” (*Stacked column*) grafiku.



1.5. attēls. **Apļa diagrammas paraugs** (attēlota sējumu struktūra 2005.gadā pēc datiem, kas redzami 1.3. attēla tabulā)

Šāda tipa grafiks (1.6. attēls) vienlaikus ļauj redzēt gan struktūru, gan dinamiku. Ja ir interese tikai par struktūru, tad var izvēlēties šī tipa grafiku ar vienādu stabiņu augstumu (dati tiks aprēķināti procentos).

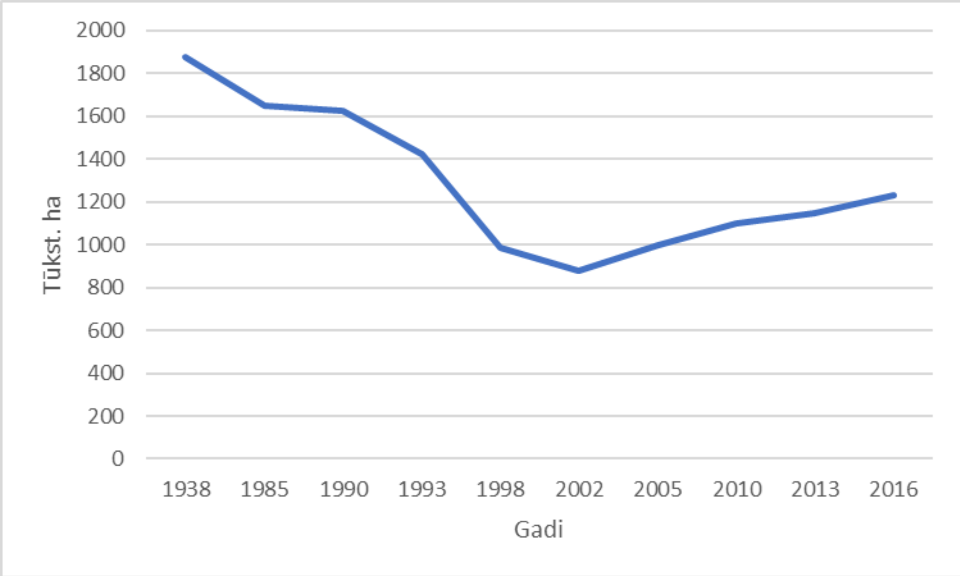


1.6. attēls. **„Stacked column" grafika paraugs** (attēlota sējumu struktūra pēc datiem, kas redzami 1.3. attēla tabulā)

Gan 1.5., gan 1.6. attēlā ir praktiski neredzamas maz pārstāvētās grupas. Ja interesē šo grupu precīzāka informācija, tad tas jāmeklē tabulās, jo attēlā ir redzams, ka dārzeņu, linu sējplatības kopējā sējumu struktūrā ir nenozīmīgas.

Ja jāsalīdzina līdzīgi objekti, tad izmanto stabiņu diagrammu, kur stabiņi novietoti blakus.

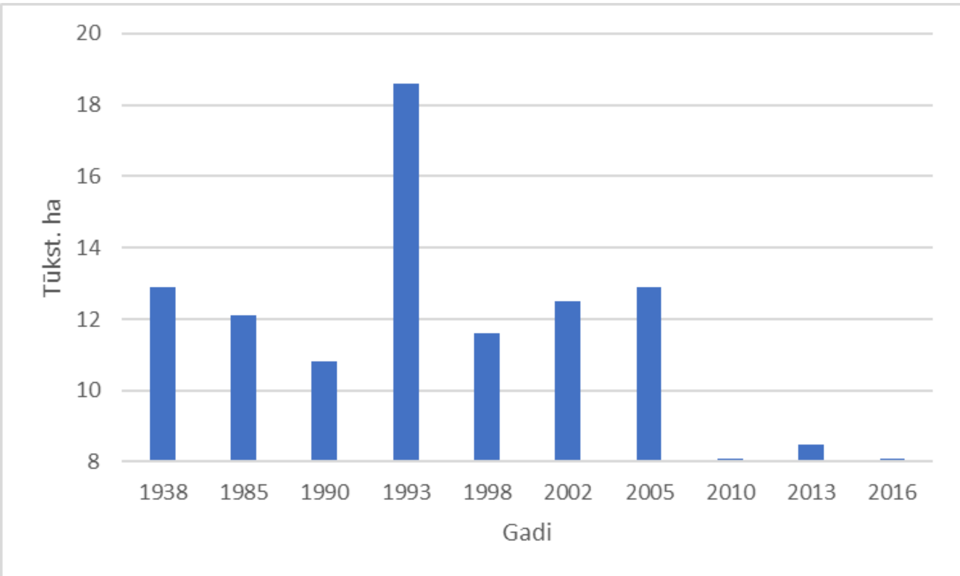
Viena rādītāja izmaiņas laikā (dinamikas rindas) labāk attēlot ar līniju grafiku.

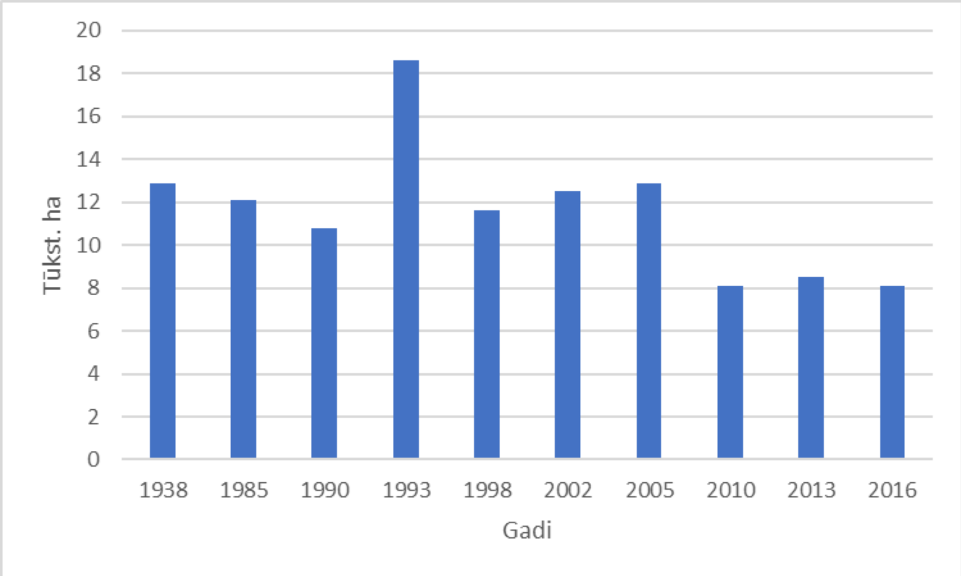


1.7. attēls. **Līniju grafiks dinamikas parādīšanai** (pēc datiem, kas redzami 1.3. attēla tabulā)

Grafikā ir trūkums, jo laika nogriežņi nav vienādi. Šajā situācijā tomēr piemērotāk būtu izmantot līniju grafiku, nevis punktu diagrammu (*XY Scatter*) un problēmu atrunāt grafika analīzes tekstā.

Grafiskajiem attēliem ordinātas skala ir jāsāk no nulles, nevis mazākās konstatētās vērtības. Ja grafika skalu sāk no mazākās konstatētā vērtības, tad rodas nepamatots priekšstats par ļoti nozīmīgām izmaiņām.





1.8. attēls. **Dārzeņu sējumu platības dinamika** (pēc datiem, kas redzami 1.3. attēla tabulā)

Augšējais attēls rada vizuālu iespaidu, ka sējplatībām ir ļoti ievērojamas svārstības. Lai saprastu situāciju, ir jāpievērš uzmanība ordinātas skaitļu asij.

Apakšējais attēls dod daudz pareizāku priekšstatu par dārzeņu sējumu platību izmaiņām laikā.

Jācenšas izvairīties no vizuālajām manipulācijām, kas rodas, grafiku saspiežot vertikāli vai horizontāli. Grafikam parasti cenšas atvēlēt laukumu ar šādu malu attiecību – vertikāle: horizontāle = 2 : 3.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
1	*2*	*3*	*4*
Alternatīva variācija	Alternative variation	Альтернативная вариация	Nominālās variācijas robežgadījums, kad pazīmei var būt tikai divas nozīmes, piemēram, dzimums – vīrietis vai sieviete. Dažreiz plašāk variējošo pazīmi apkopo divās grupās: interesējošā pazīme un citas pazīmes, piemēram, strādājošais ar augstāko izglītību vai bez tās
Atributīva pazīme	Categorical random variables	Категорийные переменные	Variantus var raksturot tikai ar jēdzieniem, piemēram, tautība, dzimums, šķirne, krāsa utt.
Biometrija	Biometry, biometrics	Биометрия	Parasti apzīmē statistisko pētījumu metožu pielietojumu bioloģiskajos, lauksaimnieciskajos un mežsaimnieciskajos pētījumos
Daļējā statistiskā novērošana	Nonprobability sampling or determinate sampling	Детермини-рованная выборка	Nepilnās statistiskās novērošanas nezinātniskais veids, kad pētījumā tiek iekļautas vienības, kas ir vieglāk pieejamas pētniekam, nav atlasītas pēc nejaušības principa
Diskrēta variācija	Discrete variation	Дискретная вариация	Kvantitatīva pazīmes variācija, ja novērojamās pazīmes vērtības var iegūt tikai kādas noteiktas vērtības, parasti tie ir veseli skaitļi, piemēram, televizoru skaits ģimenē
Ekonometrija	Econometrics	Эконометрия	Ekonometrija ir robežzinātne starp ekonomikas teoriju, statistiku un matemātiku. Kā robežzinātnei pētāmie jautājumi nav strikti noteikti un atkarībā no autora tie tiek vai nu paplašināti, vai sašaurināti. Ekonometrijā secinājumi balstās uz empīriskiem (apkārtējā pasaulē novērotiem) datiem
Faktoriālā pazīme	Factorial (independent) variable	Независимая переменная	Pazīme, kuras vērtību izmaiņas ietekmē citas pazīmes vērtības
Grupēšana	Aggregate data	Группирование	Viena no statistisko datu apstrādes pamatmetodēm. Grupējot sākotnējie dati tiek sistematizēti – izveidojas atsevišķas statistiskās kopas

1	2	3	4
Grupēšanas pazīme (pamats)	Aggregate variable	Переменная группирования	Pazīme, pēc kuras notiek kopas vienību sadalīšana grupās
Ģenerālkopa	Population	Генеральная совокупность	Visas izdalītās kopas vienības, par kurām vēlas iegūt informāciju
Hipotētiskā kopa	Hypothetic population	Гипотическая совокупность	Ģenerālkopa, kas visu laiku tiek papildināta, nav iespējams iegūt informāciju ar pilno statistisko novērošanu
Intervāla garums	Width of class interval	Ширина интервала группирования	Nepārtraukti variējošas pazīmes vērtību sadalījums grupās, starpība starp intervāla (gradācijas klases) augšējo (lielāko) un apakšējo (mazāko) vērtību. Parasti intervālu variācijas rindām ir vienāds intervālu garums, bet var būt arī vienā intervālu variācijas rindā intervāli ar atšķirīgu garumu
Intervālu skala	Interval scale	Интервальная шкала	Kvantitatīvas pazīmes aprakstīšanai, pazīmes vērtību raksturo ar skaitli. Vienādi intervāli raksturo vienādas pazīmes izmaiņas, taču nulle nenozīmē pazīmes trūkumu, piemēram, Celsija grādi temperatūras mērīšanai. Datu apstrādei var lietot visas kvantitatīvo datu apstrādes metodes
Intervālu variācijas rinda	Interval frequency array	Интервальный вариационный ряд	Divas savstarpēji saistītas skaitļu rindas, kur viena raksturo novērojumu vērtību intervālus nepārtrauktai variācijai, bet otra – novērojumu skaitu ar vērtību, kas ir pētāmā intervāla robežās
Izlase (paraugkopa)	Probability sampling	Вероятностная выборка	Nepilnās statistiskās novērošanas zinātniskais veids, kad pētījumā tiek iekļautas vienības, kas ir atlasītas pēc nejaušības principa tā, lai labi pārstāvētu (reprezentētu) visu pētāmo kopu
Kumulatīvā variācijas rinda	Cumulative frequency array	Кумулятивный ряд распределения	Uzrāda variantus vai intervālus un uzkrātos (summāros) biežumus. Tajos ieskaita visas tās kopas vienības, kurām reģistrētā pētāmās pazīmes vērtība ir mazāka vai vienāda ar attiecīgā intervāla augšējo robežu

1	2	3	4
Kvantitatīvā pazīme	Numeralical random variables	Числовые переменные	Varianšu vērtības var izteikt ar skaitļiem, piemēram, alga, ražība, pašizmaksa, atzīme eksāmenā, svars u. c.
Nepārtraukta variācija	Continuous variation	Непрерывная вариация	Kvantitatīva pazīmes variācija, ja novērojamās pazīmes vērtības var iegūt jebkuru vērtību ierobežotā vai neierobežotā apgabalā, piemēram, ķermeņa masa
Nepilnā statistiskā novērošana	Sampling	Выборка	Informācijas iegūšana par daļu no interesējošās kopas vienībām, iegūtos rezultātus vispārina un attiecina uz visu pētāmo kopu
Nominālā skala	Nominal scale	Номинальная шкала	Izmanto, lai aprakstītu kvalitatīvu (atributīvu) pazīmju (jēdzienu vai jēdzienu kodu) vērtības
Ordinārā (kārtas) skala	Ordinal scale	Порядковая шкала	Kvalitatīvas variācijas aprakstīšanai, ja pazīmes vērtības var saranžēt, bet intervāli starp atsevišķām variantēm nav vienādi, piemēram, izglītības līmeņi – pamatizglītība, vidējā, augstākā. Datu raksturošanai lieto modu, mediānu, biežumu sadalījumus, datu apstrādei – neparametriskās metodes, rangu korelācijas
Pazīme	Variable	Переменная	Objektu raksturojošās īpašības, par kurām statistiskās novērošanas procesā tiek iegūta informācija
Pilnā statistiskā novērošana	Full statistical observations	Полное статистическое наблюдение	Informācijas iegūšana par visām interesējošās kopas vienībām
Proporcionālā (attiecību) skala	Ratio scale	Шкала отношений	Kvantitatīvas pazīmes aprakstīšanai, pazīmes vērtību raksturo ar skaitli. Vienādi intervāli raksturo vienādas pazīmes izmaiņas, nulle nozīmē pazīmes trūkumu, piemēram, ķermeņa masa. Datu apstrādei var lietot visas kvantitatīvo datu apstrādes metodes
Psihometrija	Psychometric	Психометрия	Ar statistisko metožu lietošanu psiholoģiskajos un pedagoģiskajos pētījumos saistītā zinātnes nozare, kurā nodarbojas ar psiholoģisko parādību mērīšanu

1	2	3	4
Rezultējošā pazīme	Dependent variable	Зависимая переменная	Pazīmes, kuru vērtību izmaiņas ietekmē kādas citas pazīmes vērtību izmaiņas
Statistika	Statistics	Статистика	Metodes, kuras lieto datu savākšanas, attēlošanas, analīzes un interpretācijas procesā
Statistiskā novērošana	Statistical observations	Статистическое наблюдение	Datu savākšanas process statistiskajos pētījumos
Variācija	Variability	Вариация	Novēroto un reģistrēto datu dažādība
Variācijas amplitūda	Range	Диапазон вариации	Starpība starp novērojumos konstatēto lielāko un mazāko pazīmju vērtībām.
Variācijas rinda	Variable array	Вариационный ряд	Divas savstarpēji saistītas skaitļu rindas, kur viena raksturo novērojumu vērtības, bet otra – novērojumu skaitu ar šādu vērtību

2. EMPĪRISKIE SADALĪJUMI UN TO ATTĒLOŠANA

Pēc nodaļas apgūšanas studentiem:

- *jāprot grupēt datus;*
- *jāprot aprēķināt absolūtos, relatīvos un kumulatīvos biežumus;*
- *jāprot attēlot datus histogrammā, kumulātā un poligonā;*
- *jāprot apstrādāt datus ar programmu Excel.*

2.1. Grupēšana

Ja pētāmais objekts jāraksturo pēc vienas pazīmes, tad grupē pēc šīs pašas pazīmes. Piemēram, raksturojot studentu sekmību, izveido studentu grupas (grupē novērojumus) pēc atzīmju līmeņiem; raksturojot tirdzniecības apgrozījumu; sagrupē apgrozījuma novērojumus eiro. Jēdziens „**empīrisks**" nozīmē "iegūts pieredzes ceļā".[1] Veicot novērojumus dabā – apkārtējā pasaulē, iegūst empīriskos novērojumus, savukārt, tos sagrupējot, veido empīriskos sadalījumus. Nākamajās nodaļās tiks apskatīti arī teorētiskie sadalījumi, kas nav kāda konkrēta objekta novērojumi, bet gan teorētiskie modeļi, ar kuriem tiek salīdzināti empīriskie novērojumi un tiek izdarīti secinājumi, vai objekta variācija dabā atbilst kādam teorētiskam modelim.

Sadalījuma jeb **variācijas** rindu iegūst, saskaitot, cik reizes reģistrēti vienādi varianti. Ja biežumi izsaka tiešo kopas vienību skaitu, tad tie ir **absolūtie biežumi**.[2,3,4]

Biežumus var izteikt procentos vai daļās, visu kopas apjomu pieņemot par 100 % vai 1. Tā iegūst **relatīvos biežumus**.[5] Ja pazīmei ir diskrēta variācija (varianti ir nošķirti, parasti veseli skaitļi, piemēram, bērnu skaits ģimenē), tad katrs variants var būt atsevišķā pazīmes nozīme. Ja diskrētam sadalījumam nozīmju ir pārāk daudz, tad variantus apvieno intervālos. Savukārt nepārtraukti variējošajai pazīmei grupēšana ir iespējama tikai intervālos.

---

[1] Guļevska, D. (atb.red.). (1987). *Latviešu valodas vārdnīca : A-Ž.* Rīga : Avots. 221. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 21. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 24.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 13. lpp.
[5] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 21. lpp.

Kādreiz, kad nebija datortehnikas, bija ļoti svarīgi vienkāršot aprēķinus. Tad grupēšanu izmantoja, lai samazinātu kalkulācijas apjomu, kas nepieciešams, lai aprēķinātu sadalījumu raksturojošos lielumus. Tas bija ārkārtīgi svarīgi pirms kalkulatoru un pat kalkulatoru laikmetā, kad nebija pieejami personālie datori un tabulu procesori, kā tagad, kad agrāk ļoti darbietilpīgu kalkulāciju mūsdienās var ātri un ērti veikt ar programmas *Microsoft Excel* palīdzību. Neskatoties uz to, ka grupēšana ir zaudējusi savu nozīmi kā aprēķinu atvieglotāja, tā saglabā savu nozīmi kā patstāvīga analīzes metode.

Vispirms tiks apskatīts, kā grupēšanu veic manuāli (neizmantojot datortehnikas palīdzību).

2.1. piemērs. Uzņēmums, kas nodarbojas ar tiešo pārdošanu, ir ievācis datus par 52 aģentu ceturkšņa pārdošanas rezultātiem. Kā redzams, dažiem aģentiem veicas labāk, citiem sliktāk. Negrupējot šo informāciju, neko vairāk par iepriekšējo secinājumu, ka pārdošanas apjomi ir svārstīgi, izdarīt nevar.

Tālāk ir dots datu masīvs, kur ir uzskaitīti aģentu pārdošanas ieņēmumi eiro.

6365	31783	20742	9150	22798	16115	16717	38544	16876	12744
25603	18147	24735	21126	16378	42741	18585	22614	14828	23464
12865	17541	51810	21510	20980	18363	14873	47917	20548	9416
14905	19299	28083	23212	15722	10253	13833	27546	24386	24994
27109	15315	29271	19074	47964	17063	12089	48036	18564	6712
35620	48624								

Kā redzams, no šiem datiem kaut ko saprast ir ļoti grūti. Tādēļ tos ir vērts sagrupēt un izanalizēt.

1. etapā atrod minimālo (x_{min}) un maksimālo (x_{max}) reģistrēto vērtību.

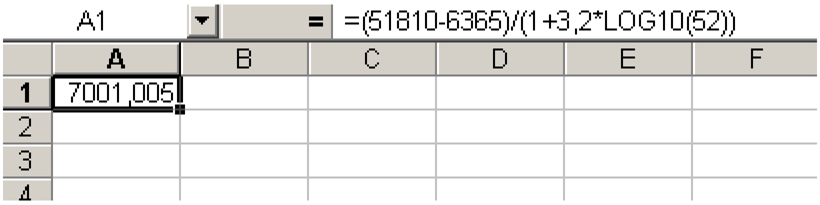
x_{min} – 6365 €.

x_{max} –51810 €.

Šī informācija jau ir nedaudz derīga analīzei, jo ekstrēmās (galējās) vērtības neraksturo galveno tendenci, bet ir atkarīgas tikai no atsevišķas veiksmes vai neveiksmes. Piemēram, no kāda aģenta lielu preču apjomu vienreiz iepērk nevis gala patērētāji, bet kāds starpnieks. Ja šis starpnieks atradīs tiešo piegādātāju, tad šis ekstrēmi augstais apgrozījums neraksturo aģenta darbības efektivitāti. Svarīgi ir izvērtēt to informāciju, kas atrodas starp galējām vērtībām – vai biežāk sastopami aģenti, kuri sasniedz lielu vai mazu apgrozījumu, cik liels apgrozījums tiek sasniegts visbiežāk.

Nākamajā etapā ir jāaprēķina grupējuma intervāla garums. Tā kā novērojumu skaits nepārsniedz 100, tad var izmantot Sterdžesa formulu

(1.1. formula). Sterdžesa formulas aprēķinos ikdienas lietošanā problēmas rada tajā ietvertais logaritms. Studentiem, kā arī praktizējošiem ekonomistiem parasti nav zinātniskie kalkulatori, bet parastajiem kalkulatoriem nav logaritma aprēķināšanas funkcijas. Gadījumos, ja zinātniskais kalkulators tomēr ir, tad tā īpašniekam bieži nav vajadzīgo iemaņu kalkulatora lietošanā, aprēķinot matemātiskās funkcijas. Daudz plašāk par zinātniskajiem kalkulatoriem ir izplatīti personālie datori, un tur šo aprēķinu var veikt vai nu programmā Excel, vai arī ar datora ieprogrammēto zinātnisko kalkulatoru. 2.1. attēlā ir parādīts intervāla garuma aprēķins Excel programmā. Formulas rindā var redzēt ievadīto formulu, lielumus var norādīt arī ar šūnu adresēm, kuras satur vajadzīgos lielumus.



2.1. attēls. **Excel tabulas fragments ar intervāla garuma aprēķināšanas formulu**

Ja nu nav nevienas no šīm iespējām, tad intervālu skaitu var izvēlēties voluntāri, tā ir pētnieka brīva izvēle. Intervālu skaits ir no 4 līdz padsmit, parasti robežās no 5 līdz 8. Ja novērojumu ir mazāk, tad arī izdalīto grupu – intervālu skaits ir mazāks. Ja novērojumu ir vairāk, tad var izdalīt vairākas grupas.

Kā redzams 2.1. attēlā, tad aprēķinātais intervāla garums ir 7001,005 €. Tāds intervāla garums, kas ir nenoapaļots skaitlis, ir ļoti neparocīgs. 1.intervāls būs no 6365 € līdz 6365 € + 7001,005 € = 13366,005 € utt. Nerunājot par to, ka lieka ir aprēķina precizitāte līdz centa daļām, arī līdz eiro noapaļots intervāla garums ir ļoti neērts grupēšanai, jo ir grūti noteikt, kuram intervālam kurš novērojums pieder. Aprēķināto intervāla garumu noapaļo līdz skaitlim, kas beidzas ar 5 vai arī vienu vai vairākām nullēm. Apaļošanas precizitāti nosaka novērojumu skaitļu mērogs, šajā gadījumā noapaļo līdz tūkstošiem. Intervāla garumu noapaļo tā, lai tas nesaturētu vairāk par diviem zīmīgiem cipariem, parasti tikai vienu zīmīgo ciparu (nulles, kas izveidojas apaļošanas rezultātā, skaitļa beigās nav zīmīgie cipari). Tātad šajā piemērā intervāla garums būs vai nu 7, vai 8 tūkstoši. Ja apaļo uz leju, tad izdalīto grupu (klašu) skaits būs lielāks, ja uz augšu – mazāks. Pirmā intervāla apakšējo robežu izvēlas tā, lai tā būtu nedaudz mazāka par zemāko reģistrēto lielumu. Ja novērojumu reģistrācija būs tikai pēc grupējuma izveides, tad grupējuma robežas nosaka iepriekšējā pieredze un loģiski apsvērumi. Apakšējā intervāla zemāko robežu nosaka ar tādu pašu precizitāti kā intervāla garumus (šajā gadījumā intervāla garums tika noteikts tūkstošos). Apskatītajā piemērā pirmā intervāla zemākā (kreisā) robeža ir 6000 € (zemākā reģistrētā vērtība ir 6365 €). Šai gadījumā varētu intervālus veidot arī no nulles. Tālāk aprēķina intervālu robežas, pie iepriekšējā intervāla augstākās (labās) robežas pieskaita intervāla garumu un iegūst nākamā intervāla labo robežu un izveido grupēšanas *tabulu (2.1. tabula). Intervālu robežu aprēķināšanu turpina tik ilgi, kamēr pēdējā intervāla labā robeža ir lielāka par augstāko reģistrēto lielumu. Intervālus sauc arī par gradācijas klasēm.*[1]

2.1. tabula

Tirdzniecības aģentu apgrozījuma grupēšanas tabulas makets

Gradācijas klases, tūkstošos €	*Novērojumi*	*Kopā*
6 – 14		
14 – 22		
22 – 30		
30 – 38		
38 – 46		
46 –54		

Klašu robežvērtība, piemēram, 14 tūkstoši €, ir norādītas divās klasēs. Parasti ir pieņemts nolasīt klases zemāko robežu "vairāk nekā ...", bet augšējo – "ieskaitot".[2] *Šajā piemērā zemākā vērtība, kura tiks iereģistrēta 2. gradācijas klasē, būs 14 001 €, bet augtākā – 22 000 €. Dažreiz šādā veidā arī norāda gradācijas klašu robežas: 2. klase – no 14 001 līdz 22 000, 3. klase – no 22 001 līdz 30 000 utt.*

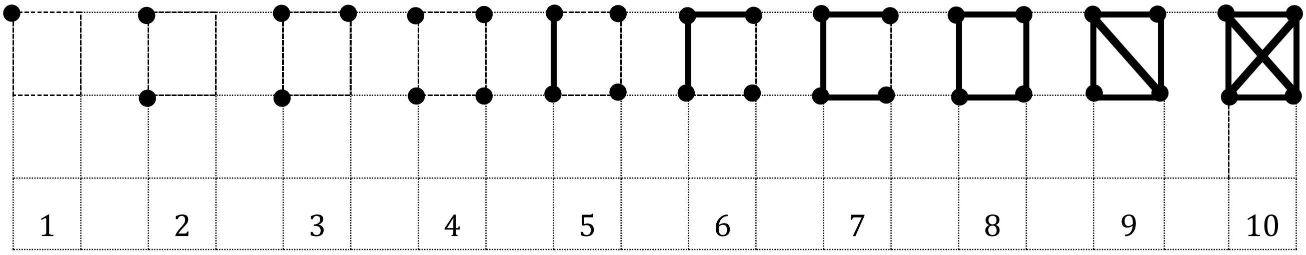
Kad ir noteiktas klašu robežas, ir jāsaskaita, cik novērojumu atbilst katrai klasei. Grupējot parasti neskata cauri visu datu masīvu un neskaita, cik tieši novērojumu atbilst 1., 2. un citām klasēm. Šādi rīkojoties, pastāv liela iespēja kļūdīties. Pareizi ir skatīt reģistrētos lielumus pēc kārtas un katru iereģistrēt atbilstošajā klasē. Novērojumu reģistrāciju pa gradācijas klasēm veic par katru novērojumu, ievelkot vienu vertikālu svītriņu.

Lai novērojumu skaitu vēlāk būtu vieglāk saskaitīt, tad četrus novērojumus atzīmē ar vertikālām svītrām (||||), bet piekto novērojumu atzīmē ar horizontālu svītru, kas šķērso pirmās četras vertikālās svītras (~~||||~~).[3] *Piemēram, ja kādā gradācijas klasē ir atzīmēta šāda svītriņu kombinācija:*

~~||||~~ ~~||||~~ |||*, tad viegli var noteikt, ka šajā klasē ir 13 novērojumi.*

Alternatīva svītriņu metodei ir novērojumu reģistrācija ar punktu un svītru kombinācijām (decimālā atzīmēšanas sistēma), kur katram novērojumu skaitam atbilst noteikta punktu un svītriņu kombinācija.[4]

[1] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 103. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 13. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 24.
[4] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 13. lpp.



2.2. attēls. **Decimālā atzīmēšanas sistēma**

Šādā veidā, aizpildot vienu kvadrātiņu, tiek uzskaitīti 10 novērojumi. Reģistrējot nākamos novērojumus, zīmē nākamo kvadrātiņu.

2.2. tabulā ir parādīts 2.1. piemērā dotā datu masīva grupējums.

2.2. tabula

Aizpildīta aģentu pārdošanas ieņēmumu grupēšanas tabula

Gradācijas klases, tūkstošos €	*Novērojumi*	*Kopā*
6 – 14	𝍸 𝍷𝍷𝍷𝍷	*9*
14 – 22	𝍸 𝍸 𝍸 𝍸 𝍷𝍷	*22*
22 – 30	𝍸 𝍸 𝍷𝍷	*12*
30 – 38	𝍷𝍷	*2*
38 – 46	𝍷𝍷	*2*
46 – 54	𝍸	*5*
Kopā		*52*

Saskaitot pa gradācijas klasēm reģistrētos novērojumus, ir jāsanāk kopējam reģistrēto novērojumu skaitam.

Šāda veida grupēšanas tabula, protams, ir darba lapa, taču tā jau dod priekšstatu par sadalījumu un par to var izdarīt pirmos secinājumus – kuras gradācijas klases ir vairāk pārstāvētas, kuras – mazāk. Dabas un sociālekonomiskajos novērojumos parasti vidējās klases ir izteikti pārstāvētas, bet klasēs ar mazām un lielām vērtībām ir sastopams mazāk novērojumu. Pētītajā piemērā visplašāk pārstāvētā klase ir aģenti ar apgrozījumu no 14 līdz 22 tūkstošiem € (22 aģenti), labi pārstāvētas ir arī 3. klase ar apgrozījumu no 22 līdz 30 tūkstošiem € (12 aģenti) un klase ar vismazāko apgrozījumu – no 6 līdz 14 tūkstošiem €. Ārpus normāli sagaidāmā „izlec" klase ar vislielāko apgrozījumu, kur ir 5 novērojumi, kamēr divās iepriekšējās klasēs ir tikai pa diviem novērojumiem. Pētniekam, analizējot to, kas ir ārpus statistikas rāmjiem, jānoskaidro, vai šis apgrozījums ir nejaušība, vai tomēr eksistē kāda likumsakarība, piemēram, zem *viena aģenta vārda darbojas vairāki izplatītāji. Sadalījums ir nedaudz asimetrisks – plašāk pārstāvētas ir klases ar mazajiem apgrozījumiem.*

Novērojumu skaitu pa gradācijas klasēm sauc par **absolūtajiem biežumiem** vai **frekvencēm** un apzīmē ar “*f*” vai “f_i”, kur indekss “*i*” ir gradācijas klases numurs pēc kārtas. Apskatītajā piemērā $f_1 = 9$, bet $f_2 = 22$ utt.

Absolūtos biežumus var aizstāt ar **relatīvajiem**, aprēķinot sadalījuma struktūru procentos vai viena daļās. Relatīvās frekvences ir piemērotākas, ja jāsalīdzina 2 vai vairāk kopas, kurā ir atšķirīgs novērojumu skaits, piemēram, sekmības līmenis divās studentu grupās (skat. 2.2. piemēru).

2.2. piemērs. Eksāmena atzīmju sadalījums divās studentu grupās

Vērtējums eksāmenā	*Studentu skaits 1. grupā*	*Studentu skaits 2. grupā*
1	0	0
2	2	0
3	4	1
4	8	3
5	18	6
6	26	12
7	20	8
8	13	5
9	5	4
10	1	0
Kopā	97	39

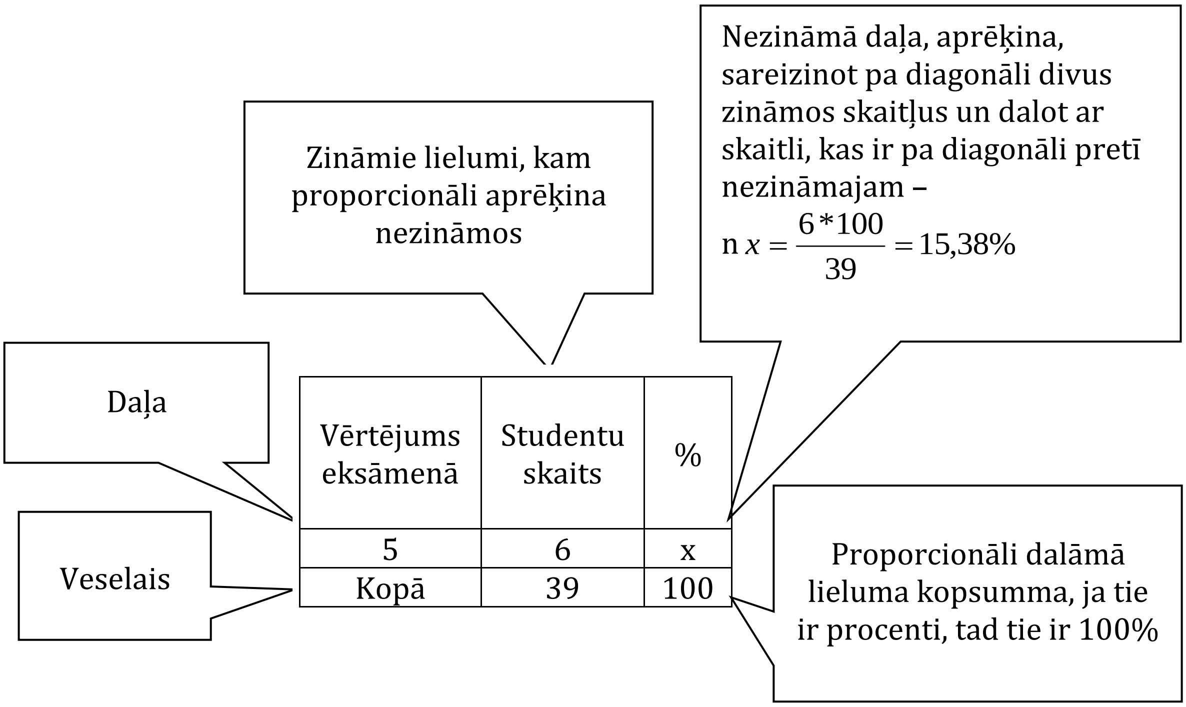
2.3. tabulā ir parādīta relatīvo biežumu aprēķināšana. Novērojumu skaitu (šajā gadījumā studentu skaitu, kas saņēmuši attiecīgo atzīmi) sauc par absolūto biežumu, bet attiecīgā novērojuma īpatsvaru procentos vai viena daļās – relatīvo biežumu.

2.3. tabula

Divu studentu grupu eksāmena atzīmju absolūtie un relatīvie biežumi

Vērtējums eksāmenā	*Absolūtais biežums*		*Relatīvais biežums, %*	
	1.grupas studentu skaits	*2.grupas studentu skaits*	*1. grupā*	*2. grupā*
1	*0*	*0*	*0*	*0*
2	*2*	*0*	*2,1*	*0,0*
3	*4*	*1*	*4,1*	*2,6*
4	*8*	*3*	*8,2*	*7,7*
5	*18*	*6*	*18,6*	*15,4*
6	*26*	*12*	*26,8*	*30,8*
7	*20*	*8*	*20,6*	*20,5*
8	*13*	*5*	*13,4*	*12,8*
9	*5*	*4*	*5,2*	*10,3*
10	*1*	*0*	*1,0*	*0,0*
Kopā	*97*	*39*	*100*	*100*

Procentu aprēķini ir proporciju aprēķināšana. Tabulā ar sākotnēji neaizpildītām šūnām ir gatavas proporcijas. Kā aprēķināt proporcijas, ir parādīts 2.3. attēlā, dati ir fragments no 2.2. piemēra.



2.3. attēls. **Daļas procentu aprēķins**

Studentu skaits 1. grupā ir vairāk nekā divas reizes lielāks par 2. grupu. Loģiski, ka visās atzīmju grupās ir vairāk novērojumu. Tas, ka eksāmenu uz 6 ballēm 1. grupā nokārtoja par 14 cilvēkiem vairāk neko nenozīmē, bet, salīdzinot relatīvos biežumus, būtiskas atšķirības starp studentu grupām neredz. Arī iepriekš minētajā (kas ir nopelnījusi 6 balles eksāmenā) grupā nav kardinālu atšķirību: 1. grupā – 27 % studentu, bet 2. grupā – 31 % eksāmenu ir nokārtojuši uz 6 ballēm. 2. studentu grupa šķiet sekmīgāka, bet atšķirības varētu būt arī nejaušības rezultātā un nevar droši teikt, ka 2. studentu grupa mācās labāk nekā 1. grupa.

Tas, kā novērtēt, vai atšķirības ir būtiskas, vai nē (hipotēžu pārbaude), tiks apskatīts nākamajās nodaļās.

Lai vizuāli novērtētu sadalījumu, to attēlo grafiski. Visbiežāk lietotais grafiskais sadalījuma attēls ir histogramma.

2.2. Empīrisko sadalījumu grafiskā attēlošana

Histogrammu lielākoties lieto intervālu variācijas rindu attēlošanai. Uz abscisu (*X*) ass atliek pazīmes vērtības – parasti attēlotie lielumi ir intervālu (gradācijas klašu) robežas. Pieņemot nogriežņus no zemākās līdz augtākajai intervāla robežai par pamatiem, konstruē taisnstūrus, kuru augstumi ir proporcionāli gradācijas klašu biežumiem. Var izmantot kā absolūtos, tā relatīvos biežumus. Šādi attēlo vienāda garuma intervālu variācijas rindu.[1,2,3]

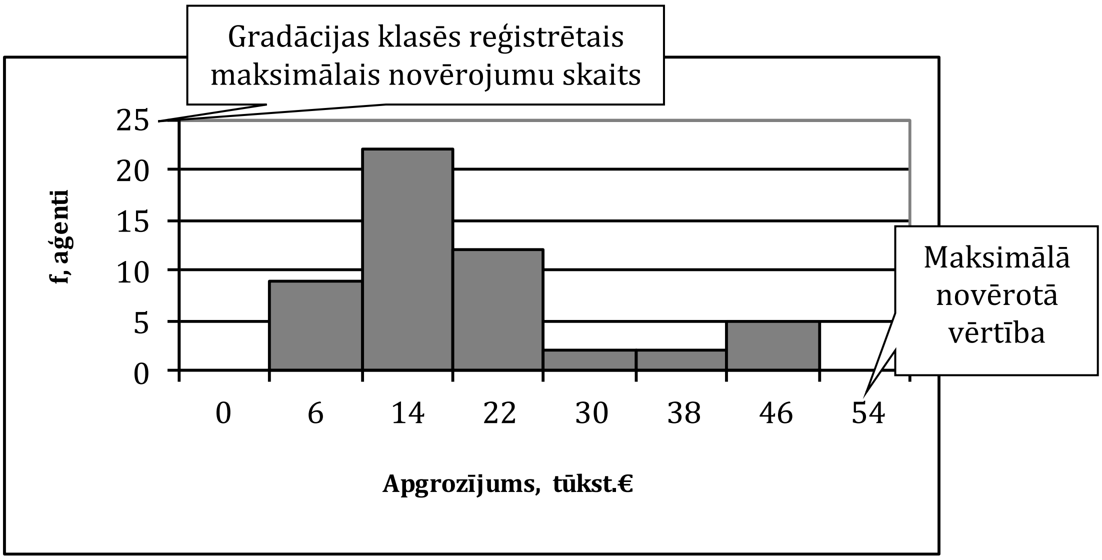
Histogrammu var uzzīmēt trīs veidos:

- manuāli (ar roku);
- ar datoru, izmantojot parastās *Excel* grafiku veidošanas iespējas grupētiem datiem;
- ar datoru, izmantojot datu analīzes (*Data Analysis*) rīku "Histogramma" (*Histogram*) negrupētiem datiem.

Ja histogrammu zīmē ar roku, tad grafikam atvēlētajā vietā uzzīmē koordinātu asis, izvēlas asu mērogu tā, lai grafiks aizņemtu visu tam atvēlēto laukumu – lai nepietrūktu vietas un nebūtu "iespiests" atvēlētā laukuma stūrītī. 2.4. attēls ir izveidots ar vienkāršu datora Excel programmu, bet tur tiek demonstrēti noteikumi, kas jāievēro, zīmējot grafiku ar roku.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 15. lpp.
[2] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 9.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 55. lpp.



2.4. attēls. **Histogramma pēc 2.2. tabulas datiem ar paskaidrojumiem**

Lai izveidotu tādu histogrammu, kā redzams 2.4. attēlā, ar parastajiem *Excel Chart* līdzekļiem iegūtais attēls ir jāformatē.

1. *Excel* darba lapā ievada datus – vienā ailē gradācijas klases, norādot robežas no … līdz … Šādi ievadītus datus dators uztvers kā tekstu un izmantos *X* ass kategoriju apzīmēšanai, 2. ailē ievada katras klases novērojumu skaitu.
2. Nospiež pogu "Chart Wizard".
3. Izvēlas grafika tipu*. Step 1 of 4. Chart Type*: izvēlas *Column*, 1.subtipu (tas tiek piedāvāts automātiski), *Next*>
4. Norāda izmantojamo datu apgabalu. *Step 2 of 4*. *Data range*. Ja darba lapā nekā cita nav, tad dators norāda automātiski pieejamos datus, atliek tos pārbaudīt un vajadzības gadījumā koriģēt. Nepieciešamos datu apgabalus norāda, iedrukājot šūnu adreses vai arī nepieciešamo datu apgabalu iezīmē ar peli, *Next*>
5. Grafika iespēju izvēlnē *Chart Options* norāda asu nosaukumus – iedrukā atbilstošo tekstu lodziņos *Category (X) axis* un *Value (Y) axis*. Ja histogrammu veido kursa- vai diplomdarbam, tad grafika nosaukumu nenorāda, jo pēc metodiskajiem noteikumiem nosaukumi ir jāraksta zem attēla. Ja histogramma tiek gatavota citām vajadzībām, tad var norādīt arī tās nosaukumu, iedrukājot to lodziņā *Chart title*. Tas parādīsies virs grafika. Izvēlnē "*Legend*" (apzīmējumi) noņem "ķeksīti" pretī lodziņam "*Show legend*". Šajā gadījumā tā ir lieka un neko neizsakoša informācija. *Next*>
6. Tiek piedāvāts nākamais solis *Step 4 of 4 – Chart Location* – grafika novietojums. Automātiski piedāvā grafiku tajā pašā darba lapā, tādēļ atliek tikai nospiest "*Finish*" – beigt.
7. Tālāk grafiku var turpināt formatēt, veicot ar peles kreiso taustiņu dubultklikšķi uz tās grafika daļas, kuru vēlas mainīt:
   - var noņemt fona krāsu (*Format Plot Area. Patterns. Area*)*,* izvēlas baltu fonu, *OK*;
   - formatējot datu sēriju (*Format Data Series*), var mainīt stabiņu krāsu pēc izvēles (*Patterns, Area*)*,* izvēlas krāsu un sabīda histogrammas stabiņus kopā – logā *Options, Gap width* uzstāda nulli, *OK*.

Grafiks ir pabeigts, to var iekopēt programmas *Word* dokumentā, aprakstīt un izanalizēt.

Sadalījumu grafiskai attēlošanai lieto arī poligonu. Poligonu galvenokārt izmanto diskrētu variācijas rindu attēlošanai. **Poligonu** zīmējot, koordinātu sistēmā atliek punktus, kuru koordinātas attiecīgi ir pazīmes vērtības (variantes) un to biežumi. Blakus esošos punktus savienojot ar taisnes nogriežņiem, iegūst poligonu.[1,2,3] Mūsdienās parasti arī diskrētu sadalījumu attēlo ar histogrammu. Tas ir pat vienkāršāk, nekā attēlojot intervālu variācijas rindu, jo dators pazīmes vērtību automātiski novieto zem stabiņa, kuru tas raksturo. Abi (histogramma un poligons) grafisko attēlu veidi ir savstarpēji aizvietojami.

Poligonu ar datoru var uzzīmēt no grupētiem datiem, izvēloties līniju (*Line*) grafika tipu, izveidošana un formatēšana līdzīga, kā zīmējot histogrammu (stabiņu grafiks – *Column*).

Dažreiz ekonomiskajā statistikā novērojumus grupē nevienāda garuma intervālos. Šādos gadījumos intervāla (gradācijas klases) absolūtos vai relatīvos biežumus aizvieto ar intervāla blīvumu. Par **intervāla blīvumu** sauc intervāla biežuma attiecību pret intervāla garumu.[4] Intervāla blīvums raksturo kopas vienību skaitu, rēķinot uz vienu variējošās pazīmes vienību. *2.4. tabulā ir pārgrupēti 2.2. piemēra 1. studentu grupas eksāmenu rezultāti.*

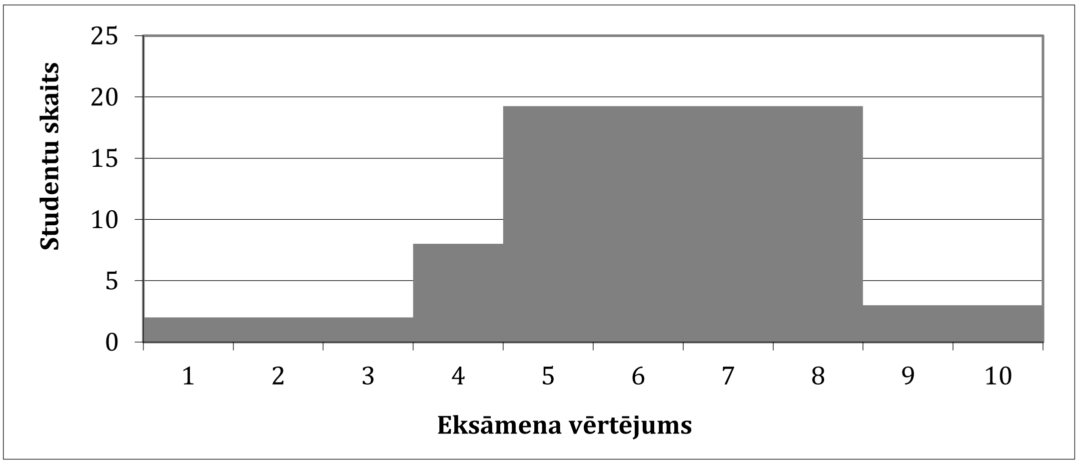
2.4. tabula

Studentu grupējums pēc sekmības līmeņiem un intervālu blīvums

Vērtējums eksāmenā	*Grupas raksturojums*	*Novērojumu skaits*	*Intervāla blīvums*
1 – 3	*Nesekmīgi*	*6*	*2*
4	*Gandrīz nesekmīgi*	*8*	*8*
5 – 8	*Normāli zinošie, programmas ietvaros*	*77*	*19,25*
9 – 10	*Labāk par programmas prasībām zinošie*	*6*	*3*
Kopā		*97*	

2.5. attēlā ir parādīta 2.4. tabulai atbilstošā histogramma.

[1] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 27. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 25.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 56. lpp.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 15. lpp.



2.5. attēls. **Studentu grupas sekmības intervālu blīvuma histogramma**

Nevienāda garuma intervālus lieto arī sociālekonomiskajos pētījumos, ja pēdējos intervālos ir maz vai nav nemaz novērojumu (tukši intervāli). Lai efektīvi parādītu visu informāciju, pēdējos intervālus apvieno. Ja nevienāda garuma intervālu lietošanu neprasa pētījuma loģika, tad lieto vienāda garuma intervālus.

Vēl empīrisko sadalījumu raksturošanai lieto kumulatīvos biežumus un gafiskiem attēliem – kumulātu. Kumulatīvie (summārie) biežumi var būt gan absolūtie, gan relatīvie, un tos var traktēt kā kopas vienību skaitu, kurām vērtība ir mazāka nekā apskatāmā intervāla augšējā robeža.[1,2,3] *2.5. tabulā parādīta kumulatīvo biežumu aprēķināšana tirdzniecības aģentu apgrozījuma 2.1. piemēram.*

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 15. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 42.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 61. lpp.

2.5. tabula

Kumulatīvo biežumu aprēķināšana

Apgrozījums, tūkstošos €	*Absolūtie biežumi*	*Kumulatīvie biežumi*
6 – 14	*9*	*9*
14 – 22	*22*	*22 + 9 = 31*
22 – 30	*12*	*12 + 31 = 43*
30 – 38	*2*	*45*
38 – 46	*2*	*47*
46 - 54	*5*	*52*
Kopā	*52*	

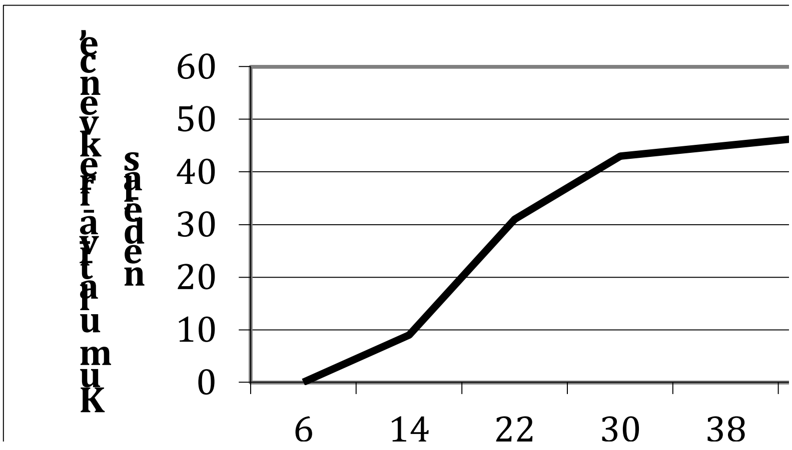
Apgrozījums līdz 22 tūkstošiem eiro ceturksnī ir bijis 31 aģentam

Kumulatīvais biežums pēdējā gradācijas klasē ir vienāds ar kopējo novērojumu skaitu

2.5. tabulā ar bultiņām un aritmētiskajām darbībām ir parādīts, kā aprēķina kumulatīvos biežumus. Kumulatīvo biežumu vērtība pieaug pa gradācijas klasēm. Ja kādā intervālā nav bijis novērojumu, tad kumulatīvais biežums paliek iepriekšējā līmenī, bet summārais (kumulatīvais) biežums nekad lielākajā gradācijas klasē nav mazāks par zemākās klases kumulatīvo biežumu.

Kumulātu iegūst, attēlojot kumulatīvo variācijas rindu. Koordinātu sistēmā atliek punktus, kuru abscisas (*X* ass vērtības) ir proporcionālas variantu lielumiem (gradācijas klašu labajām robežām), bet ordinātas (*Y* ass vērtības) – attiecīgo variantu uzkrātajiem biežumiem (absolūtajiem vai relatīvajiem). Savienojot blakus esošos punktus ar taisnes nogriežņiem, iegūst lauztu līniju – kumulātu.

2.6. attēlā ir parādīta kumulāta tirdzniecības aģentu apgrozījuma 2.1. piemēram. Kumulātu no grupētiem datiem zīmē ar Excel rīku “Chart Wizard” kā līniju tipa (Line) grafiku.



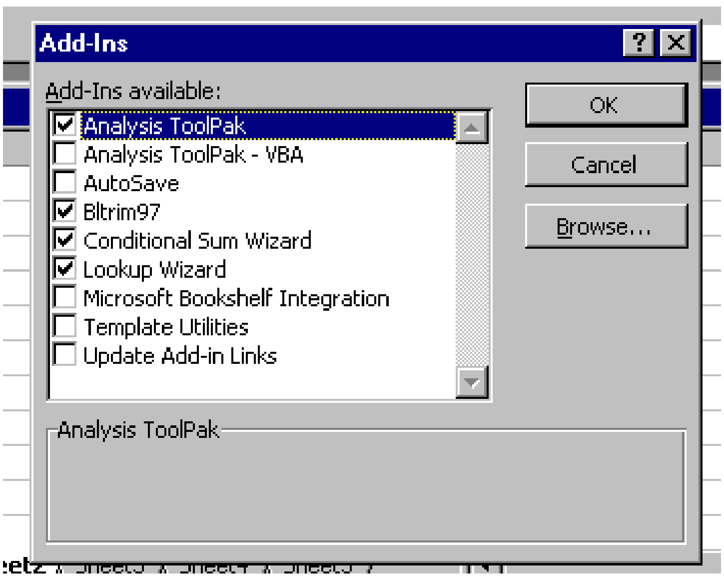
2.6. attēls. **Kumulāta aģentu apgrozījuma piemēram** (pēc 2.5. tabulas datiem)

Histogrammu un kumulātu var iegūt arī ar *Microsoft Excel* datu analīzes (*Data anlysis*) rīku "histogramma".

2.3. Empīrisko sadalījumu attēlošana un analīze ar programmas *Excel* datu analīzes rīku

Lai attēlotu empīrisko sadalījumu, ir jāveic secīgas darbības.

Jāatver programma *Excel* un jāpārliecinās, vai datu izvēlnē ir pieejams datu analīzes (*Data analysis*) rīks. Ja izvēlnē *Dati* (*Data* – angļu versijā) nav datu analīzes rīka, tad tas ir jāpievieno ar komandām. Jānospiež loga piktogramma ekrāna augšējā kreisajā stūrī/ Pievienojumprogrammas (*Add- Ins*)/ *Excel* pievienojumprogrammas *„go"*/ *Analysis ToolPak*. 2.5. attēlā ir parādīts dialoga logs, kurā ir jāieliek ķeksītis pretī analīzes rīku pakotnei (*Analysis ToolPak*). Ja vajadzīgās rindiņas burti ir gaiši pelēki un ķeksīti nevar ielikt, tas nozīmē, ka šī iespēja konkrētajā datorā nav instalēta. Šādā situācijā jālūdz palīdzība datorspeciālistam instalēt vajadzīgo *Excel* komponentu.



2.7. attēls. ***Excel* dialoga logs datu analīzes rīka aktivizēšanai**

Kad datu analīzes rīks datu izvēlnē ir pieejams, var turpināt pārējās darbības, kas nepieciešamas histogrammas zīmēšanai.

Jāievada visi skaitļi vienā ailē. Histogrammas zīmēšanai datus var norādīt arī ar datu apgabalu. Tā kā parasti tiem pašiem datiem aprēķina aprakstošās statistikas (*Descriptive statistics*) rādītājus un tos programma *Excel* rēķina katrai ailei atsevišķi, datus ievada vienā ailē. *Excel* statistiskos aprēķinus veic tikai <u>negrupētiem</u> datiem

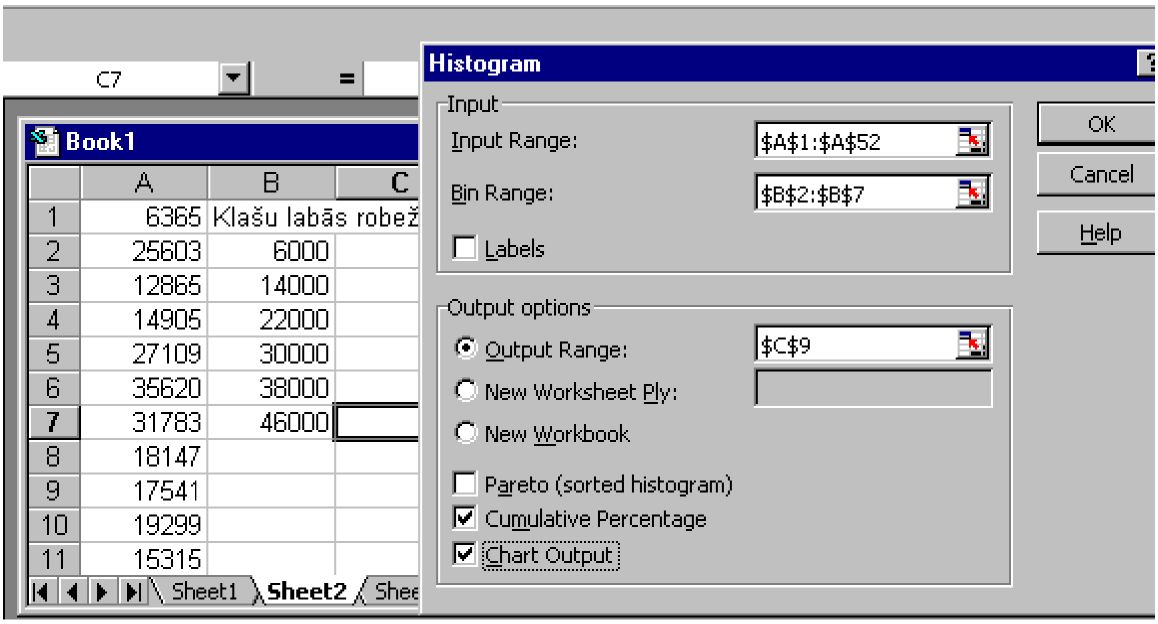
Jāizpilda komandas: *Dati/Data Analysis/Histogram*.

Input Range jānorāda datu apgabals. „*Bin Range*” norāda klašu labās robežas (šūnu adreses), ja tās ir aprēķinātas. Pēdējās klases labo robežu nenorāda, jo dators automātiski ņem vēl papildus vienu klasi (*more*). Ja klašu robežas nav aprēķinātas, tad šo lodziņu var atstāt neaizpildītu, bet tāda gadījumā intervāla garumu un klašu robežas dators aprēķina pēc formulas, un analizēt skaitļus, kas satur daudz zīmīgo ciparu, ir sarežģītāk nekā tad, ja klašu robežas ir noapaļotas.

Output Range jānorāda šūna, no kuras sākt grupējuma rezultātu un histogrammas zīmējuma izvadi.

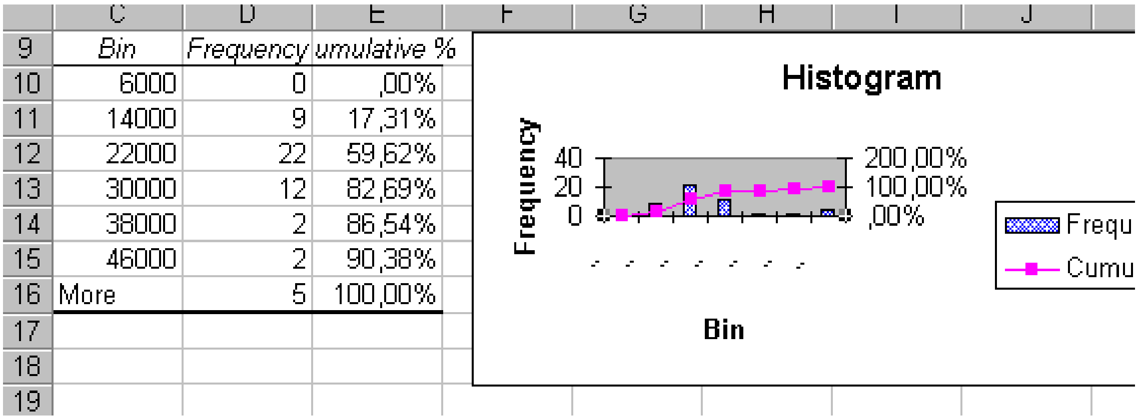
Jāieliek ķeksītis lodziņā *Cumulative Percentage* un *Chart Output/OK.*

2.8. attēlā ir redzams *Excel* darba lapas fragments ar histogrammas konstruēšanas dialoga logu.



2.8. attēls. ***Excel*** **darba lapas fragments ar dialoga logu histogrammas konstruēšanai**

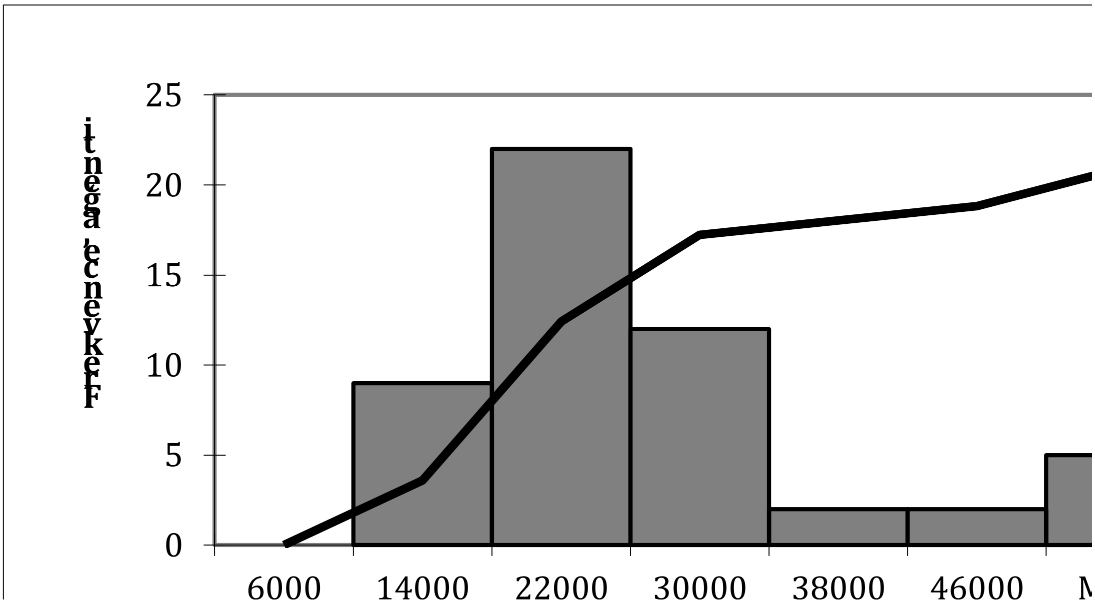
Nākamajā 2.9. attēlā ir parādīts *Excel* darba lapas fragments, kāds tiek iegūts pēc iepriekšējo komandu izpildes.



	C	D	E
9	Bin	Frequency	umulative %
10	6000	0	,00%
11	14000	9	17,31%
12	22000	22	59,62%
13	30000	12	82,69%
14	38000	2	86,54%
15	46000	2	90,38%
16	More	5	100,00%

2.9. attēls. ***Excel*** **darba lapas fragments ar grupējuma tabulu un histogrammu**

Gan tabulu, gan grafiku var un vajag formatēt pirms tālākās izmantošanas prezentācijai studiju darbā, diplomdarbā vai atskaitē. Tabulai, aktivizējot interesējošo šūnu, var aizvietot anglisko tekstu ar latvisko, var izveidot tādas robežas (*border*), kā vēlas. Noformēto tabulu var iekopēt *Word* dokumentā vai *Power Point* prezentācijā. Tabulas formatēšanu, ja nepieciešams, var veikt arī *Word* dokumentā. Grafiks jāformatē *Excel* darba lapā. Uzklikšķinot uz atbilstošās grafika daļas, to aktivizē, un var aizvietot angliskos tekstus ar latviskiem, uzrakstīt vajadzīgos (atbilstošos pētījuma būtībai) asu nosaukumus, nevajadzīgo izdzēst. 2.10. attēlā ir parādīta formatēta histogramma.



2.10. attēls. **Ar datu analīzes rīka *"Histogram"* palīdzību iegūtā un formatētā histogramma aģentu apgrozījuma 2.1. piemēram**

Tālāk iegūtā histogramma un grupējums tiek analizēts. Nosaka, kādi ir visbiežāk novērotie lielumi, kāda ir histogrammas simetrija, datu izkliedes plašums un vienmērīgums u. tml.

Datu grupēšana un grupējumu atspoguļošana tabulās un grafikos atvieglo statistiskā materiāla uztveri, ļauj to analizēt. Taču arī šādi apkopota informācija bieži vien ir pārāk apjomīga, un ir vajadzīga lielāka informācijas saspiešanas pakāpe. Lai raksturotu sadalījumu, ir vajadzīgi daži rādītāji, kas atspoguļo sadalījuma novietojumu uz skaitļu ass. Šos rādītājus sauc par **lokācijas rādītājiem**.[1] To aprēķināšana un interpretācija tiks apskatīta nākamajā nodaļā.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 17. lpp.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
1	*2*	*3*	*4*
Absolūtie biežumi	Frequency	Абсолютные частоты	Novērojumu skaits tiešajās vienībās (cilvēki, nedēļas, uzņēmumi u.tml.)
Aprakstošā statistika	Descriptive statistics	Описательная статистика	Statistikas nozare, kura apraksta atsevišķus variējošos rādītājus, izmanto grupēšanu, grafiskās metodes un lokācijas rādītājus
Empīrisks sadalījums	Empirical distribution	Эмпирическое распределение	Dabā (apkārtējā vidē) novērots kādas pazīmes sadalījums
Gradācijas klases	Class	Класс эмпирического распределения	Izdalītās grupas (varianti diskrētam vai atributīvam sadalījumam, intervāli nepārtrauktam sadalījumam), veicot grupēšanu
Histogramma	Histogram	Гистограмма	Sadalījuma grafisks atveidojums, kur klašu pārstāvniecību parāda ar stabiņu diagrammu
Intervāla blīvums	Interval density	Плотность интервала	Lieto, ja ir nevienāda garuma intervāli. Garākajiem intervāliem aprēķina vidējo novērojumu skaitu uz īsākā intervāla apjomu vai arī visiem intervāliem var rēķināt uz vienu pazīmes vienību
Kumulāta	Cumulative frequency curve	Кумулята (кривая накопленных частот)	Kumulatīvo biežumu grafisks attēlojums ar lauztu līniju, kumulatīvos biežumus atliek pret klašu augšējām robežām
Kumulatīvie biežumi	Cumulative frequency	Кумулятивная (накопленная) частота	Novērojumu, kas ir mazāki nekā attiecīgā intervāla augstākā robeža, skaits – summējas visi novērojumi no zemākajām klasēm
Lokācijas rādītāji	Statistics	Описательные статистики	Skaitļi, kuri raksturo sadalījuma izvietojumu uz skaitļu ass
Poligons	Frequency polygon	Полигон	Sadalījuma grafisks atveidojums, kur klašu pārstāvniecību atliek ar punktiem, kurus savienojot iegūst lauztu līniju – poligonu
Relatīvie biežumi	Relative frequency	Относительные частоты	Novērojumu skaits izteikts viena daļās vai procentos
Sadalījuma (variācijas) rinda	Distribution series	Ряд распределения	Sadalījuma grupu un tām atbilstošo novērojumu uzskaitījums

3. LOKĀCIJAS RĀDĪTĀJI

Pēc nodaļas apgūšanas studentiem:

- *jāzina pakāpju vidējo lietošanas situācijas;*
- *jāprot aprēķināt svērto aritmētisko vidējo, pamatot tā lietošanu;*
- *jāzina struktūras vidējie (moda un mediāna) un citi struktūras rādītāji (kvantiles), to lietošana;*
- *jāzina dažādi variācijas rādītāji, to lietošana un interpretācija;*
- *jāprot aprēķināt standartnovirzi, lietojot norādīto formulu;*
- *jāsaprot asimetrijas un ekscesa koeficientu interpretācija;*
- *jāprot aprēķināt aprakstošās statistikas rādītājus ar Microsoft Excel līdzekļiem un tos interpretēt.*

Statistisko informāciju analizē, grupējot un grupējumu atspoguļojot tabulā un grafiskajos attēlos. Taču, ja nepieciešama informācija vēl īsākā formā, tad ir vajadzīgs viens vai vairāki rādītāji, kas raksturotu sadalījumu novietojumu uz skaitļu ass. Šim nolūkam noder **lokācijas rādītāji.**

Statistiskie dati variē, un šajā dažādībā interesi rada biežāk pārstāvētie lielumi un datu izkliedes plašums, to koncentrācija. Sadalījuma lokāciju (novietojumu) raksturo dažādi vidējie lielumi. Parasti tos sauc **par sadalījuma centrālās tendences rādītājiem**.[1]

Sadalījumi var atšķirties ar apgabalu (datu variācija), ko tie aizņem uz abscisu (X) ass. Pie šīs grupas pieder tādi rādītāji kā **variācijas apjoms** jeb **amplitūda, vidējā absolūtā novirze, standartnovirze** jeb **vidējā kvadrātiskā novirze, variācijas koeficients** u. c.[2,3,4]

Sadalījumi atšķiras ar savām simetrijas īpašībām. Dažreiz vairāk pārstāvētas ir variantes ar mazākām vērtībām, citreiz – gluži otrādi. Tādēļ ir vajadzīgi rādītāji, kas raksturo variācijas rindas **asimetriju**.[5,6]

Sadalījuma smailumu sauc par **ekscesu**. Ja vairums novērojumu ir tuvu centrālajai tendencei, tad ir pozitīvs ekscess – sadalījuma līkne ir smaila, bet, ja

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 63. lpp.

[2] Turpat, 68. lpp.

[3] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 44.lpp.

[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 35. lpp.

[5] Turpat, 53. lpp.

[6] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 74.

novērojumi ir sastopami puslīdz vienmērīgi pa visu datu apgabalu, tad sadalījuma līkne ir plakana un ekscess negatīvs.[1,2]

Tālāk tiks detalizēti iztirzāti lokācijas rādītāji, to aprēķināšana un interpretācija.

3.1. Centrālās tendences rādītāji

Statistiskajos pētījumos dati ir variējoši, piemēram, studenti uzrāda dažādu sekmības līmeni, aģentu pārdošanas apjoms ir atšķirīgs u.tml., bet ir jānoskaidro, kāda ir galvenā tendence, vidējā atzīme eksāmenā, vidējais pārdošanas apjoms, biežāk sastopamā eksāmena atzīme. Ir jāatrod lielums, kuru varētu izmantot, lai prezentētu visu datu kopumu. Šādus rādītājus sauc par centrālās tendences rādītājiem. Vienkāršākais no tiem ir aritmētiskais vidējais. To plaši lieto sadzīviskā līmenī bez statistikas zināšanām, un tas ir arī biežāk lietotais centrālās tendences rādītājs statistiskajos pētījumos. Statistiskajos pētījumos lieto arī citus centrālās tendences rādītājus.

Visi centrālās tendences rādītāji tiek iedalīti divās grupās:

- pakāpju vidējie;
- struktūras vidējie.[3,4,5]

Pakāpju vidējo aprēķināšanas vispārējā formula vienkāršajiem vidējiem:

$$\bar{x} = \sqrt[z]{\frac{\sum_{i=1}^{n} x_i^z}{n}} \tag{3.1.}$$

un svērtajiem vidējiem:

$$\bar{x} = \sqrt[z]{\frac{\sum_{i=1}^{k} x_i^z f_i}{\sum_{i=1}^{k} f_i}} \tag{3.2.}$$

kur x_i – i novērojums (variante), bet, rēķinot variācijas rindas vidējo, variants – intervāla centrs vai klases vidējais;

n – kopas vienību skaits;

k – gradācijas klašu skaits;

f_i - i gradācijas klases vienību skaits jeb varianta statistiskais svars;

z – pakāpes rādītājs, kas katram vidējā tipam ir atšķirīgs.

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem.* Rīga: Izglītības soļi. 76. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 50. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 18. lpp.
[4] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 68. lpp
[5] Lasmanis, A. (1999). *Datu ieguves, apstrādes un analīzes metodes pedagoģijas un psiholoģijas pētijumos*. Rīga : Mācību apgāds NT. 94. lpp.

Biežāk izmantotie pakāpju rādītāji ir šādi:

- –1 – harmoniskais vidējais;
- 0 – ģeometriskais vidējais;
- 1 – aritmētiskais vidējais;
- 2 – kvadrātiskais vidējais utt.

Katram pakāpes vidējam ir sava darba formula un pielietojuma joma.

3.1.1. Aritmētiskais vidējais

Vienkāršo aritmētisko vidējo lieto negrupētu datu vidējā lieluma aprēķināšanai – novērojumu kopsummu izdala ar novērojumu skaitu (3.3. formula).

Vienkāršais aritmētiskais vidējais:

$$\bar{x} = \frac{\sum x_i}{n} \qquad (3.3.)$$

Ir situācijas, kad ir jālieto svērtais aritmētiskais vidējais (3.4. formula).

Svērtais aritmētiskais vidējais:[1,2,3]

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \qquad (3.4.)$$

Statistikā ir daudz formulu, un to visu atcerēšanās no galvas nav nepieciešama, bet ir jāsaprot formula un jāvar to pielietot atbilstošās situācijās.

Pazīmes lielumus parasti apzīmē ar *x*. Vispārējā formā novērojumu apzīmē ar x_i – *i* novērojums, konkrētos novērojuma apzīmējumus (x_1, x_2 utt. parasti nelieto).

$\overline{x}$ – vidējā pazīmes vērtība.
f_i – *i* grupas novērojumu skaits (frekvence).
Σ – summas zīme, kas nozīmē saskaitīšanu pēc kārtas, dažreiz tiek norādītas summēšanas robežas, parasti vispārējā formā summēt no *i* = 1 (pirmā) līdz *n* (pēdējam) novērojumam.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 19. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 66.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 36. lpp.

Svērto aritmētisko vidējo visbiežāk lieto, aprēķinot relatīvo lielumu vidējo. Ir trīs uzdevumu grupas, kuru ietvaros lieto svērtos vidējos.[1]

Kā pirmā tiks apskatīta uzdevumu grupa, kur pazīme *x* ir diskrēta. Svērto aritmētisko vidējo aprēķina no grupētiem datiem, par statistiskajiem svariem ņemot vienību skaitu katrā grupā. Aritmētisko vidējo šādā gadījumā var aprēķināt arī ar vienkāršā aritmētiskā vidējā formulu (3.3.), bet tas nav racionāli.

Aprēķina būtību var izanalizēt studentu atzīmju piemēram (2.3. tabula). 1. studentu grupai jāaprēķina vidējais eksāmena vērtējums.

Ja rēķina ar vienkāršā vidējā formulu, tad ir jāsummē 97 atzīmes un ir viegli kļūdīties, jo ļoti rūpīgi jāseko tam, vai, piemēram, ir saskaitīti visi 26 sešinieki utt.

$$\bar{x} = \frac{2+2+3+3+3+3+4+\ldots+8+9+9+9+9+9+10}{97} = 6{,}11$$

Izmantojot vidējo svērto, aprēķins kļūst ievērojami vienkāršāks, samazinās kļūdīšanās iespēja.

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n}$$

$$\bar{x} = \frac{1*0+2*2+3*4+4*8+5*18+6*26+7*20+8*13+9*5+10*1}{0+2+4+8+18+26+20+13} = 6{,}11$$

Statistikā (arī citos ekonomiskajos aprēķinos) bieži sastopami iepriekšējai formulai līdzīgi aprēķini. Šādas formulas satur dažādas darbības – reizināšanu, dalīšanu, kāpināšanu u. c., un iegūtie starprezultāti ir jāsasummē. Šādas formulas nerēķina izvērstā veidā, kā tas ir parādīts iepriekšējā aprēķinā, bet gan tabulas formā (3.1. tabula).

Otrā uzdevumu grupa, kur jālieto aritmētiskā vidējā svērtā formula, ir tad, ja, vadoties pēc kvalitatīviem (profesionāliem) apsvērumiem, visas statistiskā objekta vienības nav vienādi lielas (nozīmīgas, informatīvas) vidējā lieluma veidošanā. Situācijas izpratnei var apskatīt divus piemērus. Lai saprastu, kā sistēma darbojas, jāizpēta tās reakcija ekstrēmās situācijās. Šajā sakarā pirmais piemērs: pagastā ir divas graudu ražošanas saimniecības – viena liela (400 ha sējplatība un graudu ražība 40 cnt/ha), bet otra maza (20 ha sējplatība un graudu ražība 20 cnt/ha). Ja vidējo ražību aprēķina kā vienkāršo aritmētisko vidējo, tad vidējā ražība būtu 30 cnt/ha ((20+40)/2). Tas nav pareizi, jo saimniecības nav vienādi lielas. Vidējā ražība būs tuvāk 40 cnt/ha, jo lielākā sējplatība ir devusi tādu ražību. Vidējo ražību aprēķina ar vidējo aritmētisko svērto, par frekvencēm ņemot sējplatības. (400*40+20*20)/(400+20) = 39,05 cnt/ha. Šajā gadījumā aprēķina kopražu, kuru vēlāk izdala ar kopējo sējumu platību. Tā ir jārīkojas vienmēr, ja ir jārēķina relatīvo rādītāju vidējie. Relatīvos

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 20. lpp.

rādītājus var pazīt pēc lieluma mērvienības, tā būs attiecība (€/v., st./v., cnt/ha, u. c., kā arī %).

3.1. tabula

Aritmētiskā vidējā svērtā aprēķins tabulas veidā studentu eksāmenu rezultātu piemēram

Atzīme eksāmenā, x_i	*Studentu skaits, f_i*	*$x_i f_i$*
1 x_1	*0* f_1	*0* $=x_1*f_1$
2 x_2	*2*	*4*
3	*4*	*12*
4	*8*	*32*
5	*18*	*90*
6	*26*	*156*
7	*20*	*140*
8	*13*	*104*
9	*5*	*45*
10	*1*	*10*
Kopā	*97* Σf_i	*593* Σx_i*f_i

Lai aprēķinātu aritmētisko vidējo, atliek tikai izdalīt 593 ar 97.

Tagad tiks apskatīts piemērs, kam nepieciešama lielāka kalkulācija, apgūstot arī iemaņas strādāt ar kalkulatoru. 3.2. tabulā ir parādīta gan sākotnējā, gan aprēķinu informācija.

3.2. tabula

Zemnieku saimniecību vidējās graudu ražības aprēķins

Saimniecības numurs pēc kārtas, i	*Graudu ražība, cnt/ha, xi*	*Graudu sējplatība, ha, fi*	*Kopraža, cnt, xifi*
1.	*22,4*	*13,6*	*304,64*
2.	*12,8*	*23,2*	*296,96*
3.	*30,5*	*5,8*	*176,9*
4.	*40,1*	*28,2*	*1130,82*
5.	*42,7*	*15,4*	*657,58*
Kopā	*X*	*86,2*	*2566,9*

$\overline{X}$ = 2566,9/86,2 = 29,78 cnt/ha. Tātad vidējā ražība apskatāmajā saimniecību grupā ir nepilni 30 centneri no hektāra.

Pirms apskatīt, kā jāveic aprēķini ar kalkulatoru, nedaudz jāmin par kalkulatora izvēli, ja tāda nav vai esošais neapmierina. Mūsdienās ir ļoti plaša kalkulatoru izvēle, sākot no maziem un beidzot ar sarežģītiem zinātniskiem kalkulatoriem, kuriem ir grafiskās un pat drukas iespējas. Var būt daudz un dažādi apsvērumi, izvēloties kalkulatoru, tāpēc jau tie ir tik dažādi, bet tomēr ir daži ieteikumi ekonomikas studentiem racionāla kalkulatora izvēlei:

1) kalkulatoram ir jābūt pietiekami lielam, lai, ātri veicot aprēķinus, vienlaikus nespiestos vairāki taustiņi;
2) ekonomiskajos aprēķinos parasti nav nepieciešamas sarežģītas matemātiskās funkcijas, tāpēc diez vai būtu jāizvēlas zinātniskais kalkulators. Turklāt zinātniskā kalkulatora lietošana prasa zināmas iemaņas, kuras ir jāapgūst, patstāvīgi, rūpīgi studējot klāt pievienoto instrukciju. Ja instrukcija netiek studēta, tad ir iegūts dārgs, bet neērts kalkulators;
3) statistikas kursa vajadzībām (dažreiz arī citos ekonomikas priekšmetos) ir vajadzīga kvadrātsaknes funkcija, tādēļ, pērkot kalkulatoru, jāskatās, lai būtu šī funkcija;
4) saprātīga cena un kalkulatora ilgmūžība.

Laba izvēle ir “*semi – desk*” klases kalkulatori, kas var kalpot visus studiju gadus un vēl ilgāk.

Viena no iespējām, kas ļoti samazina kalkulācijas apjomu, ir kalkulatora atmiņas izmantošana. Daudzās statistikas formulās ir summas zīme (Σ). Tas nozīmē, ka ir jāveic kādas darbības (reizināšana, dalīšana, kāpināšana) un iegūtie starprezultāti ir jāsummē. Ja prot izmantot kalkulatora atmiņu, tad šīs darbības var veikt paralēli. Pirms uzsākt kalkulatora atmiņa izmantošanu, ir jāpārliecinās, ka kalkulatora atmiņā nekas nav ieskaitīts (kalkulatora displejā nav redzams burts “M” vai vārds “*memory*”). Ja atmiņā kaut kas ir, tad pirms kalkulatora atmiņas lietošanas tā ir jāizdzēš. To izdara, nospiežot taustiņu “MC” vai divreiz – taustiņu “MRC”, ja atmiņas dzēšanas taustiņš atsevišķi nav izdalīts. Pirmo reizi nospiežot, displejā parādās atmiņā esošais skaitlis, bet, otrreiz nospiežot, tas tiek no atmiņas izdzēsts. Zinātniskajiem kalkulatoriem dažreiz šādu taustiņu nav, tad informāciju no atmiņas dzēš, nospiežot taustiņu “X→M” – displejā esošais skaitlis aizvietos atmiņā esošo skaitli, tādēļ, atmiņu dzēšot, displejā ir jābūt nullei.

Tālāk tiks apskatītas veicamās darbības atbilstoši 3.2. tabulas aprēķiniem.

Kalkulatorā ievada “22,4 X 13,6 =”, pieraksta tabulā iegūto rezultātu un nospiež taustiņu “M+”. Nākamā darbība – “12,8 X 23,2 =”, pieraksta rezultātu, nospiež “M+”. Tā turpina līdz pēdējais reizinājums (starprezultāts) ir ievadīts kalkulatora atmiņā, tad nospiež taustiņu “MR” vai “MRC” un displejā parādās summa. Iegūto rezultātu ieraksta tabulā, pārbauda aprēķinu pareizību, saskaitot piereģistrētos starprezultātus. Ja rezultāts abos gadījumos sakrīt, tad turpina aprēķinus, ja nē – meklē kļūdu. Rēķinot, neizmantojot atmiņu, tiktu iegūts tikai rezultāts, bet, šādi rēķinot, ar vienādu darbietilpību tiek iegūts rezultāts, un tas tiek pārbaudīts. Kad šīs iemaņas nostiprinās, tad starprezultātus var arī nepiereģistrēt. Ja starprezultātus pieraksta, tad svarīgi to izdarīt vienmēr vienā konkrētā momentā (pirms vai pēc ievadīšanas kalkulatora atmiņā). Ja to tā nedara un starprezultātu atmiņā ievada, kad pagadās, tad pieaug kļūdīšanās iespējas. Kalkulatora atmiņu izmanto arī, ja manuāli jāaprēķina summa lielam datu masīvam. Steigā ievadot skaitļus, var gadīties kādu ievadīt nepareizi. Lai varētu veikt pārbaudi un samazinātu kopējo aprēķinu daudzumu, aprēķinus veic pa grupām (parasti stabiņiem), starprezultātus pieraksta un summē kalkulatora atmiņā. Ja, otrreiz pārrēķinot, kāda summa nesakrīt, tad jāpārrēķina tikai kļūdainā grupa, nevis viss datu masīvs.

Trešā aprēķinu grupa, kur lieto vidējo svērto, ir aritmētiskā vidējā aprēķins pēc intervālu variācijas rindas. Ja ir grupēti dati, bet sākotnējo datu nav vispār vai arī nav laika, lai veiktu pilnīgi precīzus aprēķinus pēc sākotnējiem datiem, tad apmierinošas precizitātes aprēķinu var iegūt, aprēķinot vidējo svērto no grupētiem datiem.

3.3. tabulā ir parādīts, kā aprēķina vidējo tirdzniecības aģentu apgrozījuma piemēram (2.1. piemērs).

3.3. tabula

Vidējā apgrozījuma aprēķina tabula tirdzniecības aģentu apgrozījuma piemēram

Gradācijas klases, tūkstošos €	*Nedēļas, fi*	*Intervāla centrs, xi*	*xi*fi*
6 – 14	*9*	*10*	*90*
14 – 22	*22*	*18*	*396*
22 – 30	*12*	*26*	*312*
30 – 38	*2*	*34*	*68*
38 – 46	*2*	*42*	*84*
46 –54	*5*	*50*	*250*
Kopā	*52*		*1200*

*Vidējais apgrozījums ir 23 tūkstoši 77 € (1200/52). Ja to pašu aprēķinu veic pēc sākotnējiem datiem (2.1. piemērs), tad iegūst: x = 1183552/52 = 22761 €. Kļūda ir 316 € (23077–22761) jeb apmēram 1 % (316/22761*100 %). Parasti ieguvums no aprēķinu vienkāršošanas šāda lieluma kļūdu attaisno. Ja ir zināmi gradācijas klašu vidējie, tad par intervālus pārstāvošajiem lielumiem izmanto tos, kļūdas šajā gadījuma praktiski nav. Ja gradācijas klasēm vidējie nav zināmi, tad par klases pārstāvošajiem lielumiem izmanto intervālu centrus.*

3.1.2. Aritmētiskā vidējā īpašības

Aritmētiskā vidējā īpašībām[1] ir gan teorētiska, gan praktiska nozīme. Daudzas īpašības savu praktisko nozīmi mūsdienās sakarā ar vispārējo datorizāciju ir zaudējušas, jo ir orientētas uz aprēķinu vienkāršošanu, ja aprēķinus veic ar vienkāršu skaitļošanas tehniku.

1. Noviržu summa no aritmētiskā vidējā ir nulle. Parasti, aprēķinot lokācijas rādītājus, bez aritmētiskā vidējā tiek rēķināta arī standartnovirze (tā tiks apskatīta vēlāk). Standartnovirzi aprēķinot manuāli, kā starprezultāts tiek iegūtas novirzes . Ja šo noviržu summa nav nulle vai skaitlis ļoti tuvs nullei (ja starprezultāti tiek apaļoti), tad aritmētiskais vidējais ir aprēķināts nepareizi un tas ir jāpārrēķina. Šī īpašība kalpo aritmētiskā vidējā aprēķina pareizības pārbaudei.
2. Noviržu kvadrātu summa no aritmētiskā vidējā ir minimāla (vismazāko kvadrātu īpašība): ja . Šai īpašībai ir teorētiska nozīme, daudzas statistikas formulas balstās uz šo īpašību.
3. Ja no visiem lielumiem, kuru vidējo aprēķina, atņem vienu un to pašu konstantu lielumu un starpībām aprēķina aritmētisko vidējo, tad tas ir par to pašu konstanto lielumu c mazāks nekā sākotnējo datu aritmētiskais vidējais. Šī īpašība ļoti atviegloja aprēķinus bez kalkulatora. Iespēju samazināt sākotnējos datus par konstantu lielumu plaši izmantoja arī standartnovirzes un citos aprēķinos. Lai izprastu šīs īpašības būtību, var minēt nelielu piemēru. Uzņēmumā ir trīs strādnieki, kuri gatavo izstrādājumus *A*. Pirmais strādnieks izgatavo 21, otrais – 25, bet trešais strādnieks – 23 izstrādājumus. Aritmētiskais vidējais būs

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 23. lpp.

(21+25+23)/3 = 23 izstrādājumi. Izmantojot apskatāmo aritmētiskā vidējā īpašību, no katra novērojuma atņem konstanti (šajā gadījumā piemērots lielums ir 20) un vidējo aprēķina tikai starpībām: (1+5+3)/3+20 = 23 izstrādājumi. Starpības var būt arī negatīvas, mīnus zīme ir jāņem vērā aprēķinos.

4. Ja visus nesagrupētos datus (variantus vai intervālu centrus) samazina reizes (izdala ar Δ) un dalījumiem aprēķina aritmētisko vidējo, samazināto lielumu vidējais ir jāreizina ar Δ. Svarīga īpašība mūsdienās un ne tikai statistikas, bet arī citos ekonomiskajos aprēķinos. Ekonomikā bieži vien ir jāoperē ar skaitļiem tūkstošos vai miljonos, kas reizēm ir arī noapaļoti ar iepriekšminēto (līdz tūkstošiem vai miljoniem) precizitāti. Katram skaitlim ievadot trīs vai pat sešas nulles, ievērojami pieaug kļūdīšanās iespēja, aprēķins ir gausāks. Standartnovirzi rēķinot, sākotnējie dati ir jākāpina kvadrātā, ja skaitļi ir lieli, tad kāpinātais skaitlis vairs nav attēlojams parasta kalkulatora displejā. Lielums var būt jebkurš skaitlis, bet parasti samazina par veselām kārtām (10, 100, 1000 utt. reizes). Visbiežāk samazina tūkstoš vai miljons reižu, jo tad iegūtais rezultāts ir interpretējams (lasāms) bez pārveidojumiem (15000 € ir 15 tūkstoši €, skaitlis, ko izmantos aprēķinos, būs 15, bet mērvienība tam būs tūkstoši €). Ja samazināšanu veic 100 vai 10000 reizes, aprēķinus var veikt, bet, lai iegūto rezultātu nosauktu, ir jāpāriet uz parastām mērvienībām, ir jāreizina iegūtais rezultāts ar sākotnējo datu dalīšanai izmantoto skaitli.

5. Ja visu datu (variantu) statistiskie svari ir vienādi, svērtais aritmētiskais vidējais ir vienāds ar vienkāršo vidējo. Apskatot svērto aritmētisko vidējo, tika uzsvērts, ka relatīvajiem rādītājiem (cnt/ha, €/v, % u. c.), rēķinot vidējo, nedrīkst lietot vienkāršo aritmētisko vidējo. Šo īpašību var izmantot, lai vienkāršotu aritmētiskā vidējā aprēķinus relatīvajiem rādītājiem. Piemēram, agronomiskajos pētījumos pārbauda 4 šķirņu ražību, ar katru šķirni apsēti 5 ha. Ražība:
   - 2. šķirne – 28 cnt/ha;
   - 3. šķirne – 26 cnt/ha;
   - 4. šķirne – 20 cnt/ha.
   - 1. šķirne – 24 cnt/ha;

   Kopējo izmēģinājuma lauka vidējo ražību var aprēķināt kā vienkāršo aritmētisko vidējo: (24+28+26+20)/4 = (24*5+28*5+26*5+20*5)/ (5+5+5+5) =24,5 cnt/ha.

6. Ja visu svaru sistēmu reizina vai dala ar vienu un to pašu skaitli, svērtais aritmētiskais vidējais nemainās. Ja svaru summa ir vienāda ar 1, aritmētisko vidējo var aprēķināt pēc formulas , kur ν – relatīvie biežumi. Praktiska nozīme šai īpašībai ir tad, ja ir aprēķināti relatīvie biežumi viena daļās. Teorētiski un praktiski šī ideja tiek izmantota varbūtību aprēķinos (tie tiks apskatīti tālāk).

7. Ja statistikā kopa ir sadalīta *k* grupās un ir izrēķināti grupu aritmētiskie vidējie, tad visas kopas svērto aritmētisko vidējo var aprēķināt kā grupu vidējo svērto aritmētisko vidējo.

Veicot parasto grupēšanu, velkot svītriņas, vai decimālajā punktu – svītriņu metodē, tiek zaudēta iekšgrupu precizitāte. Aprēķinot aritmētisko vidējo no grupētiem datiem, par grupu pārstāvošu lielumu lieto intervāla centru, taču rezultāts var nedaudz atšķirties, ja vispirms saskaita grupā reģistrētos novērojumus un tad aprēķina grupas vidējo.

Ir vēl viena grupēšanas metode, kas ļauj grupēt datus, nezaudējot precizitāti. Metode saucas **"stumbrs – lapas"**.[1] Lai lietotu metodi, ir jābūt

[1] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 15.

diskrētam sadalījumam ar variantu skaitu līdz 100 vai nedaudz vairāk. Ja variantu ir daudz vairāk, tad sanāk ļoti sīks grupējums. Šāda grupēšanas metode var uzlabot precizitāti arī nepārtrauktam sadalījumam, bet tad novērojumu skaitļi ir jānoapaļo.

Grupēšanas metode tiks apskatīta 3.1. piemērā.

3.1. piemērs. Studentu zināšanas eksāmenā tiek vērtētas, piešķirot noteiktu punktu skaitu par katru atbildi vai uzdevuma risinājumu. Iespējamā vērtējuma amplitūda ir no 0 līdz 120 punktiem. Tālāk ir dots datu masīvs par studentu grupas uzrādītajiem rezultātiem.

89	64	58	60	57	114	69	68	53	83
94	67	63	69	69	102	29	58	68	28
72	74	63	67	105	117	85	48	61	52
70	9	18	83	81	85	79	69	65	78
89	72	71	57	101	61	41	87	79	34
18	46	114	82	70	38	76	72	85	68
16	65	34	43	83	87	93	74	76	52
87	73	85	60	72	113	60	64	89	
45	50	72	39	83	30	40	76	47	
48	25	108	26	85	79	62	69	85	

“Stumbrs” ir desmiti, bet lapas ir vieni. “Stumbrs” veido grupas, bet novērojumus atzīmē ar ciparu (pēdējais cipars skaitlī) atbilstošajā grupā. 3.4.tabulā ir parādīts grupējums un aprēķināti grupu vidējie.

3.4. tabula

Eksāmena rezultātu grupējums “Stumbrs – lapas”

“Stumbrs”	*“Lapas”*	*Novērojumu skaits, fi*	*Ciparu summa, Σ(xij-c)*	*Grupas vidējie,* $\overline{x_i}$
0	*9*	*1*	*9*	*9*
1	*868*	*3*	*22*	*17,33*
2	*5698*	*4*	*28*	*27*
3	*49804*	*5*	*25*	*35*
4	*58631087*	*8*	*38*	*44,75*
5	*08778322*	*8*	*37*	*54,63*
6	*4753309709190289498158*	*22*	*111*	*65,05*
7	*204231202996246968*	*18*	*75*	*74,17*
8	*997532133557575953*	*18*	*93*	*85,17*
9	*43*	*2*	*7*	*93,5*
10	*8512*	*4*	*16*	*104*
11	*4473*	*4*	*18*	*114,5*
Kopā			*97*	

Novērojumu reģistrāciju pa grupējuma klasēm veic līdzīgi, kā ar svītriņu vai punktu – svītru kombinācijas metodi, tikai svītriņu vai punktu vietā gradācijas klasē atzīmē pēdējo ciparu. Pirmais novērojums ir 89 punkti, ieraksta 8.klasē ciparu 9, otrais novērojums ir 94 – 9.klasē ieraksta ciparu 4. Tā turpina, kamēr tiek reģistrēti visi novērojumi. Atšķirībā no "klasiskās" grupēšanas, kas tika apskatīta iepriekš, šai metodei klase sākas ar nulli un beidzas ar 9, piemēram, 7.klasē mazākais ieskaitītais skaitlis ir 60, bet lielākais – 69. Gradācijas klašu vidējos aprēķina, izmantojot aritmētiskā vidējā 3.īpašību (aprakstu skatīt iepriekš). Tā 2.gradācijas klasē ir reģistrēti trīs novērojumi ar vērtībām – 18, 16 un 18. Nemainīgā daļa – konstante (stumbrs) šajā klasē ir 10, lai aprēķinātu vidējo, saskaita pēdējo ciparu summu, izdala ar novērojumu skaitu un pieskaita konstanti (stumbru): (8+6+8)/3+10 =17,33. Analoģiski rīkojas arī pārējās gradācijas klasēs, trešajā klasē pieskaitāmā "stumbra" daļa ir 20 utt.

Excel aritmētisko vidējo rēķina negrupētiem datiem ar funkciju ***AVG – average***. Ja aritmētisko vidējo jārēķina no grupētiem datiem, tad tas jādara, ievadot formulas.

3.1.3. Harmoniskais vidējais

Harmonisko vidējo iegūst, pakāpju vidējā formulā ņemot $z = -1$. Tā kā $x^{-1} = \frac{1}{x}$, tad harmoniskais vidējais ir:

$$\bar{x}_{-1} = \sqrt[-1]{\frac{\sum x_i^{-1}}{n}} = \frac{n}{\sum \frac{1}{x_i}} \tag{3.5.}$$

svērtais harmoniskais vidējais:

$$\bar{x}_{-1} = \frac{\sum f_i}{\sum \frac{f_i}{x_i}} \tag{3.6.}$$

Harmonisko vidējo lieto tad, ja jāaprēķina apgrieztu relatīvo lielumu vidējie[1], piemēram, darba ražīguma, kapitālietilpības, kapitāla aprites vidējie rādītāji u.tml. Aprēķinu būtību ir atklāta nākamajā piemērā.

Detaļas gatavo divi strādnieki. viens detaļas izgatavošanai patērē 4 minūtes, bet otrs – 6 minūtes. Aprēķinot aritmētisko vidējo: (4+6)/2 = 5, tiek iegūts nepareizs rezultāts, jo ražīgākais strādnieks laika vienībā saražos vairāk detaļas nekā otrs, un vidējais laika patēriņš detaļas gatavošanai būs tuvāk 4 minūtēm nekā 6 minūtēm. Līdz pareizam rezultātam var nonākt, veicot aprēķinus no otra puses – izvēlas kādu laika momentu, aprēķina kopā saražoto vienību skaitu,

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 25. lpp.

kopā nostrādāto laiku izdala ar kopā saražoto izstrādājumu skaitu. Piemērā tiek pieņemts, ka strādnieki strādā vienu stundu. Pirmais strādnieks stundas laikā izgatavo 15 detaļas (60 minūtes/4 minūtēm/detaļa), bet otrs – 10 detaļas (60/6). Kopā strādnieki ir saražojuši 25 detaļas un iztērējuši 120 minūtes. Vidējais laika patēriņš detaļas izgatavošanai būs: 120 min./25 det. = 4,8 minūtes/detaļa jeb 4 minūtes 48 sekundes.

Šo pašu rezultātu var iegūt tieši, izmantojot harmonisko vidējo:

$$\overline{x_{-1}} = \frac{2}{\frac{1}{4}+\frac{1}{6}} = \frac{2}{0{,}42} = 4{,}8\,min$$

Lielākus aprēķinus veic tabulas formā. Uzskatāmībai var aprēķināt vidējo darba patēriņu detaļas izgatavošanai 3.2. piemēram.

3.2. piemērs. Ir savākti šādi dati par uzņēmuma strādnieku darba ražīgumu:

- 2 min./det. – 4 strādnieki;
- 3 min./det. – 9 strādnieki;
- 4 min./det. – 10 strādnieki;
- 5 min./det. – 3 strādnieki.

3.5. tabula

Vidējā svērtā harmoniskā aprēķināšana darba ražīguma piemēram

Detaļas izgatavošanai patērētais laiks, min./det. xi	*Strādnieku skaits, fi*	
2	*4*	*2*
3	*9*	*3*
4	*10*	*2,5*
5	*3*	*0,6*
Kopā		

$$\overline{x}_{-1} = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}$$ *= 26/8,1 =3,21 min./det.*

Aprēķinu pareizību var pārbaudīt līdzīgi, kā tas tika darīts iepriekšējā piemērā. Rēķinot vidējo darba patēriņu ar harmonisko vidējo, aprēķins ir ātrāks.

Katram no vidējo veidiem ir konkrēts pielietojums. Var jebkuriem skaitļiem aprēķināt katru no vidējiem, bet tikai vienam no tiem būs ekonomiska jēga un iegūtos rezultātus varēs izskaidrot. Tādēļ pirms vidējā aprēķināšanas ir jāizvērtē aprēķināmā rādītāja būtība un jāizvēlas piemērots vidējā veids.

Ar *Excel* harmonisko vidējo var aprēķināt no negrupētiem datiem ar funkciju *Harmean*. Ja dati ir sagrupēti, tad harmonisko vidējo vieglāk ir aprēķināt ar kalkulatoru, vai ievadot vajadzīgās formulas ar roku *Excel* darba lapā.

3.1.4. Ģeometriskais un kvadrātiskais vidējais

Ģeometrisko vidējo izskaitļo, lietojot logaritmus:

$$lg\bar{x}_0 = \frac{\sum lg\, x}{n};\ \overline{x_0} = 10^{lg\bar{x}_0} \tag{3.7.}$$

Ja ir jāaprēķina svērtais ģeometriskais vidējais, lieto formulu

$$lg\bar{x}_0 = \frac{\sum(lgx)\cdot f}{f} = \frac{f_1\cdot lgx_1 + f_2\cdot lgx_2 + \cdots + f_n\cdot lgx_n}{f_1 + f_2 + \cdots + f_n} \tag{3.8.}$$

Ģeometrisko vidējo galvenokārt lieto **dinamikas rindu** apstrādē. Ar to aprēķina vidējos augšanas tempus.[1]

Vidējo augšanas tempu no negrupētiem datiem var aprēķināt, izmantojot *Excel* funkciju *GEOMEAN*. Plašāk šis jautājums tiek skatīts kursā „Statistika II”.

Kvadrātisko vidējo iegūst, pakāpju vidējā formulā ņemot z=2. Tad:

$$\overline{x_2} = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} \tag{3.9.}$$

Svērto kvadrātisko vidējo aprēķina pēc formulas:

$$\overline{x_2} = \sqrt{\frac{\sum x^2 f}{\sum f}} = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \cdots + x_n^2 f_n}{f_1 + f_2 + \cdots + f_n}} \tag{3.10.}$$

Kvadrātisko vidējo plašāk lieto divos gadījumos:

1) ir jārēķina vidējais datiem, kuri ir izteikti novirzēs no kāda cita vidējā. Aprēķinot vidējo kvadrātisko novirzi, īstenībā tiek raksturota ne pazīmes centrālā tendence, bet pazīmes variācija. Vidējā kvadrātiskā novirze vairāk tiks analizēta pie variācijas rādītājiem;

2) ir jāaprēķina kādu lineāru mērījumu (kvadrātu malu, apļu diametru utt.) vidējais lielums, turklāt zinātniski praktiska interese ir ne tik daudz par pašiem šiem mērījumiem, bet par figūru laukumiem, ko šie

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 29. lpp.

mērījumi raksturo. Uz šīs idejas pamata ir izstrādāta mežsaimniecības metodika koksnes krājas (daudzuma) noteikšanai mežā.[1]

3.1.5. Struktūras vidējie un citi rādītāji

Par centrālās tendences rādītājiem pareizi saukt ir tikai modu un mediānu, bet kvantiles ir variācijas rādītāji.

Par **modu** (*Mo*) sauc sadalījuma rindā visbiežāk sastopamās variantes vērtību. Diskrētā vai atributīvā variācijas rindā moda nolasāma tieši kā variants ar vislielāko absolūto vai relatīvo biežumu.[2,3,4,5] 2.2. piemērā 1. studentu grupā *Mo*=6, bet 2. studentu grupā *Mo*=7. Atributīva sadalījuma gadījumā moda ir vienīgais centrālās tendences rādītājs, piemēram, nav iespējams aprēķināt studentu grupā vidējo tautību, bet var noteikt modālo (vislabāk pārstāvēto) tautību. Arī diskrēta kvantitatīva sadalījuma gadījumā moda var būt labāks centrālās tendences rādītājs nekā aritmētiskais vidējais, piemēram, veikala vadītāju vairāk interesēs modālais kurpju izmērs nekā vidējais kurpju izmērs.

Intervālu rindā vispirms jānosaka **modas** jeb **modālais intervāls**. Tas ir intervāls ar vislielāko biežumu.

Intervālu variācijas rindai modu aprēķina pēc **interpolācijas** (nezināmu datu meklēšana zināmā datu apgabalā) **formulas**:

$$M_O = x_O + \Delta \frac{f_{M_O} - f_{M_{O-1}}}{\left(f_{M_O} - f_{M_{O-1}}\right) + \left(f_{M_O} - f_{M_{O+1}}\right)} \quad (3.11.)$$

kur M_O – moda;

x_O – modālā intervāla apakšējā robeža;

Δ – modālā intervāla garums;

f_{M_O} – modālā intervāla biežums;

$f_{M_{O-1}}$ – pirmsmodālā intervāla biežums;

$f_{M_{O+1}}$ – pēcmodālā intervāla biežums.

Parasti analīzē aprobežojas ar modālā intervāla noteikšanu.

Programmā *Excel* modu negrupētiem datiem atrod ar funkciju *MODE*. Ja vairākām vērtībām ir vienāds varianšu skaits, tad nav atbildes. Nepārtrauktam vai diskrētam sadalījumam ar daudz

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 30. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis*. Rīga: Datorzinību Centrs. 39. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 75. lpp
[4] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 73.
[5] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 33. lpp.

variantiem modas noteikšanai ar *Excel* nav jēgas, jo tās atrašana ir atkarīga no nejaušām sakritībām un šādi noteikta moda neraksturo centrālo tendenci. Piemēram, tiek analizēti studentu grupas fiziskie rādītāji – augums, svars. Nosakot svaru (nepārtraukts sadalījums) visi mērījumi kaut nedaudz, bet ir atšķirīgi, izņemot divus, kas ar doto precizitāti izrādās identiski. Neskatoties uz to, ka studentu svars pārsvarā ir 50-60 kg, *Excel* ir atrasta moda, kas ir 102,4 kg. Šāda moda neraksturo šī sadalījuma centrālo tendenci.

Par **mediānu** sauc augošā vai dilstošā kārtībā sakārtotas variācijas rindas vidū esošo variantes vērtību.[1,2,3] Piemēram, jāatrod mediāna studentu eksāmena atzīmei. Ir šādi eksāmena rezultāti: 5; 7; 7; 9; 6; 4; 5; 8; 6 – tā ir nesakārtota empīriskā variācijas rinda. Lai atrastu mediānu, vispirms ir jāsakārto (jāsaranžē) visi novērojumi. Tie paši rezultāti, saranžēti augošā secībā, būs:

4; 5; 5; 6; 6; 7; 7; 8; 9.

5. novērojuma vērtība
no 9 būs mediāna

Ja mediāna jāatrod no variācijas rindas, kurā ir pāra novērojumu skaits, tad variācijas rindas vidu pārstāv divi novērojumi, mediāna ir šo divu novērojumu aritmētiskais vidējais. Ja iepriekš apskatīto studentu grupu papildina vēl viens students, kurš eksāmenu nokārto uz 7 ballēm, tad ranžēta variācijas rinda būs šāda:

4; 5; 5; 6; 6; 7; 7; 7; 8; 9.

Mediāna ir 5. un 6.
variantes aritmētiskais
vidējais: (6+7)/2 = 6,5

Ja sadalījums neatbilst normālajam sadalījumam (teorētiskajā normālajā sadalījumā vidējās vērtības ir sastopamas biežāk nekā lielās vai mazās vērtības, par teorētiskajiem sadalījumiem plašāk būs tālāk), tad mediāna ir piemērotāks centrālās tendences rādītājs par aritmētisko vidējo.

Pētot studentu grupas nodrošinātību ar kabatas naudu, iegūti šādi rezultāti (studenta kabatas nauda vidēji vienā dienā): 0,5; 1; 1; 1,5; 1,5; 2; 2,5; 3; 50 €/dienā.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 31. lpp.
[2] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 35.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 40.lpp.

Aritmētiskais vidējais kabatas naudas apjoms ir 7 €/dienā, bet to nevar nosaukt par centrālo tendenci, jo lielo vidējo kabatas naudas lielumu nosaka viena studenta ekstrēmi lielā kabatas nauda, visiem pārējiem kabatas nauda dienā ir mazāka par trim eiro. Mediāna ir 1,5 €/dienā, un tā labāk raksturo studentu materiālā stāvokļa centrālo tendenci nekā aritmētiskā vidējā kabatas nauda – 7 €/dienā. Reizēm statistiskajos aprēķinos neiekļauj no vienas līdz trim ekstrēmajām (mazākajām un lielākajām) variantēm no abiem galiem. Reizēm šādas ekstrēmi lielas vai mazas variantes var būt reģistrācijas vai novērošanas kļūdas.

Mediānas intervāls ir tas, kurā uzkrātie absolūtie biežumi pirmo reizi pārsniedz pusi no kopas vienību skaita vai kurā uzkrātie relatīvie biežumi pirmo reizi pārsniedz 50 %.

Mediānas interpolācijas formula:

$$M_e = x_0 + \Delta \frac{\frac{\sum f}{2} - S_{M_{e-1}}}{f_{M_e}} \qquad (3.12.)$$

kur x_0 – mediānas intervāla apakšējā robeža;
Δ – mediānas intervāla garums;
$S_{M_{e-1}}$ – uzkrātais biežums intervālā, kas atrodas pirms mediānas intervāla;
f_{M_e} – mediānas intervāla biežums;
$\sum f$ – kopējais novērojumu skaits variācijas rindā.

3.6. tabula

Mediālā un modālā intervālu noteikšana (2.1. piemēra dati)

Apgrozījums, tūkstošos eiro	*Absolūtie biežumi*	*Kumulatīvie biežumi*
6 – 14	*9*	*9*
14 – 22	*22*	*22 + 9 = 31*
22 – 30	*12*	*12 + 31 = 43*
30 – 38	*2*	*45*
38 – 46	*2*	*47*
46 - 54	*5*	*52*
Kopā	*52*	

Modālo intervālu nosaka lielākais reģistrēto novērojumu skaits vienā gradācijas klasē

Mediālais intervāls ir tas, kur kumulatīvie biežumi pārsniedz pusi no kopējo novērojumu skaita (52/2 =26)

Izmantojot modas interpolācijas formulu (3.11.), aprēķina modu:

$$M_O = 14 + 8\frac{22-9}{(22-9)+(22-12)} = 18{,}52 \text{ tūkstoši €},$$

kā arī mediānu (3.12.):

$$M_e = 14 + 8\frac{52/2 - 9}{22} = 20{,}18 \text{ tūkstoši €}$$

Programmā Excel, pielietojot funkciju MODE 2.1. piemēram, nav atbildes, bet MEDIAN ir 19923,5 €, rezultāti iegūti no negrupētiem datiem. Kā redzams, iepriekšējā aprēķinā nav būtiskas kļūdas salīdzinājumā ar aprēķiniem no negrupētiem datiem.

Ja grupēšanai izmanto metodi “stumbrs – lapas”, tad mediānas vērtību var atrast precīzi. Informācija par eksāmena punktu piemēra skaidrojumu ir dota 3.7. tabulā.

3.7. tabula

Kumulatīvo biežumu aprēķins un mediālā intervāla noteikšana eksāmena punktu 3.1. piemēram

“Stumbrs”	*“Lapas”*	*Novērojumu skaits, fi*	*Kumulatīvie biežumi*
0	*9*	*1*	*1*
1	*868*	*3*	*4*
2	*5698*	*4*	*8*
3	*49804*	*5*	*13*
4	*58631087*	*8*	*21*
5	*08778322*	*8*	*29*
6	*4753309709190289498158*	*22*	*51*
7	*204231202996246968*	*18*	*69*
8	*997532133557575953*	*18*	*87*
9	*43*	*2*	*89*
10	*8512*	*4*	*93*
11	*4473*	*4*	*97*

Kopējais novērojumu skaits ir 97, ranžētas variācijas rindas variante, kas dala rindu uz pusēm, būs 49. variante – mazākas un lielākas par to būs pa 48 variantēm. 49. variante ir 7. gradācijas klasē. Atliek tikai saranžēt 7. gradācijas klases variantes un atrast 49. variantes vērtību. 6. gradācijas klasē lielākajai variantei ir 29. numurs, bet mazākā vērtība 7. gradācijas klasē ir ar numuru 30. Tālāk 6. gradācijas klases ciparus pārraksta augošā secībā – tur ir trīs nulles, divi vieninieki utt.

0	0	0	1	1	2	3	3	4	4	5	5	7	7	8	8	8	9	9	9	9	9
30.	31.	32.	33.	34.	35.	36.	37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.	49.	50.	51.

Mediānas vērtība ir 69 (“stumbrs ir 6 desmiti, bet lapas vērtība ir 9).

Sadalījuma struktūras raksturošanai var izmatot kvantiļu.[1,2,3] (kvartiles, kvintiles, deciles, procentiles) grupējumus. Grupēšanas princips kvantiļu grupējumam atšķiras no iepriekš apskatītā intervālu grupējuma. Intervālu grupējumā tika fiksētas pazīmes robežvērtības, bet sadalījums tika raksturots ar to, cik novērojumu ir katrā intervālā. Kvantiļu grupējumā sadala visu novērojumu skaitu vienādās daļās, bet kvantiles raksturo ar robežvarianšu vērtībām.

Kvartiles ir pētītās pazīmes vērtības, kas sakārtotu variācijas rindu dala četrās līdzīgās daļās tā, ka katrā no tām nonāk 25 % kopas vienību. Ir trīs kvartiles: pirmā Q_1, otrā Q_2 un trešā Q_3. Otrā kvartile ir mediāna.

Sakārtotu variācijas rindu desmit vienādās daļās dala **deciles**, kuras aprēķina līdzīgi kvartilēm.

Variācijas rindu piecās līdzīgās daļās dala **kvintiles**, bet 100 vienādās daļās – **procentiles**.

Kvartiles, kvintiles, deciles u. c. šādus rādītājus apzīmē ar kopējo terminu – **kvantiles**.

Kvantiļu grupējumu plaši lieto, raksturojot iedzīvotāju ienākumu sadali, piemēram, "10 % nabadzīgāko valsts iedzīvotāju iztiek ar naudas summu uz vienu mājsaimniecības locekli, kas ir mazāks nekā … € (1. decile)" u.tml.

3.2. Variācijas rādītāji

Statistika pēta tikai to, kas variē, līdz ar to ir nepieciešams kāds rādītājs, kas raksturotu datu izkliedi. Par variāciju var spriest pēc grupējuma rezultātiem tabulā vai histogrammā. Taču ir nepieciešams universāls rādītājs, kas koncentrētā veidā informētu par datu izkliedi.

Variācijas raksturošanai var lietot vairākus rādītājus. Viena no tiem – variācijas amplitūda – jau tika lietota, aprēķinot intervālu garumu.

Par **variācijas amplitūdu** sauc starpību starp pazīmes lielāko un mazāko novēroto skaitlisko vērtību:[4,5,6]

---

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel*: mācību līdzeklis. Rīga: Datorzinību Centrs. 42. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 32. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 79.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 35. lpp.
[5] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 82. lpp.
[6] Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 60. lpp.

$$I = x_{max} - x_{min}, \quad (3.13.)$$

kur x_{max} – pazīmes lielākā;

x_{min} – mazākā vērtība.

Variācijas amplitūda nav labs izkliedes rādītājs, jo ir atkarīga tikai no ekstrēmajām (lielākās un mazākās) vērtībām. Kā novērojumā gadās kāda krasi atšķirīga variante, tā strauji palielinās variācijas amplitūda. Tādēļ variācijas amplitūdu kā datu izkliedes pamatrādītāju nelieto.

Noviržu summa no aritmētiskā vidējā ir nulle (aritmētiskā vidējā īpašība), tādēļ arī šo rādītāju nevar izmantot datu izkliedes raksturošanai. Ja novirzēm no aritmētiskā vidējā ignorē zīmi (vai konkrētais novērojums ir bijis lielāks, vai mazāks nekā aritmētiskais vidējais), tad noviržu moduļu summu var izmantot datu izkliedes raksturošanai. **Vidējā absolūtā jeb vidējā lineārā novirze** ir aritmētiskais vidējais no atsevišķu novērojumu (varianšu) absolūtajām novirzēm no aritmētiskā vidējā.[1,2,3]

Negrupētiem datiem:

$$\alpha = \frac{\sum |x_i - \bar{x}|}{n}, \quad (3.14.)$$

grupētiem datiem:

$$\alpha = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f}, \quad (3.15.)$$

kur α – vidējā lineārā novirze;

n – kopas vienību skaits;

f_i – statistiskie svari.

Vidējo lineāro novirzi lieto samērā reti (novērtējot prognožu kļūdas). Praksē pieņemts datu izkliedi raksturot ar vidējo kvadrātisko novirzi (parasti to sauc par standartnovirzi). Novirzes no aritmētiskā vidējā kāpinot kvadrātā, zūd to zīme (gan pozitīva, gan negatīva skaitļa kvadrāts ir pozitīvs). Atšķirībā no vidējās lineārās novirzes vidējā kvadrātiskā novirze akcentē lielās novirzes (kāpinot kvadrātā lielāku skaitli, tas kļūst relatīvi lielāks, nekā kāpinot kvadrātā mazāku skaitli). Standartnovirzes lietošana arī ir labāk teorētiski pamatota.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 35. lpp.

[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 70. lpp.

[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība. 83. lpp.

Datu izkliedes raksturošanai izmanto rādītāju, ko sauc par **dispersiju**. To apzīmē ar s^2, ja to aprēķina pēc izlases datiem, vai ar σ^2, ja tā ir ģenerālkopas dispersija vai dispersija teorētiskos sadalījumos.[1,2,3]

Ja novērojumu statistiskie svari ir vienādi vai arī tos var neņemt vērā (negrupēti dati), dispersiju aprēķina pēc formulas:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n} \qquad (3.16.)$$

vai svērtās dispersijas formulas:

$$s^2 = \frac{\sum(x_i - \bar{x})^2 f_i}{\sum f_i} \qquad (3.17.)$$

Samērā bieži kā starprezultātu fiksē noviržu kvadrātu summu, kas ir dispersijas formulas skaitītājs un kuru izmanto citos aprēķinos. To apzīmē ar Q.

$$Q = \sum(x_i - \bar{x})^2 \qquad (3.18.)$$

vai grupētiem datiem:

$$Q = \sum(x_i - \bar{x})^2 f_i \qquad (3.19.)$$

Dispersija ir grūti interpretējams rādītājs, tās mērvienība ir sākotnējo datu mērvienības kvadrāts. Ja analizē tirdzniecības aģentu apgrozījumu, tad dispersijas mērvienība ir kvadrāteiro. Izvelkot kvadrātsakni no dispersijas, iegūst **standartnovirzi (vidējo kvadrātisko novirzi),** un to apzīmē ar s vai σ.[4,5,6] Dispersijas formulu, līdzīgi kā citas formulas, kuras satur summas zīmi, rēķina tabulas veidā.

Tālāk tiks aprēķināta vidējā lineārā novirze, dispersija un standartnovirze 2.1. piemēram.

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 91.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja.36. lpp.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 44. lpp.
[4] Turpat, 46. lpp.
[5] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 92.
[6] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 37. lpp.

3.8. tabula

Aritmētiskā vidējā, vidējā lineārās novirzes un dispersijas aprēķins

Apgrozījums, tūkstošos eiro	*xi*	*fi*	*xi*fi*	*xi-x*	*(xi-x)*fi*	*(xi-x)2*fi*
6 – 14	*10*	*9*	*90*	*-13,077*	*-117,69*	*1539,05*
14 – 22	*18*	*22*	*396*	*-5,077*	*-111,69*	*567,05*
22 – 30	*26*	*12*	*312*	*2,923*	*35,08*	*102,53*
30 – 38	*34*	*2*	*68*	*10,923*	*21,85*	*238,63*
38 – 46	*42*	*2*	*84*	*18,923*	*37,85*	*716,17*
46 - 54	*50*	*5*	*250*	*26,923*	*134,62*	*3624,26*
Kopā		*52*	*1200*		*0* *458,769*	*6787,69*

Noviržu summai ir jābūt nullei. Ja tā nav, tad aritmētiskais vidējais ir aprēķināts nepareizi

$\Sigma|x_i - x|*f_i$, ailes kopsumma, neņemot vērā zīmes

$\sum (x_i - \bar{x})^2 f_i$

Vidējā lineārā novirze (3.14.) ir 458,769/52 = 8,822 tūkst. €.

Dispersija (3.16.) s^2 = 6787,69/52 = 130,533 un standartnovirze = 11,425 tūkst. €.

Standartnovirzi parasti min kopā ar aritmētisko vidējo un pieraksta šādi: $\bar{x} \pm s$ = 23,077±11,425 tūkst. € apgrozījums firmā A. Standartnovirze iegūst skaitlisku interpretāciju, salīdzinot to ar cita novērojuma standartnovirzi vai ar novērojuma vidējo aritmētisko. Ja citā novērojumā (firma B) ir iegūti šādi dati: $\bar{x} \pm s$ = 21,95±7,55 tūkst. €, tad var teikt, ka firmā B vidējais apgrozījums ir nedaudz zemāks nekā firmā A, bet tas ir stabilāks, nav tik lielu svārstību kā firmā A. Ja salīdzināmās kopas ir atšķirīgas mērogā, tad absolūtais lielums – standartnovirze –datu izkliedes salīdzināšanai neder. Ja vēlas veikt salīdzinājumu ar nekustamā īpašuma firmu (C), kuras vidējais apgrozījums un standartnovirze ir: $\bar{x} \pm s$ = 230,77±114,25 tūkst. €, tad absolūtos skaitļos, protams, datu svārstības ir 10 reizes lielākas nekā firmā A, bet relatīvi datu izkliede abām firmām ir vienāda.

Lai raksturotu relatīvo datu izkliedi, lieto variācijas koeficientu.

Variācijas koeficientu aprēķina, dalot standartnovirzi ar aritmētisko vidējo:[1,2]

$$V = \frac{S}{\bar{x}} \quad (3.20.)$$

V_A=11,425/23,077=0,495; V_B=7,55/21,95=0,344; V_C=114,25/230,77=0,495

Kā redzams pēc variācijas koeficienta, firmai B datu izkliede ir relatīvi mazāka, bet firmām A un C relatīvā datu izkliede ir vienāda.

Dispersijas īpašības[3,4]

1. Konstanta lieluma dispersija ir vienāda ar nulli. Būtībā neaktuāla īpašība, jo statistikā pēta tikai to, kas varie. Piemēram, muļķīgi ir pētīt augstskolas absolventu studiju parādus. Ja students ir beidzis augstskolu, tas nozīmē, ka viņam nav studiju parādu un visas variantes ir nulle, nav nekādas datu izkliedes.
2. Ja no visiem novērojumiem atņem kādu konstantu lielumu, tad dispersija nemainās.
3. Ja visus novērojumus reizina (dala) ar konstantu lielumu Δ, tad dispersija pieaug (samazinās) Δ^2 reizes. 2. un 3. īpašība ir aktuāla, ja aprēķinus veic manuāli, īpašības ir analoģiskas aritmētiskā vidējā īpašībām.
4. Ja, aprēķinot dispersiju, aritmētisko vidējo aizstāj ar kādu citu konstantu lielumu *c*, tad dispersija pieaug par lielumu $(\bar{x}-c)^2$. No tā izriet, ka dispersiju var aprēķināt pēc formulas.

   Ja c=0, tad iegūst formulu:

$$s^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 \quad (3.21.)$$

   Ja izmanto statistiskos svarus, analoga formula ir šāda:

$$s^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - (\bar{x})^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{\sum f_i}\right)^2 \quad (3.22.)$$

Šīs formulas ir aktuālas manuālajiem aprēķiniem, tās ievērojami samazina nepieciešamo kalkulāciju apjomu. Pēdējās divas formulas sauc par momentu formulām, sīkāks skaidrojums būs tālāk.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 35. lpp.
[2] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 86. lpp
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 39. lpp.
[4] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 72. lpp.

3.9. tabula

Standartnovirzes aprēķina palīgtabula momentu 3.22.formulai

Apgrozījums, tūkstošos eiro	*xi*	*fi*	*xi*fi*	*xi2*fi*
6 – 14	*10*	*9*	*90*	*900*
14 – 22	*18*	*22*	*396*	*7128*
22 – 30	*26*	*12*	*312*	*8112*
30 – 38	*34*	*2*	*68*	*2312*
38 – 46	*42*	*2*	*84*	*3528*
46 - 54	*50*	*5*	*250*	*12500*
Kopā		*52*	*1200*	*34480*

$\Sigma x_i^2 * f_i$

Veicot kalkulāciju tabulās, jāatceras par iespēju izmantot kalkulatora atmiņu – veicot darbības pa rindiņām, iegūtos rezultātus iesummē atmiņā ar M+ taustiņu.

$s^2 = 34480/52 - (1200/52)^2 = 130{,}533.$

Kā redzams, dispersijas lielums ir tieši tāds pats kā, aprēķinot to ar 3.17. formulu. Kalkulācijas daudzums ir ievērojami mazāks – ar 3.17.formulu ir jāatņem (rēķina novirzes), iegūtās novirzes jākāpina kvadrātā, jāsareizina ar frekvencēm un jāsasummē, bet, rēķinot standartnovirzi ar 3.22. formulu, ir nepieciešams tikai aritmētiskā vidējā rēķināšanai iegūtos starprezultātus ($x_i * f_i$) sareizināt ar atbilstošajiem x_i un iegūtos rezultātus sasummēt. Summēšanai vienmēr jāizmanto kalkulatora atmiņa, tas ievērojami samazinās nepieciešamo kalkulāciju apjomu.

Bez centrālās tendences un variācijas rādītājiem ir vēl divu veidu rādītāji, kurus izmanto, lai raksturotu sadalījumu. Empīriskā sadalījuma simetriskumu raksturo asimetrijas rādītājs, bet sadalījuma smailumu ekscesa rādītājs.

3.3. Asimetrijas un ekscesa rādītāji

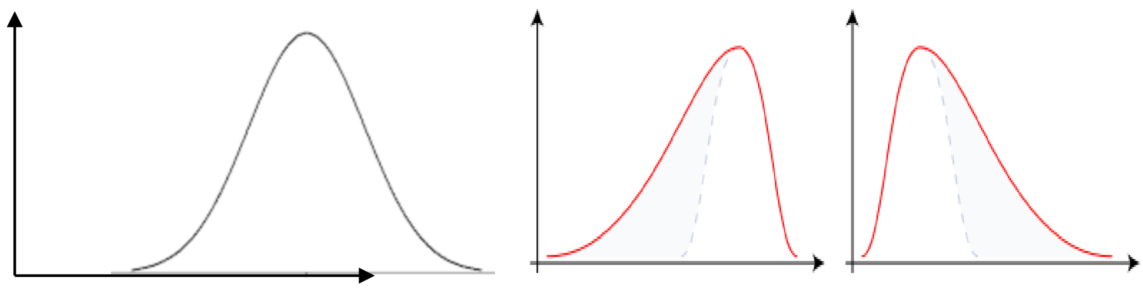
Lai saprastu šo rādītāju būtību, ir jāiedomājas sadalījuma histogramma. Ja, konstruējot histogrammu, palielina novērojumu skaitu un tajā pašā laikā samazina intervālu garumu, tiek iegūta nevis sadalījuma piramīda, bet gan nolaidena sadalījuma teorētiskā līkne. Zinot asimetrijas koeficienta vērtību, var iedomāties, kāda izskatās sadalījuma līkne. Asimetrijas koeficienta pamatformula (atklāj aprēķina būtību) ir standartizētais noviržu kubiskais vidējais:[1,2]

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 53. lpp.

[2] Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 75. lpp.

$$A = \frac{\sum(x_i - \bar{x})^3}{n * s^3} \tag{3.23.}$$

Kāpinot novirzes kubā, saglabājas novirzes zīme, līdz ar to summa var būt gan negatīva, gan pozitīva. Liela nozīme ir variantēm, kas būtiski atšķiras no centrālās tendences. Ja novērojumu nav daudz, tad asimetrijas koeficienta aprēķins var dot kļūdainu priekšstatu – papildinot novērojumus ar ekstrēmu varianti, ļoti mainās asimetrijas koeficienta vērtība. 3.1. attēlā ir parādītas sadalījuma līknes ar dažādām asimetrijas koeficienta vērtībām.



$\bar{x} = Mo = Me$, A=0 $\bar{x} \rangle Me \rangle Mo$, A<0 $x < Me < Mo$, A>0

3.1. attēls. **Sadalījuma līknes izskats pie dažādām asimetrijas koeficienta vērtībām**

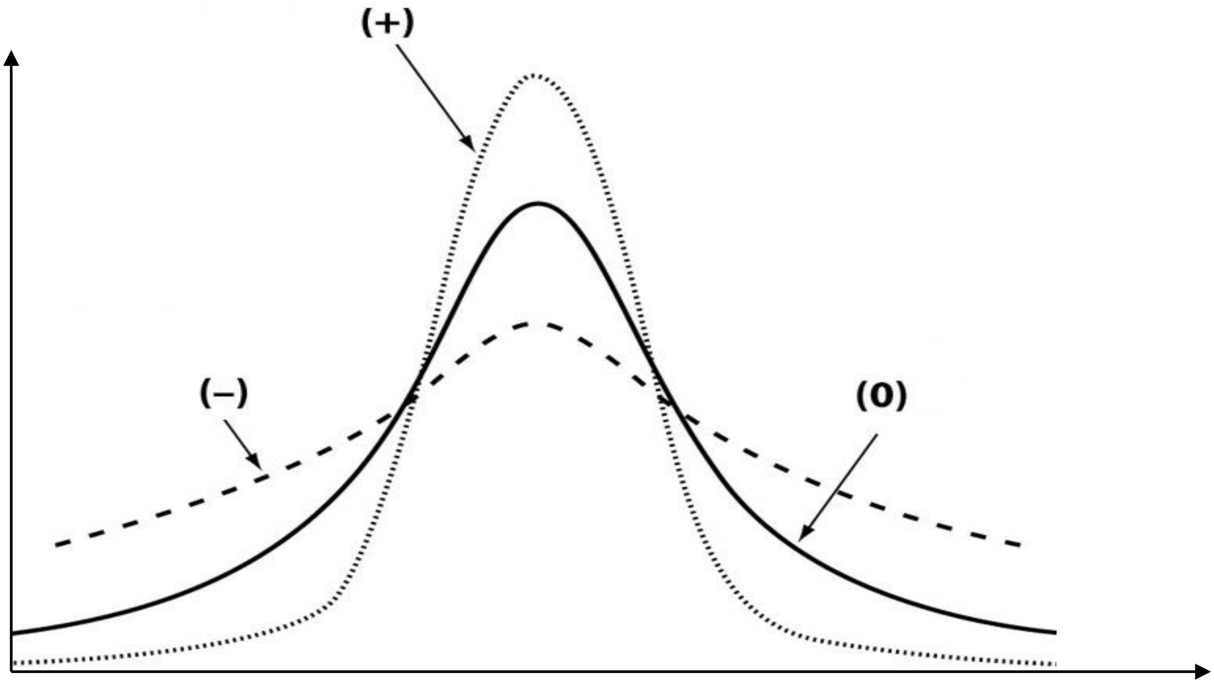
Ja novērojumu biežums abpus aritmētiskā vidējā ir simetrisks, tad asimetrijas koeficienta vērtība ir nulle. Ja ir "izstiepts" empīriskā sadalījuma labais zars, tad asimetrijas koeficients ir lielāks par nulli – sadalījumam ir labā jeb pozitīvā asimetrija. Ja salīdzina sadalījuma, kuram ir pozitīva asimetrija, centrālās tendences rādītājus, moda ir ar vismazāko vērtību, tad mediāna, bet aritmētiskais vidējais ir pats lielākais. Ja ir "izstiepts" kreisais sadalījuma zars, tad asimetrijas koeficients ir negatīvs un centrālas tendences rādītāju izvietojums ir pretējs – vismazākā vērtība ir aritmētiskajam vidējam, tad mediānai un vislielākā ir moda. Noteikt zīmi asimetrijas koeficientam pēc histogrammas vai arī iedomāties histogrammu, ja ir zināma asimetrijas koeficienta vērtība, var, nostājoties pretī (reāli vai domās) sadalījuma histogrammai un ar rokām to attēlojot. Ja, attēlojot sadalījuma līkni, ir jāceļ labā roka, tad ir pozitīvā asimetrija, savukārt kreisās rokas pacelšana liecina par negatīvo asimetriju.

Novēroto vērtību biežumu izvietojumu raksturo ekscesa koeficients.

Ekscesa koeficients ir standartizēts noviržu ceturtajā pakāpē summas vidējais mīnus trīs. Novirzes kāpinot ceturtajā pakāpē, vēl lielāka nozīme ir lielajām novirzēm. Kāpinot ceturtajā pakāpē gan pozitīvu, gan negatīvu skaitli, iegūst pozitīvu rezultātu. Lai šo rezultātu būtu ērtāk salīdzināt, atņem konstanti "3", tad teorētiskajam normālajam sadalījumam koeficients ir nulle.[1,2,3]

$$E = \frac{\sum(x_i - \bar{x})^4}{n * s^4} - 3 \tag{3.24.}$$

Ja ekscesa koeficientu vēlas vizualizēt, tad tas raksturo sadalījuma smailumu (3.2. attēls).



3.2.attēls. **Sadalījuma līknes izskats dažādām ekscesa koeficienta vērtībām**

Standarts, ar kuru salīdzina gan ekscesa, gan asimetrijas koeficientus, ir normālais teorētiskais sadalījums, kas tālāk tiks apskatīts plašāk. Ja novērotās vērtības visā datu apgabalā ir sadalītas samērā vienmērīgi, tad ekscesa koeficients ir negatīvs, ja tās ir tuvu centrālās tendences (aritmētiskā vidējā) vērtībām, tad ekscesa koeficients ir pozitīvs.

3.23. un 3.24. formula labi izskaidro rādītāju būtību, bet manuālajiem aprēķiniem tās nav ērtas, aprēķini ir ļoti darbietilpīgi. Lai atvieglotu aprēķinu veikšanu ar vienkāršu skaitļošanas tehniku, ir izstrādātas tā saucamās momentu formulas. Moments ir mehānikas jēdziens, kas raksturo mehāniskai sistēmai pieliktā

---

1 Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis*. Rīga: Datorzinību Centrs. 50. lpp.
2 Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 76. lpp.
3 Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 54. lpp.

spēka daudzumu, kas varētu iekustināt to rotācijas kustībai. Atkarībā no noteiktā sistēmas centra izšķir:[1]

- sākuma momentus (centrs ir koordinātu sistēmas sākumpunkts – nulle, apzīmē ar lielajiem burtiem *M*);
- nosacītie momenti (centrs - jebkurš brīvi izvēlēts punkts uz abscisu ass, apzīmē ar mazajiem burtiem *m*);
- centrālie momenti (centrs - aritmētiskais vidējais, apzīmē ar $\overline{m}$).

$$A = \frac{\overline{m}_3}{s^3} \quad (3.25.)$$

$$\overline{m}_3 = M_3 - 3M_2M_1 + 2M_1^3 \quad (3.26.)$$

$$E = \frac{\overline{m_4}}{s^4} - 3 \quad (3.27.)$$

$$\overline{m_4} = M_4 - 4M_3M_1 + 6M_2M_1^2 - 3M_1^4 \quad (3.28.)$$

$M_1 = \frac{\sum xf}{\sum f}$, aritmētiskais vidējais (3.4.)

$$M_2 = \frac{\sum x^2 f}{\sum f}, \quad (3.29.)$$

$$M_3 = \frac{\sum x^3 f}{\sum f}, \quad (3.30.)$$

$$M_4 = \frac{\sum x^4 f}{\sum f}, \quad (3.31.)$$

- $s = \sqrt{M_2 - M_1^2}$, standartnovirze, pēc būtības tā ir 3.22. formula.

Ja dati nav grupēti, tad sākuma momentu formulās skaitītājs nav jāreizina ar gradācijas klašu biežumiem, bet saucējā ir novērojumu skaits – *n*.

Ja kāpināšanas darbības veic ar kalkulatoru, tad jāizdara šādas darbības:

- kāpinot kvadrātā – ievada skaitli, nospiež taustiņus X un =;
- kāpinot 3 pakāpē – ievada skaitli, nospiež taustiņus X,= un =;
- kāpinot 4 pakāpē – ievada skaitli, nospiež taustiņus X,=,= un = vai arī ievada skaitli un nospiež taustiņus X,=,X un =.

Starprezultātus ieraksta darba tabulā un summē kalkulatora atmiņā. Ja kalkulācijas iemaņas ir pietiekami labi attīstītas, tad starprezultātus var arī nepierakstīt. Veicot aprēķinus ar grupētiem datiem, vispirms kāpina pazīmes vērtības vajadzīgajā pakāpē, tad sareizina ar novērojumu skaitu gradācijas klasē, summē kalkulatora atmiņā. 3.10. tabulā rindā "Kopā" ir formulām nepieciešamās kopsummas.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 47. lpp.

3.10. tabula

Darba tabula lokācijas rādītāju aprēķināšanai

Apgrozījums, tūkstošos €	*xi*	*fi*	*xi*fi*	*xi2*fi*	*xi3*fi*	*xi4*fi*
6 – 14	*10*	*9*	*90*	*900*	*9000*	*90000*
14 – 22	*18*	*22*	*396*	*7128*	*128304*	*2309472*
22 – 30	*26*	*12*	*312*	*8112*	*210912*	*5483712*
30 – 38	*34*	*2*	*68*	*2312*	*78608*	*2672672*
38 – 46	*42*	*2*	*84*	*3528*	*148176*	*6223392*
46 - 54	*50*	*5*	*250*	*12500*	*625000*	*31250000*
Kopā		*52*	*1200*	*34480*	*1200000*	*48029248*

$\bar{x} = M_1 = 1200/52 = 23{,}077$ tūkst.€

$s = \sqrt{34480/52-23{,}077^2} = 11{,}425$ tūkst. €

$M_2 = 34480/52 = 663{,}077$

$M_3 = 1200000/52 = 23076{,}92$

$M_4 = 48029248/52 = 923639{,}38$

$\bar{m}_3 = 23076{,}92 - 3*663{,}077*23{,}077 + 2*23{,}077^3 = 1750{,}569$

Arī iepriekšējās formulas aprēķinu var racionalizēt, izmantojot taustiņus M+ un M-.

$A = 1750{,}569/11{,}425^3 = 1{,}17$

Asimetrijas koeficients ir skaitlis bez mērvienības, jo tas ir lielāks, jo nesimetriskāks ir sadalījums. Par to, ka sadalījumam tiešām ir "izstiepts" labais (lielo vērtību) zars, var pārliecināties, apskatot histogrammu 2.2. attēlā.

$\bar{m}_4 = 923639{,}38 – 4*23076{,}92*23{,}077 + 6*663{,}077*23{,}077^2 – 3*23{,}077^4 = 61358{,}65$

$E = 61358{,}65/11{,}425^4 – 3 = 0{,}6$

Sadalījums salīdzinājumā ar normālo ir nedaudz smailāks, apgrozījums vairāk svārstās ap aritmētisko vidējo, liela un maza apgrozījuma nedēļas ir sastopamas retāk.

Ja tiek analizēta viena pazīme vienai kopai, tad var gan uzzīmēt histogrammu, gan tieši analizēt grupējumu. Bet, analizējot lielu datu apjomu – vairākas kopas pēc vairākām pazīmēm, ir vajadzīga lielāka datu kompresijas pakāpe, un tad lieto lokācijas rādītājus. Tos var aprēķināt katru atsevišķi ar funkcijām, kas tika nosauktas pie katra no rādītājiem, bet parasti visus rādītājus aprēķina ar *Excel* darba burtnīcas datu analīzes rīku "*Decription Statistics*" (aprakstošās statistikas rādītāji). Aprēķinus veic tikai negrupētiem datiem. Ja dati ir grupēti, tad aprēķini jāveic, ievadot formulas atbilstošajās šūnās.

Lai aprēķinātu aprakstošās statistikas rādītājus tirdzniecības aģentu apgrozījuma piemēram, ir jāizpilda šādas darbības:

- jāatver *Excel* programma;
- jāievada visi skaitļi vienā ailē;

- jāizpilda komandas: *Data/Data Analysis/Descriptive statistics/OK*;
- logā *Input Range* jānorāda datu apgabals;
- logā *Output Range* jānorāda šūna, no kuras sākt aprēķinus;
- jāieliek ķeksīti pret *Summary Statistics/OK*.

Pēc izpildītajām darbībām tiek iegūta 3.11. tabula.

3.11. tabula

Ar *Excel* iegūtie rezultāti ar tulkojumiem un komentāriem latviski

Mean	22760,6154	Aritmētiskais vidējais
Standard Error	1566,86435	Standartkļūda, būtība un aprēķināšanas metodika tiks skatīta tēmā par izlases metodi
Median	19923,5	Mediāna
Mode	#N/A	Moda nav atrasta, nepārtrauktiem sadalījumiem tā gadās bieži, ja neviena vērtība nav konstatēta biežāk nekā citas
Standard Deviation	11298,8195	Standartnovirze
Sample Variance	127663322	Dispersija
Kurtosis	0,77456975	Ekscess
Skewness	1,1597472	Asimetrija
Range	45445	Variācijas amplitūda
Minimum	6365	Minimālā novērotā vērtība
Maximum	51810	Maksimālā novērotā vērtība
Sum	1183552	Novērojumu kopsumma
Count	52	Novērojumu skaits

Kā redzams, aprēķinot rādītājus no grupētiem datiem, nedaudz tiek zaudēta precizitāte, bet iegūtie rādītāji arī no grupētiem datiem rezultātus interpretē tāpat, kā aprēķinot rādītājus no negrupētiem datiem.

Uzdevums ir aprakstīt iegūtos rezultātus (centrālo tendenci, asimetriju, ekscesu, datu izkliedi utt.).

Šajās nodaļās tika apgūtas prasmes aprakstīt masu objektus, konstatēt situāciju, bet ir jautājums „Nu, un tad?”. Lai izdarītu secinājumus, ir nepieciešami kādi salīdzinājumi. Par tiem tiks runāts nodaļās par secinošo statistiku, bet teorētiskie pamati secinājumu izdarīšanai tiks doti nākamajā nodaļā.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
1	*2*	*3*	*4*
Aritmētiskais vidējais	Arithmetic mean (average)	Арифметическое средее	Novērojumu summa, dalīta ar novērojumu skaitu
Asimetrija	Skewness	Асимметрия	Rādītājs, kurš raksturo sadalījuma simetriskumu. Simetriskam sadalījumam ir nulle
Centrālā tendence	Central tendency	Центральная тенденция	Pamatrādītāji, raksturojot sadalījumu – dažādi vidējie
Deciles	Decile	Децили	Varianšu vērtības, kas ranžētu variācijas rindu dala 10 vienādās daļās
Dispersija	Variance	Дисперсия	Datu izkliedes rādītājs, standartnovirze kvadrātā
Ekscess	Kurtosis	Эксцес	Sadalījuma smailumu raksturojošs rādītājs, normālam sadalījumam tas ir nulle
Ģeometriskais vidējais	Geometric mean	Среднее геометрическое	Izmanto vidējo augšanas tempu aprēķināšanai
Harmoniskais vidējais	Harmonic mean	Гармоничное среднее	Izmanto intensitātes rādītāju vidējās vērtības aprēķināšanai
Interpolācija	Interpolation	Интерполяция	Nezināmu datu vērtības noteikšana zināmā datu apgabalā
Kvadrātiskais vidējais	Squared mean	Среднее квадратичное	Lieto vidējā laukumu lieluma noteikšanai pēc lineāriem mērījumiem
Kvantiles	Fractile	Квантили	Kopīgs apzīmējums kvartilēm, decilēm uc. Variančšu vērtībām, kas dala ranžētu variācijas rindu vienādās noteikta skaita novērojumu daļās
Kvartiles	Quartile	Квартили	Variančšu vērtības, kas dala ranžētu variācijas rindu četrās vienādās daļās
Kvintiles	Quintile	Квинтили	Variančšu vērtības, kas dala ranžētu variācijas rindu piecās vienādās daļās
Mediāna	Median	Медиана	Variantes vērtība, kas dala ranžētu variācijas rindu uz pusēm
Moda	Mode	Мода	Biežāk sastopamā varianta vērtība diskrēta vai atributīva sadalījuma rindā
Modālais intervāls	Modal interval	Модальный интервал	Plašāk pārstāvētais intervāls intervālu variācijas rindā nepārtrauktā sadalījumā
Procentiles	Percentile	Процентили	Variančšu vērtības, kas ranžētu variācijas rindu dala 100 vienādās daļās

1	2	3	4
Standart-novirze	Standard deviation	Стандартное отклонение	Datu izkliedi raksturojošais pamat rādītājs
Stumbrs - lapas	Steam - leaves	Ствол - листья	Grupēšanas metode, kas ļauj sagrupēt datus, nezaudējot precizitāti
Svērtais aritmētiskais vidējais	Weighted arithmetic mean	Взвешанное арифметическое среднее	Jālieto, aprēķinot vidējo vērtību grupētiem datiem vai relatīviem rādītājiem
Variācijas amplitūda	Range	Амплитуда вариации	Starpība starp lielāko un mazāko novēroto vērtību
Variācijas koeficients	Coefficient of variation	Коэффициент вариации	Relatīvā standartnovirze (standartnovirze dalīta ar aritmētisko vidējo)
Vidējā absolūtā novirze	Mean absolute deviation	Среднее абсолютное отклонение	Noviržu no aritmētiskā vidējā moduļu summa, dalīta ar novērojumu skaitu

4. VARBŪTĪBU TEORIJAS PAMATI UN TEORĒTISKIE SADALĪJUMI

Pēc nodaļas apgūšanas studentiem:
- *jāzina notikuma jēdziens, jāizprot klasiskā un statistiskā varbūtība;*
- *jāprot veikt vienkāršākās darbības ar varbūtībām;*
- *jāzina nodaļā apskatītie teorētiskie sadalījumi un to pielietošanas jomas;*
- *jāprot lietot standartizētā normālā sadalījuma funkcijas vērtības, aprēķinot datu intervāla varbūtību, vai noteikt intervāla robežas ar iepriekš definētu varbūtību;*
- *jāzina izlases metodes priekšrocības;*
- *jāsaprot standartnovirzes un standartkļūdas interpretācija;*
- *jāprot aprēķināt matemātiskās cerības ticamības intervāls;*
- *jāzina izlases metodes (izlasē iekļaujamo vienību atlases principi).*

4.1. Varbūtības teorijas pamati

Jau iepriekšējās tēmās, analizējot empīriskos sadalījumus, tika minēts teorētiskā sadalījuma jēdziens. Varbūtību teorija un teorētiskie sadalījumi ir pamats salīdzinājumiem un secinājumiem par dažādu kopu atšķirību statistisko nozīmību. Lai varētu izprast secinājumu būtību izlases metodē, ir jāapgūst varbūtību teorijas pamati.

4.1.1. Notikumu jēdziens

Notikums varbūtību teorijā ir jebkurš fakts, kuru var konstatēt novērojuma vai izmēģinājuma rezultātā.[1,2,3] Notikumu piemēri: uzņēmums gūst peļņu, pārdota prece, noslēgts līgums, no rīta uz ielas pirmais sastaptais cilvēks ir vīrietis (sieviete), pārbaudāmais izstrādājums atbilst kvalitātes prasībām vai nē u.tml.

Par **novērojumu** vai **izmēģinājumu** sauc zināmu apstākļu realizāciju, kā rezultātā var iestāties notikums.[4] Lai kādu notikumu (faktu) varētu konstatēt, ir nepieciešams kāds apstākļu kopums, kuru rezultātā notikums var rasties.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 59. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 93. lpp.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.

Piemēram, lai varētu konstatēt, vai cilvēks nopirks preci, ir jābūt atvērtam veikalam un potenciālajam pircējam šis veikals ir jāapmeklē. Lai novērotu, ka akmens gravitācijas iespaidā nokrīt uz zemes, tad tas ir jāpamet uz augšu. Akmens pamešana gaisā ir tā zināmo apstākļu realizācija, kuru rezultātā būs notikums – akmens nokritīs uz zemes. Ja nav šo apstākļu realizācijas (neviens nemet akmeni), tad akmeņi no gaisa nekrīt. Parasti nepieciešamo apstākļu realizācija tiek uztverta kā pašsaprotama lieta un to īpaši neuzsver, bet, plānojot novērojumu, to ņem vērā.

Izmēģinājums nozīmē aktīvu interesējošo apstākļu kompleksa radīšanu[1,2]. Ekonomikas un uzņēmējdarbības vadības pētījumos pārsvarā ir novērojumi. Pētnieks iegūst informāciju no parastas vides, piemēram, tirgzinības pētījumos pētnieks intervē faktiskos vai potenciālos pircējus tad, kad viņi patiešām iet iepirkties. Izmēģinājumus plaši lieto tehnoloģiskajos pētījumos. Pārbaudot jaunas tehnoloģijas, tās uzreiz neievieš visā ražošanā, bet ražo nelielas izmēģinājuma partijas un salīdzina ar kontroli (kontrolpartijas, kas ražotas ar līdzšinējo ražošanas tehnoloģiju), un tikai pēc pozitīva, statistiski nozīmīgi atšķirīga rezultāta iegūšanas jaunās tehnoloģijas ievieš pilna apjoma ražošanā. Izmēģinājumu plaši izmanto arī produkcijas kvalitātes kontrolē. Izmēģinājumu ekonomikas parādību pētījumos var realizēt ar lietišķo spēļu palīdzību.

Droši sagaidāms notikums iestājas vienmēr, ja ir izveidojusies zināma apstākļu kopa.[3] Piemēram, uz augšu pamesta monēta zemes pievilkšanas spēka ietekmē vienmēr nokritīs uz zemes, nekad tā nepaliks, karājoties gaisā, vai neaizlidos kosmosā.

Neiespējams notikums noteikti neiestājas, ja ir izveidojusies zināma apstākļu kopa.[4] Turpinot iepriekšējo piemēru par monētu, var teikt, ka ir neiespējami, ka uz augšu pamesta monēta nenokrīt zemē.

Gadījumnotikums, pastāvot zināmai apstākļu kopai, var notikt un var arī nenotikt.[5] Uz augšu pamesta monēta var nokrist gan ar aversu (ģerboni), gan reversu uz augšu. Konkrētais notikums, piemēram, pamestā monēta nokrīt ar aversu uz augšu, ir gadījumnotikums.

Nesavienojami notikumi nevar notikt viena novērojuma vai izmēģinājuma rezultātā.[6] Ja uz augšu pamestā monēta nokrīt ar aversu uz augšu, tad tajā pašā metienā šī monēta nevar nokrist ar reversu uz augšu. No rīta pirmais satiktais

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.
[2] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 93. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.
[4] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 93. lpp.
[5] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 59. lpp.
[6] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.

cilvēks ir vai nu vīrietis, vai sieviete – vienlaicīgi tas nevar būt gan vīrietis, gan sieviete. Pretim nākošais cilvēks var būt latvietis vai citas tautības cilvēks.

Savienojami notikumi var notikt kopēji viena novērojuma vai izmēģinājuma rezultātā[1]. Piemēram, no rīta pirmais satiktais cilvēks ir vīrietis un latvietis. Šos abus faktus var konstatēt viena novērojuma rezultātā.

Pilnu notikumu kopu veido visi notikumi, kuri ir zināmi vai tos var iedomāties, no kuriem vienam ir jārodas izmēģinājuma rezultātā[2]. Metot monētu gaisā un skatoties, ar kuru pusi tā kritīs, pilna notikumu kopa būs reverss un averss. Ja students kārto eksāmenu, tad šīs apstākļu realizācijas rezultātā ir iespējams viens no 10 vērtējumiem – pilna notikumu kopa ir eksāmenā saņemtā atzīme (1; 2; 3; 4; 5; 6; 7; 8; 9; 10).

Savstarpēji pretēji (alternatīvi) notikumi ir pilnu notikumu kopu veidojoši, tikai divi nesavienojami notikumi.[3] No apskatītajiem piemēriem alternatīvi notikumi ir sastaptais vīrietis vai sieviete, monēta nokritusi ar reversu vai aversu uz augšu. Bieži vien alternatīvi notikumi tiek vispārināti no plašākas pilnas notikumu kopas. Iegūto vērtējumu eksāmenā var vispārināt kā – students ir nokārtojis eksāmenu, ja ir saņemts vērtējums no 4 līdz 10 ballēm, vai nenokārtojis eksāmenu, ja vērtējums ir 1,2 vai 3 balles. Vai izstrādājums ir kvalitatīvs, nosaka daudz un dažādas pazīmes, bet alternatīvie notikumi ir vai nu izstrādājums atbilst kvalitātes prasībām, vai arī ir brāķis.

Novērojums vai izmēģinājums ir jāorganizē tā, lai visiem notikumiem būtu objektīvi vienāda iespēja notikt katra izmēģinājuma rezultātā, notikumi ir **vienādi iespējami.**[4] Pārbaudot produkcijas partijas kvalitāti, speciāli netiek atlasītas kvalitatīvās vai nekvalitatīvās vienības. Veicot pircēju aptauju sieviešu veļas veikalā, aptaujā gan jaunas, gan vecas sievietes, gan vīriešus, ja pēc nejaušības principa sanāk tos aptaujāt (pētnieks, pabeidzis intervēt iepriekšējo respondentu, aptaujā nākamo satikto pircēju). Notikumu vienāda iespējamība ir pamats pētījuma rezultātu objektivitātei.

Vienāda iespējamība attiecas uz **elementāriem notikumiem**. Tie ir notikumi, kurus nevar vairs tālāk detalizēt.[5] Vai konkrētais aptaujātais pircējs sieviešu veļas veikalā ir vīrietis vai sieviete ir elementārs notikums, bet kopējais aptaujāto vīriešu un sieviešu skaits atšķirsies, sagaidāms, ka pārsvarā tiks aptaujātas sievietes.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 59. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 56. lpp.
[4] Arhipova, I., Bāliņa, S. (2003). Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis. Rīga: Datorzinību Centrs. 60. lpp
[5] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 57. lpp.

Praktiski elementāru notikumu notikšanas vienādu iespēju nodrošināšanai nepieciešama **novērotāja neitralitāte** pret novērojuma rezultātiem. Pie tās pieder arī apstākļu izvērtējums, kas varētu ietekmēt notikuma notikšanu – vai pētījums ir labi sagatavots, un pētījuma apstākļi atbilst parastajiem. Tikai objektīva informācija ir derīga lēmumu pieņemšanai. Pētnieks var būt apzināti vai neapzināti tendēts iegūt kādus noteiktus rezultātus. Tādiem pētījumiem ir sistemātiskā kļūda, un tie neatspoguļo īstenību. Šāda statistiskā izpēte pārvēršas par vislielākajiem meliem, jo aiz zinātniskuma vairoga tiek nomaskētas neapstiprinātas sakarības, kuras vai nu nepastāv vispār, vai to nozīmība nav tāda, kā to uzdod pētnieks. No statistikas viedokļa arī negatīvs rezultāts ir rezultāts, bet problēma ir apstāklī, ka ar negatīviem rezultātiem zinātnieki nevar aizstāvēt disertācijas un profesionāļi cerēt uz izaugsmi karjerā.

Iespējas ir potenciāli notikumi, kas ir pamats kāda reāla notikuma notikšanai.[1]

4.1.2. Varbūtības definīcijas

Varbūtība ir skaitlis, kas raksturo, cik droši ir sagaidāma kāda notikuma notikšana.

Klasiskā varbūtība ir labvēlīgo iespēju (potenciālu notikumu) saskaitīšana un to attiecināšana pret visiem iespējamajiem notikumiem. Notikumam **labvēlīgās iespējas** ir tās iespējas, kas nodrošina notikuma notikšanu. Par notikuma A varbūtību sauc šim notikumam labvēlīgo iespēju skaita M attiecību pret visu vienādi un vienīgi iespējamo, nesavienojamo, elementāro notikumu skaitu N, kuri var rasties viena izmēģinājuma vai novērojuma rezultātā.[2,3]

$$P(A) = \frac{M}{N} \tag{4.1.}$$

Varbūtību apzīmē ar burtu P (*probabilities* – varbūtība angliski), aiz tā iekavās liekot gadījumnotikuma simbolu. Pieraksts P(A), nozīmē notikuma A notikšanas varbūtība. Klasisko varbūtību plaši izmanto spēļu biznesā, ir pat vesela spēļu torija. Spēļu galdi un automāti ir būvēti un spēles organizētas tā, lai spēlētāji kopumā pazaudētu nedaudz vairāk naudas nekā iegūtu. Primitīvs piemērs klasiskajai varbūtībai ir monētas mešana gaisā un minēšana, ar kuru pusi tā nokritīs. Katrs zina, ka iespēja uzkrist aversam uz augšu ir 1/2 jeb 0,5. Ja met spēļu kauliņu, kuram ir 6 skaldnes, tad katra skaitļa uzkrišanas iespēja ir 1/6. Šīs varbūtības var noteikt pirms pieredzes, neveicot izmēģinājumu sēriju,

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 57. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 82. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 95. lpp.

lai noteiktu monētas aversa uzkrišanas iespēju. Tāpēc klasisko varbūtību sauc par ***apriori*** (pirms pieredzes) varbūtību.[1]

Tomēr ekonomikā un uzņēmējdarbībā, atskaitot spēļu biznesa organizēšanu, klasiskā varbūtība parasti nav pielietojama. To nosaka lietojamā izziņas procesa būtība. Klasiskā varbūtības definīcija asociējas ar **deduktīvo** izziņas procesu – vispirms ir definīcija, un tad tā tiek pielietota dzīvē.[2] Daudz biežāk ir jāsaskaras ar pretēju procesu, kad vispirms dabā novēro datu variāciju (veic empīriskos novērojumus) un tad no plašās informācijas ir jāizdara vispārinājumi. Šādu izziņas procesu dēvē par **indukciju** (vispirms ir atsevišķo gadījumu pētījums, un tad tiek vispārināta likumsakarība)[3]. Indukciju pārstāv statistiskā varbūtība.

Skaitli, ap kuru svārstās relatīvais biežums atsevišķās izmēģinājumu vai novērojumu sērijās, sauc par notikuma A **statistisko varbūtību**.[4]

$$V(A) = \frac{m}{n} \,, \qquad (4.2.)$$

kur *V(A)* ir notikuma *A* statistiskā varbūtība,
m – interesējošā notikuma iestāšanās reizes,
n – kopējais novērojumu skaits.

Klasiskās varbūtības atbilstību praksei var pārbaudīt, veicot izmēģinājumus. Elementārus izmēģinājumus ir vērts veikt arī tāpēc, lai saprastu, ka varbūtības darbojas tikai masveida novērojumos. Ir zināms, ka katra skaitļa uzkrišanas varbūtība, metot spēļu kauliņu, ir 1/6. Metot kauliņu vienreiz, uzkritīs kāds viens no sešiem skaitļiem – viena iespēja būs realizējusies, bet pārējās ne. Ja met kauliņu vairāk reižu, tad sagaidāms, ka katrs skaitlis uzkritīs 1/6 daļu no kopējo metienu skaita. Ja kauliņu metīs 10 reizes, tad notikumu sadalījums noteikti nebūs pa 1/6 daļai (tas nemaz nav iespējams, jo 10 nedalās ar 6), pilnīgi iespējams, ka kāds skaitlis nebūs uzkritis vispār, jo novērojumu ir pārāk maz, lai realizētos likumsakarība. Metot kauliņu 100 reizes, katra skaitļa uzkrišanas relatīvais biežums tuvosies 1/6 daļai, bet ļoti iespējams, ka būs kāds skaitlis, kas krīt biežāk, kāds cits – retāk. Atkārtojot izmēģinājumu, proporcijas starp uzkritušajiem skaitļiem būs nedaudz savādākas. Pie ļoti liela novērojumu skaita relatīvais biežums tieksies uz klasisko varbūtību. Vēl jāsaprot, ka vienas iespējas realizācija pat vairākas reizes pēc kārtas nepalielina citas iespējas varbūtību nākamajā mēģinājumā. Visi zina un tic, ka uz augšu pamesta monēta vidēji vienu reizi no divām krīt ar aversu (ģerboni) uz augšu, bet tas nenozīmē, ka vienreiz monēta ir kritusi ar aversu uz augšu, tad nākamajā metienā tā kritīs

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 57. lpp.
[2] Turpat, 59. lpp.
[3] Turpat, 59. lpp.
[4] Turpat, 59. lpp.

ar reversu uz augšu. Varbūtība nākamajā metienā monētai krist ar reversu uz augšu saglabājas 1/2 pat tad, ja ir izdarīti 10 metieni un visos monēta ir nokritusi ar aversu uz augšu. Taču, ja pie ļoti liela novērojumu skaita statistiskā varbūtība netiecas uz klasisko, tas nozīmē, ka notikumu realizācijas iespējas nav vienādas, piemēram, monēta ir saliekta, vai vienā pusē ir blīvāks metāls, spēļu kauliņa skaldnes nav pilnīgi vienādas, kauliņa dažādās malās ir dažāds materiāla blīvums. Vēl viens iemesls, kādēļ statistiskā varbūtība var nesakrist ar klasisko varbūtību pat pie ļoti liela novērojumu skaita, ir tas, ka kaut kādu iemeslu dēļ eksperimenta veicējs nav neitrāls pret iespējamo iznākumu un ar savām darbībām veicina kāda konkrēta iznākuma rašanos. Tāda rīcība pēc savas būtības ir krāpšanās kā spēlē tā, arī zinātniskajā darbā.

Ekonomiskajos pētījumos pārsvarā dominē statistiskās varbūtības, un izvirzītās hipotēzes (loģisks, nepierādīts pieņēmums) pamato ar statistiskajiem datiem.

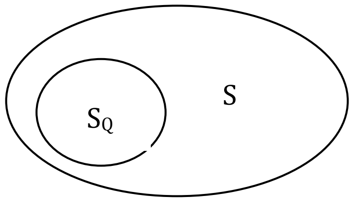
Uzņēmējdarbības vadībā ļoti bieži ir situācija, kad derīgus statistiskos datus, kas ļautu aprēķināt statistiskās varbūtības, nav iespējams iegūt. Ļoti daudzi notikumi uzņēmumā norit tikai vienu reizi. Neskatoties uz to, uzņēmējdarbības vadībā varbūtības, it īpaši sadzīvisku spriedumu veidā, pielieto diezgan plaši, varbūtību pielietojums aprēķinos ir retāks. Eksperts (uzņēmuma vadītājs, zemākā līmeņa vadītāji vai pieaicinātie speciālisti) novērtē varbūtības bez atbilstošas datu bāzes un aprēķiniem. Šādi noteiktas varbūtības sauc par **subjektīvām varbūtībām.**[1] Statistiskā un klasiskā varbūtība ir objektīvas varbūtības. Subjektīvās varbūtības balstās uz dažādiem spriedumiem, loģisko analīzi. Subjektīvo varbūtību objektivitāti var pārbaudīt tikai pēc notikumiem. Izdarīt secinājumus pēc viena notikuma nav pareizi. Piemēram, eksperts paredz nelielu varbūtību, ka projekts nesīs zaudējumus, ticamāk, ka projekts būs rentabls. Realizējot šo projektu, tas izrādās nerentabls. Šis fakts vēl nenozīmē eksperta nekompetenci. Varbūtību likumsakarības parādās tikai masveida novērojumos. Šīs nostādnes labāku izpratni sniegs klasisks varbūtību piemērs: metot spēļu kauliņu, zaudē, ja uzkrīt vieninieks, citos gadījumos vinnē. Tātad zaudējuma varbūtība ir 1/6, bet vinnesta varbūtība ir 5/6. Metot kauliņu tikai vienu reizi (projekts arī tiek realizēts tikai vienu reizi), var uzkrist vieninieks, tādēļ nav pamata apstrīdēt noteiktās varbūtības. Ja daudzu reālo notikumu sadalījums apmēram atbilst prognozētajām varbūtībām, var teikt, ka ir labs eksperts. Subjektīvo varbūtību lietošanu pamato tas, ka cilvēka smadzenes apstrādā ļoti lielu informācijas apjomu, pašas atrod informācijas apstrādes algoritmus. Ļoti liela informācija daļa tiek apstrādāta zemapziņā. Noniecināt šādi iegūtu informāciju tikai tādēļ, ka tā nepakļaujas formāliem zinātniskuma kritērijiem, nebūtu pareizi. Galvenais labums no subjektīvo varbūtību un atbilstošo kalkulāciju lietošanas ir tas, ka vadītāji labāk izprot savu

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 148.

uzņēmējdarbību un īsākā vai ilgākā laikā noteikti pieņems pareizākus lēmumus nekā tad, ja šāda analīze netiku veikta.

Vēl varbūtībām ir ģeometriskā interpretācija. Tā nav pielietojama formula, bet varbūtības vizualizācija. **Ģeometriskā varbūtība** pilno notikumu kopu attēlo ar vienu ģeometrisku figūru, bet interesējošo jeb labvēlīgo notikumu kopu – ar citu pirmajā figūrā ietvertu figūru.[1]

$$P(Q) = \frac{S_Q}{S} \tag{4.3.}$$



4.1.attēls. **Varbūtību ģeometriskā interpretācija**

4.1.3. Varbūtības galvenās īpašības

1. Droši sagaidāma notikuma varbūtība ir viens.
2. Neiespējama notikuma varbūtība ir nulle.

   Statistikā nepēta nevariējošus lielumus, tāpēc šīs īpašības nav aktuālas. Izprast šīs īpašības var, apskatot piemēru ar droši sagaidāmu un neiespējamu notikumu. Droši sagaidāms notikums ir, ka uz augšu pamests akmens nokritīs zemē, bet neiespējams ir tas, ka akmens nenokritīs. Metot akmeni, piemēram, 5 reizes (*n*), akmens nokrīt zemē 5 reizes (*m*). Varbūtība, ka mests akmens nokritīs zemē, ir *V* (nokritīs) = 5/5 = 1 (pēc 4.2. formulas). Savukārt no 5 metieniem ne reizi akmens nepaliek, karājoties gaisā *V* (nenokritīs) = 0/5 = 0.
3. Gadījumnotikuma varbūtība ir skaitlis, lielāks par nulli un mazāks par vienu.

   0 < P(A) < 1.
4. Pilnas notikumu kopas varbūtību summa vienmēr būs viens. Students, kurš atnācis kārtot eksāmenu, noteikti saņems vērtējumu (iegūs atzīmi no 1 līdz 10), metot spēļu kauliņu, noteikti uzkritīs kāds skaitlis u.tml.
5. Praktiskajā interpretācijā varbūtību var izteikt procentos vai arī kā attiecību (1:10, 2:3 u. tml.). Ar viena daļās izteiktām varbūtībām ir ērtāk veikt matemātiskās darbības, tāpēc kalkulācijām varbūtības izsaka decimāldaļās.
6. Praktiskā darbā notikumu uzlūko par droši sagaidāmu, ja tā iestāšanās varbūtība ir tuva vienam.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 59. lpp.

Pārbaudot statistiskās hipotēzes, maksimālo varbūtību, ar kuru kādu apgalvojumu var pieņemt par praktiski neiespējamu, sauc arī par **nozīmības līmeni (α)**.[1] Ekonomikas un uzņēmējdarbības vadības pētījumos pieļauj lielāku kļūdīšanās iespēju (α ir 0,05, vai pat 0,1). To nosaka tas, ka daudzus pētījumus nevar atkārtot, apstākļi ļoti strauji mainās, vai arī nav ekonomiski pamatoti (pārāk augstas izmaksas) iegūt augstāku pārliecības līmeni. Nozīmības līmenis nozīmē to, ka par praktiski neiespējamu notikumu uzskatīs to, kura notikšanas varbūtība ir mazāka par nozīmības līmeni. Ir jārēķinās ar to, ka par praktiski neiespējamu atzīts notikums tomēr reizēm notiek. Bioloģiskajos un it īpaši medicīnas pētījumos pieļaujamais kļūdīšanās slieksnis ir noteikts 0,01 vai pat 0,001. Tas nozīmē, ka nevēlams notikums var notikt ne biežāk kā vienā gadījumā no 100 vai pat 1000 reizēm. Tas ir saistīts ar diviem momentiem. Pirmkārt, nav pieļaujamas zāles, kuras rada būtisku kaitējumu. Otrkārt, tehnoloģiskajos pētījumos, bioloģijā, medicīnā var veikt rūpīgus eksperimentus, kuros tiek ievērots *ceteris paribus* (pie pārējiem nemainīgiem nosacījumiem) princips, un iegūt statistiski nozīmīgus datus. Uzņēmējdarbībā katrs notikums ir vienreizējs, un, pat ja līdzīgi notikumi atkārtojas, ***ceteris paribus***[2] princips pilnībā netiek ievērots, jo mainās arī citi nosacījumi, ne tikai pētāmais nosacījums.

4.1.4. Vienkāršākās darbības ar varbūtībām

Darbības var veikt gan ar klasiskajām, gan statistiskajām varbūtībām, bet vieglāk ir saprast klasiskās varbūtības, tāpēc darbības ar varbūtībām ir vērts izanalizēt ar klasiskās varbūtības piemēriem.

Varbūtību saskaitīšanu lieto tad, ja ir jāaprēķina gadījuma notikumu apvienojuma jeb loģiskās summas varbūtība:

$$C = A \cup B, \qquad (4.4.)$$

kur $\cup$ ir loģiskās summas zīme, ko lasa kā “vai”.

Ar notikumu *C* saprot gan notikumu *A*, gan notikumu *B*.[3,4,5] Piemēram, daudzās kauliņu spēlēs atkārtoti kauliņu met, ja uzkrīt 1 vai 6.

Varbūtību, ka viena izmēģinājuma rezultātā notiks viens no šiem notikumiem, aprēķina kā abu notikumu varbūtību summu:

$$P(A \cup B) = P(A) + P(B) \qquad (4.5.)$$

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 60. lpp.
[2] McConnell, C., Brue, S. (1996). *Economics: principles, problems, and policies*. (13th ed.) New York: McGraw-Hill, Inc. p.4.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 61. lpp.
[4] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 150.
[5] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 61. lpp.

Katra notikuma varbūtība ir 1/6, bet abu šo notikumu apvienojums (varbūtība, ka būs jāmet atkārtoti) būs 1/6 + 1/6 = 2/6 = 1/3. Tātad vidēji spēles dalībniekiem katrā trešajā gājienā met kauliņu atkārtoti.

Vispārīgā formā nenoteikta notikumu skaita notikumu apkopojums ir:

$$P(A \cup B \cup \ldots \cup Z) = P(A) + P(B) + \ldots P(Z) \tag{4.6.}$$

Pilnu kopu veidojošu gadījuma notikumu varbūtību summa ir viens (varbūtības īpašība):

$$P(A_1) + P(A_2) + \ldots P(A_n) = 1 \tag{4.7.}$$

Ja pilnu notikumu kopu veido divi alternatīvi notikumi, tad 4.4. formulu saīsina un pieraksta formā, kas izceļ to, ka tie ir divi pretēji notikumi:

$$P(A) + P(\overline{A}) = 1 \tag{4.8.}$$

jeb

$$p + q = 1 \tag{4.9.}$$

Neatkarīgu notikumu varbūtību reizināšana

Varbūtību reizināšanu lieto, ja ir jāaprēķina gadījuma notikumu kopējās notikšanas varbūtība.

Divus gadījuma notikumus sauc par **savstarpēji neatkarīgiem**, ja viena notikuma iestāšanās varbūtība nav atkarīga no tā, vai otrs notikums jau noticis, vai nē (un arī otrādi).

Divu neatkarīgu gadījumnotikumu kopējās notikšanas varbūtība ir vienāda ar abu šo notikumu varbūtību reizinājumu:

$$P(A \cap B) = P(A)P(B)\,, \tag{4.10.}$$

kur $\cap$ – loģiskā reizinājuma summa, ko lasa kā "un".[1,2,3]

Piemēram, ja bundžā sakrata divas monētas un izber uz galda, tad varbūtība, ka abas monētas būs ar ģerboni uz augšu, ir:

$$P(A \cap B) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4} = 0{,}25$$

Katras monētas uzkrišanas varbūtība ar ģerboni uz augšu ir ½, bet abu monētu kopējā varbūtība uzkrist ar ģerboni uz augšu ir tikai ¼.

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 158.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 62. lpp.
[3] Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 97. lpp.

Līdzīgi var aprēķināt arī varbūtību diviem spēļu kauliņiem uzkrist ar vieniniekiem uz augšu: 1/6*1/6 = 1/36

Darbības ar savstarpēji atkarīgu notikumu varbūtībām

Ja viena notikuma varbūtība ir atkarīga no tā, vai otrs notikums ir noticis, vai nē, tad šie notikumi ir **savstarpēji atkarīgi**.

Ja notikumi A un B ir savstarpēji atkarīgi, tad par **nosacīto varbūtību** $P_A(B)$ sauc notikuma B varbūtību, pieņemot, ka A jau noticis.

Divu atkarīgu notikumu **kopējās notikšanas varbūtību** aprēķina, reizinot pirmā notikuma parasto varbūtību ar otra notikuma nosacīto varbūtību:[1,2,3]

$$P(A \cap B) = P(A) P_A(B) \qquad (4.11.)$$

> *4.1. piemērs*. Studentu dienas svinībām Jelgavā ir pieteikušies 28 studenti, bet ir tikai 13 vietas augstskolas mikroautobusā. Braucējus ir nolemts izvēlēties ar izlozes palīdzību. Aprēķināt, kāda ir varbūtība, ka divi draugi, kas velk lozes pirmie, abi brauks uz Studentu dienas svinībām.

Varbūtība, ka pilnu lozi izvilks pirmais draugs, ir 13/28 (13 pilnās lozes no 28 lozēm kopā). Ja pirmais draugs izvelk pilno lozi, tad pastāv varbūtība, ka arī otrs draugs izvilks pilno lozi – viņam tā ir 12/27 (viena pilnā loze ir izvilkta). Šī ir nosacītā varbūtība, jo tai ir nosacījums, ka pirms tā ir izvilkta pilnā, nevis tukšā loze. Ja pirmais draugs ir izvilcis tukšo lozi, tad varbūtības, ka abi draugi brauks (pēc izlozes rezultātiem), nav.

Divus gadījumnotikumus sauc par **savienojamiem**, ja viena notikšana neizslēdz otra notikšanu tajā pašā izmēģinājumā.[4]

Varbūtība, ka notiks vismaz viens no savstarpēji savienojamiem notikumiem, ir vienlīdzīga šo notikumu varbūtību summai, no kuras atņemta abu notikumu kopīgas iestāšanās varbūtība.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) \qquad (4.12.)$$

Piemēram, par vinnētāju atzīs tad, ja no 36 kāršu kavas tiks izvilkts vai nu kungs, vai kreicis. Kāršu kavā ir 4 kungi, varbūtība izvilkt kungu ir 4/36, un kavā ir 9 kreiči – varbūtība izvilkt kreici ir 9/36, taču var izvilkt arī kreiču kungu, kas ir abu šo notikumu apvienojums. Aprēķināt vinnesta varbūtību ar 4.12. formulu.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 65. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 169.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 61. lpp.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 66. lpp.

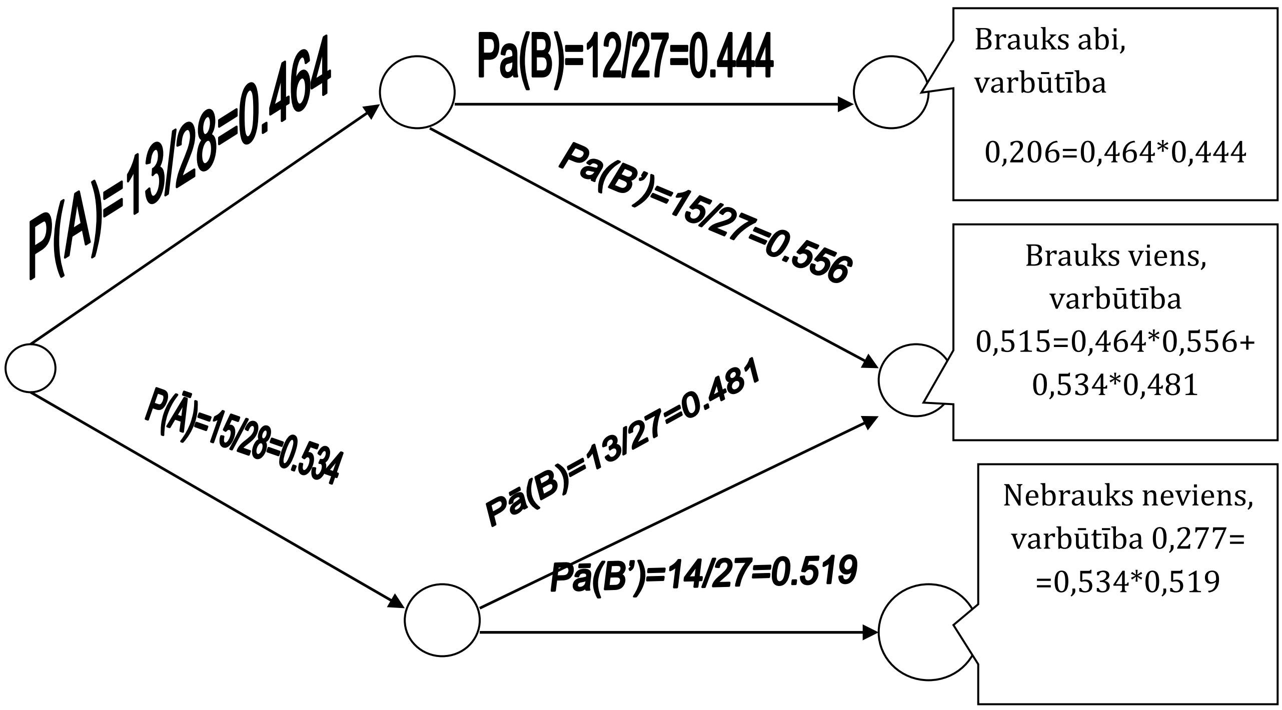
$$P(A \cup B) = \frac{4}{36} + \frac{9}{36} - \frac{4}{36} * \frac{9}{36} = \frac{12}{36} = \frac{1}{3}$$

Aprēķinu pareizību ir viegli pārbaudīt, atdalot „vinnējošās” kārtis un tās saskaitot.

Lēmumu koks

O.Krastiņš to sauc par **virzīto grafu attēlu,** un tas ir grafisks iespējamo notikumu un lēmumu attēlojums.[1,2]

Apskatot piemēru par izlozi braucienam uz Studentu dienas svinībām, var iegūt visu iespējamo notikumu uzskaitījumu (4.2. attēls).



4.2. attēls. **Pilnas notikumu kopas uzskaitījums 4.1. piemēram**

Saskaitot visas varbūtības (0,206+0,515+0,277), ir 0,998. Pilnas notikumu kopas varbūtība ir 1, jo viens no šiem notikumiem noteikti notiks. 2 tūkšdaļas iztrūkst rezultātu apaļošanas dēļ.

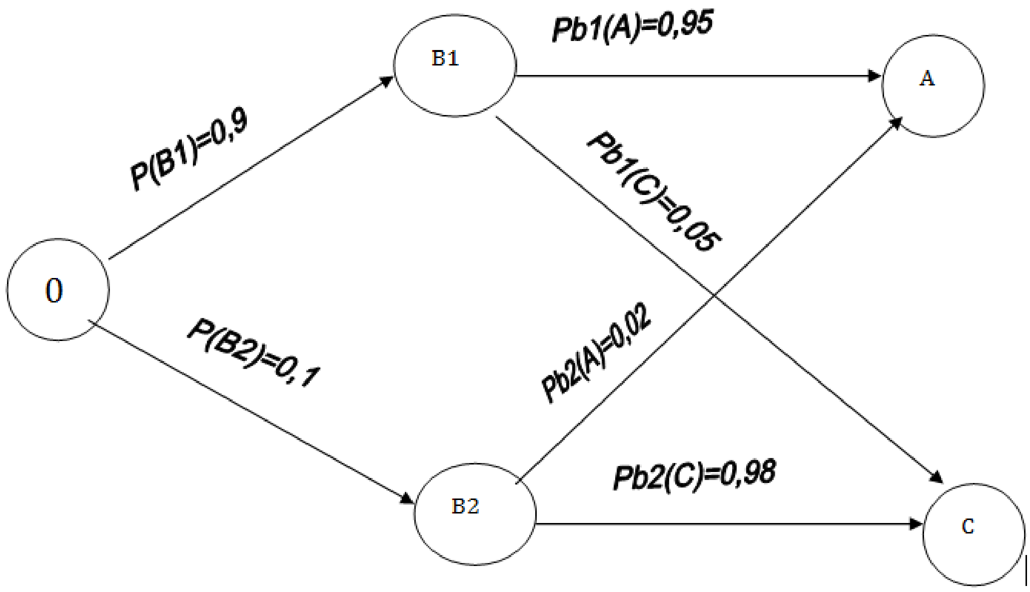
[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 89. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 181.

Beijesa teorēma

Nosacītajās varbūtībās rēķinās ar to, ka kāds notikums jau ir noticis. Līdz vienādam secinājumam (rezultātam) var nonākt pa dažādiem ceļiem, taču šajā gadījumā interesē viena ceļa varbūtība.[1,2]

> *4.2. piemērs*. Ir zināms, ka no saražotās produkcijas kāds noteikts procents ir brāķis (brāķa varbūtība – 0,1). Kontrolieris, kas šķiro detaļas, pārsvarā sašķiro pareizi, bet dažreiz kļūdās. Precīzākas pārbaudes atklāj, ka kvalitātes kontrolē nepareizi izbrāķē 5 % derīgu detaļu (varbūtība – 0,05), bet 2 % (varbūtība – 0,02) no kvalitātei neatbilstošajām detaļām atzīst par derīgām. Uzņēmuma vadībai interesē, kāda ir brāķa varbūtība par derīgajām atzītajās detaļās.

Uzdevuma risinājuma izpratnei labāk izmantot virzīto grafu attēlu (skat. 4.3. att.).



4.3. attēls. **Notikumu un varbūtību uzskaitījums 4.2. piemēram**

„0” ir produkcijas ražošanas punkts, „B” – kvalitātes kontrole: „B1” – derīgo detaļu šķirošana, „B2” – brāķa detaļu šķirošana, „A” – par derīgām atzītās detaļas, „C” – par nederīgām atzītās detaļas.

Varbūtību sauc par **hipotēzes varbūtību** tad, ja reālais plūsmas ceļš ir OBA, bet, pieņemot, ka notikums A jau ir noticis, jānovērtē atpakaļvirzienā, kurā no ceļiem tas varētu būt noticis.

---

1 Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 70. lpp.
2 Goša, Z. (2003). *Statistika. Rīga: SIA "Izglītības soļi"*. 192. lpp.

Šādus uzdevumus vispārina **Beijesa formula**. 4.2. piemēram pilno varbūtību veido divi saskaitāmie, un Beijesa formula ir šāda:[1,2]

$$P_A(B_2)=\frac{P(B_2)P_{B_2}(A)}{P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A)}, \qquad (4.13.)$$

Formulas skaitītājā ir interesējošā ceļa (notikumu secības – nederīgu, bet par derīgām atzītu detaļu) varbūtība pret kopējo rezultāta varbūtību saucējā (norādītas divas iespējamo notikumu secības, kas rada vienādu rezultātu – derīgas detaļas, kas par tādām arī atzītas, un nederīgas, bet par derīgām atzītas detaļas).

$$P_A(B_2)=\frac{0{,}1*0{,}02}{0{,}9*0{,}95+0{,}1*0{,}02}=0{,}00233$$

Brāķa varbūtība ir nedaudz lielāka par 0,2%.

<u>Beijesa</u> formula vispārīgā veidā ir[3]:

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+\ldots+P(B_k)\cdot P_{B_K}(A)} \qquad (4.14.)$$

Nākamajā apakšnodaļā tiks apskatīti teorētiskie sadalījumi. Tie kalpo par bāzi salīdzinājumiem un secinājumiem par ģenerālkopu, veicot nepilno statistisko novērošanu.

4.2. Teorētiskie sadalījumi

4.2.1. Diskrēti sadalījumi

Ja izmēģinājuma vai novērojuma rezultātu var izteikt ar skaitli, tad sadalījums ir kvantitatīvs. Konstatējamās pazīmes var būt diskrētas[4,5,6] (tas ir, tikai konkrētas, parasti veseli skaitļi, vērtības, piemēram, ģimenes locekļu skaits vai istabu skaits dzīvoklī u. tml.).

Diskrēti sadalījumi dabas un ekonomiskajos pētījumos nav bieži sastopami. Bet, tā kā matemātiskās darbības ar diskrētu sadalījumu varbūtībām ir vienkāršākas, tad ekonomiskajās prognozēs reāli nepārtraukts sadalījums bieži tiek aizvietots ar diskrētu sadalījumu. Piemēram, projekta realizācijas gaitā

[1] Goša, Z. (2003). *Statistika. Rīga: SIA "Izglītības soļi"*.192. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 70. lpp.
[3] Turpat.
[4] Turpat, 75.lpp.
[5] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 205.
[6] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 63. lpp.

iegūtā peļņa ir nepārtraukti variējošs lielums, bet savos aprēķinos bezgalīgi daudzo variantu vietā var izmantot dažus scenārijus. Projekta realizācijas gaitā var iegūt jebkuru peļņu – no absolūta ieguldījuma zaudējuma līdz milzu ieguvumam. Tiesa, ekstrēmo vērtību varbūtības ir ļoti niecīgas, bet arī ticamākajā intervālā variantu skaits ir neierobežots. Taču, vērtējot projektu, tiek izstrādāti daži (3–5) scenāriji. Piemēram, pie sliktākās notikumu attīstības secības (pesimistiskais scenārijs) rodas 15 tūkstošu € zaudējums. Šim skaitlim piemēro varbūtību, parasti tā ir subjektīvā varbūtība, kas daļēji balstās uz neformālu pagātnes datu un pašreizējo notikumu analīzi. Līdzīgā veidā novērtē ticamākos un optimistiskos scenārijus (trīs scenāriju variantā, ja ir vairāk, tad tiem izdomā atbilstošus nosaukumus, piemēram, mēreni piesardzīgais scenārijs u. tml.).

4.3. piemērs. Ar statistisko varbūtību palīdzību ir iegūtas bankas nodaļas nedēļā izsniegto kredītu skaita varbūtības.

Izsniegto kredītu skaits	Varbūtība
0	0,10
1	0,15
2	0,25
3	0,30
4	0,10
5	0,05
6	0,05
	1,00

Šī sadalījuma parametru aprēķināšanai var izmantot formulas empīrisko novērojumu raksturotāju aprēķināšanai. Atšķirība ir tā, ka relatīvie biežumi ir vispārināti statistiskajās varbūtībās, un var uzskatīt, ka šāds sadalījums saglabāsies arī turpmāk.

Aprēķiniem tiks izmantota 3.4. formula, par frekvencēm – varbūtības.

Teorētisko sadalījumu vidējo aritmētisko apzīmē ar μ (grieķu burts, izrunā "mī") un sauc par **matemātisko cerību**.[1,2,3]

3.4. formula diskrētiem teorētiskajiem sadalījumiem kļūst šāda:

$$\mu = \sum x_i P_i \,, \qquad (4.15)$$

kur x_i – pazīmes vērtība;

Pi – notikuma (pazīmei būs x_i vērtība) varbūtība.

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 65. lpp.

[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 81. lpp.

[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 219.

Teorētisko sadalījumu standartnovirze arī tiek saukta par standartnovirzi, bet tiek apzīmēta ar grieķu burtu σ (sigma). Aprēķinu formulas ir šādas:[1,2,3]

$$\sigma = \sqrt{\sum \left(x_i - \overline{x}\right)^2 P_i} \tag{4.16.}$$

vai

$$\sigma = \sqrt{\sum x_i^{\ 2} P_i - \mu^2} \tag{4.17.}$$

Aprēķini 4.3. piemēram ir parādīti 4.1. tabulā

4.1. tabula

Matemātiskās cerības un standartnovirzes aprēķins 4.3. piemēram

Izsniegto kredītu skaits, xi	*Varbūtība, Pi*	*xi*Pi*	*xi2*Pi*
0	*0,10*	*0,00*	*0,00*
1	*0,15*	*0,15*	*0,15*
2	*0,25*	*0,50*	*1,00*
3	*0,30*	*0,90*	*2,70*
4	*0,10*	*0,40*	*1,60*
5	*0,05*	*0,25*	*1,25*
6	*0,05*	*0,30*	*1,80*
	1,00	*2,50*	*8,50*

$$\mu = \sum x_i P_i$$

$$\sum x_i^{\ 2} P_i$$

Matemātiskā cerība (vidējo gaidāmo izsniegto kredītu skaits nedēļā) ir aprēķināta tieši tabulā.

Standartnovirzes aprēķins ir šāds:

$$\sigma = \sqrt{8{,}5 - 2{,}5^2} = 1{,}5$$

Šāda tipa sadalījumu praksē lieto bieži bez uzsvara uz tā teorētiskumu. Tālāk tiks apskatīts binomiālais sadalījums, kas arī ir diskrēts teorētiskais sadalījums un kam ir lielāka nozīme teorētisko sadalījumu izpratnē salīdzinājumā ar iepriekš apskatītajiem, kurus var uztvert kā empīrisko sadalījumu vispārinājumu.

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 220.

[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 82. lpp.

[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 65. lpp.

4.2.2. Binomiālais sadalījums

Īpašības:

1) *bi* – nozīmē divi. Pazīmei ir tikai divas vērtības, kuras varētu apzīmēt – “veiksme”, “neveiksme”;
2) ja veiksmes (notikums tiek konstatēts) varbūtība ir *p*, tad neveiksmes varbūtība ir $q=1-p$ (4.9. formula, pilnu notikumu kopu veido divi notikumi, kuru varbūtību summa vienmēr būs viens – vismaz viens notikums noteikti iestāsies);
3) katra individuālā izmēģinājuma rezultāts nav atkarīgs no cita izmēģinājuma rezultāta. Lai garantētu rezultātu neatkarību, izlase tiek veidota no milzīgas ģenerālkopas, vai arī jau pārbaudītās kopas vienības tiek atgrieztas ģenerālkopā, un tām atkal ir iespēja iekļūt izlasē atkārtoti. Tas nodrošina, ka vienības nokļūšana izlasē ir ar tādu pašu varbūtību katrā mēģinājumā;
4) izlase ir kāds noteikts mēģinājumu skaits – *n*.

Binomiālo sadalījumu lieto, lai novērtētu, cik veiksmes gadījumu būs *n* mēģinājumos.[1,2,3]

> *4.4. piemērs*. *Coca–cola* produktu virzīšanas kampaņā izmantoja reklāmas saukli “Pudeles rodas zem korķa”. Vidēji katrai trešajai kolas pudelei korķa iekšpusē ir pudeles zīmējums, kas ļauj korķa īpašniekam saņemt vēl vienu pudeli bez maksas. Pudeļu veikalā ir ļoti daudz, un atsevišķi pirkumi ar to veiksmēm vai neveiksmēm neietekmē kopējo veiksmju un neveiksmju varbūtību.
>
> Jāaprēķina binomiālā sadalījuma varbūtību sadalījums, ja tiek veikti trīs (*n*=3), pieci un desmit mēģinājumi (tas ir, vienā pirkšanas reizē tiek nopirktas trīs, piecas vai desmit pudeles).
>
> Jāapprēķina sadalījuma matemātiskā cerība un standartnovirze.
>
> Iegūtie sadalījumi jāattēlo ar poligoniem.

Ja tiek nopirktas 3 pudeles, tas nenozīmē, ka automātiski tiks saņemta arī ceturtā. Iespējamie notikumi:

- visas tukšas (ne vienai no trim pudelēm zem korķa nav pudeles attēla);
- viena „pilna”;
- divas „pilnas”;
- visas trīs „pilnas”.

---

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 68. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 90.lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 230.

Binomiālā sadalījuma varbūtības aprēķina ar Bernulli formulu.[1,2,3]

Par **atkārtotu novērojumu** vai **izmēģinājumu** sauc vienas un tās pašas apstākļu kopas atkārtotu radīšanu, kuras rezultātā notikums A var notikt vai nenotikt.

Ir divas situācijas.

1. Ja interesē varbūtība, ka notikums A notiks m reizes iepriekš fiksētā secībā, piemēram, visas m reizes novērojumu sērijas sākumā:

$$P_n(m, fiks.) = p^m q^{n-m}, \tag{4.18.}$$

kur $P_n(m, fiks.)$ – varbūtība, ka notiks *m* notikumi no *n* mēģinājumiem, iepriekš noteiktā secībā;
p – labvēlīgā notikuma varbūtība;
q – pretējā notikuma varbūtība;
m – labvēlīgā notikuma skaits iepriekš fiksētā secībā;
n-m – pretējā notikuma skaits.

2. Biežāk interesē varbūtība, ka notikums A iestāsies tieši m reizes, bet iestāšanās un neiestāšanās gadījumu secība nav noteikta, tā var būt jebkura. Tad iepriekšējā varbūtība ir jāpareizina ar kombināciju skaitu, ko var sastādīt no n elementiem pa m elementiem:

$$P_n(m) = C_n^m p^m q^{n-m}, \tag{4.19.}$$

kur C_n^m – kombināciju skaits, kas var veidoties m labvēlīgiem rezultātiem n mēģinājumos.

Kombināciju skaitu var aprēķināt pēc 4.20.formulas:

$$C_n^m = \frac{n!}{m!(n-m)!}, \tag{4.20.}$$

Izsaukuma zīme ir matemātisks simbols, kas apzīmē faktoriālu. Jāzina, ka 0!=1, bet pārējo skaitļu faktoriālus nosaka , sareizinot visus naturālos skaitļus no 1 līdz tam skaitlim, kura faktoriālu meklē. Tā, piemēram, 3!=1*2*3=6.

Bernulli formula ir:

$$P_n(m) = \frac{n!}{m!(n-m)!} \cdot p^m q^{n-m} \tag{4.21.}$$

Bernulli formula ir vienāda ar attiecīgo Ņūtona binoma izvirzījuma locekli, ja binomu ņem pakāpē n, kas atbilst izmēģinājumu skaitam.[4]

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 234.
[2] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 195. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 234.
[4] Turpat.

$$(p+q)^n = p^n + np^{n-1}q + C_n^{n-2}p^{n-2}q^2 + C_n^{n-3}p^{n-3}q^3 + \\ + \ldots + \underbrace{C_n^m p^m q^{n-m}}_{\text{Bernulli formula ir Ņūtona binoma izvirzījuma vispārējais loceklis}} + \ldots + q^n. \tag{4.22.}$$

Kombināciju skaita aprēķināšana ekonomikas un vadības fakultāšu studentiem var sagādāt grūtības, jo specialitātes aprēķini neprasa zinātniskā kalkulatora lietošanu, bet parastajiem grāmatvedības kalkulatoriem nav faktoriāla aprēķināšanas funkciju.

Faktoriālus var aprēķināt secīgi, zinot, ka 0! =1; 1!=1 un 2!-2, tālākos faktoriālus aprēķina pēc formulas n!=(n-1)!*n. 3!=2!*3=2*3=6; 4!=3!*4=9*4=24 utt.

Bet visvienkāršākais veids kombināciju skaita aprēķināšanai ir Paskāla trīsstūra izmantošana.[1]

4.4. attēlā ir dots aprēķināts Paskāla trīsstūris līdz 10 mēģinājumiem. 4.5. attēlā ar bultiņām ir parādīts, kā aprēķina Paskāla trīsstūra skaitļus.

									1		1									
								1		2		1								
							1		3		3		1							
						1		4		6		4		1						
					1		5		10		10		5		1					
				1		6		15		20		15		6		1				
			1		7		21		35		35		21		7		1			
		1		8		28		56		70		56		28		8		1		
	1		9		36		84		126		126		84		36		9		1	
1		10		45		120		210		252		210		120		45		10		1

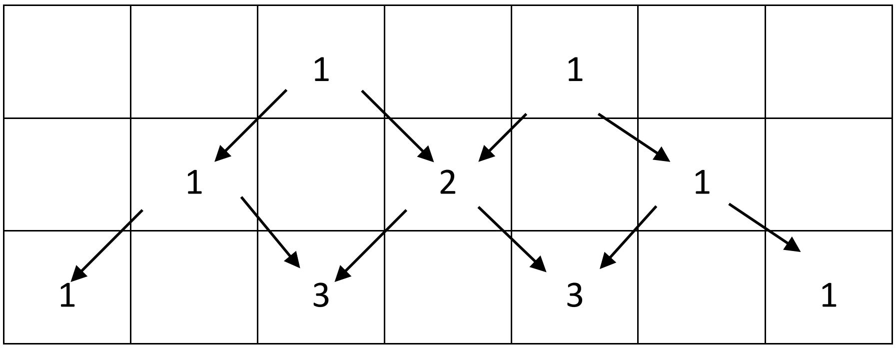
4.4. attēls. **Paskāla trīsstūris līdz desmit mēģinājumiem** (*n* – no 1 līdz 10)

Paskāla trīsstūrim ir noteikta aizpildīšanas kārtība.

1. Lapas vidū ar vienas rūtiņas atstarpi ieraksta divus vieniniekus. Tas nozīmē, ja ir viens mēģinājums, tad ir viena iespēja neveiksmei un viena iespēja veiksmei.

[1] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 53. lpp.

2. Nākamās rindiņas aizpilda, sasummējot pa diagonāli pieskarošajās rūtiņās esošos skaitļus. Katra nākamā rindiņa ir par vienu rūtiņu uz abām pusēm plašāka, tajās ieraksta vieninieku, bet rūtiņas, kurām pieskaras divas skaitļu rūtiņas no iepriekšējās rindas, ir šo pieskarošos rūtiņu skaitļu summa. 4.5. attēlā ir parādīts, kā veidojas nākamo skaitļu (kombināciju rindas).



4.5. attēls. **Paskāla trīsstūra veidošanas princips**

3. Interpretē iegūtos skaitļus. Piemēram, trešajā rindā (n=3) viena kombinācija ir iespējama visām trim neveiksmēm, 3 kombinācijas – vienai veiksmei. To var pārbaudīt arī ar 4.20. formulu:

$C_n^m = \frac{n!}{m!(n-m)!} = \frac{3!}{1!(3-1)!} = \frac{6}{1*2} = 3$, 3 kombinācijas 2 veiksmēm un viena kombinācija visām trim veiksmēm.

Binomiālā sadalījuma matemātisko cerību aprēķina šādi:

$$\mu = n * p \tag{4.23.}$$

un standartnovirzi:

$$\sigma = \sqrt{n * p * q} \tag{4.24.}$$ [1,2,3]

Tālāk tiks aprēķināts 4.4. piemēra uzdevums. 4.2. tabulā ir dots varbūtību sadalījums, ja n=3.

Pēc tabulas redzams, ka, pērkot trīs pudeles, varbūtība nevinnēt ne vienu papildus pudeli ir 0,29. Vislielākā varbūtība ir vinnēt vienu pudeli, un tā ir 0,44, visas trīs pudeles vinnēt varbūtība ir tikai nepilni 4 % (0,037). Varbūtību summa ir 1 (varbūtības īpašības – pilnas notikumu kopas varbūtību summa ir viens, izpildīšanās).

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 92. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 69. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 243.

4.2. tabula

Binomiālā sadalījuma varbūtību aprēķins 4.4. piemēram (n=3)

Veiksmju skaits, m	*Kombināciju skaits,* C_n^m	p^m = *0,33m*	q^{n-m} = *0,67n-m*	$P_n(m)$	$x * f$	$x^2 * f$
0	*1*	*1*	*0,296296*	*0,296296*	*0*	*0*
1	*3*	*0,333333*	*0,444444*	*0,444444*	*0,444444*	*0,444444*
2	*3*	*0,111111*	*0,666667*	*0,222222*	*0,444444*	*0,888889*
3	*1*	*0,037037*	*1*	*0,037037*	*0,111111*	*0,333333*
				1	*1*	*1,666667*

Pēdējās divas ailes ir pievienotas, lai parādītu, ka matemātiskās cerības un standartnovirzes aprēķiniem var izmantot 4.15. un 4.17. formulu ($\mu = \sum x_i P_i = 1$ *;* $\sigma = \sqrt{\sum x_i P_i - \mu^2} = \sqrt{1{,}67 - 1^2} = 0{,}816$ *). Rezultāti sakrīt ar 4.23. un 4.24. formulas dotajiem rezultātiem:*

$$(\mu = n * p = 3 * 0.33 = 1; \ \sigma = \sqrt{n * p * q} = \sqrt{3 * 0{,}33 * 0{,}67} = 0{,}816)$$

4.3. tabula

Binomiālā sadalījuma varbūtību aprēķins 4.4. piemēram (n=5)

Veiksmju skaits, m	*Kombināciju skaits,* C_n^m	p^m *= 0,33m*	q^{n-m} *= 0,67n-m*	$P_n(m)$
0	*1*	*1*	*0,131687*	*0,131687*
1	*5*	*0,333333*	*0,197531*	*0,329218*
2	*10*	*0,111111*	*0,296296*	*0,329218*
3	*10*	*0,037037*	*0,444444*	*0,164609*
4	*5*	*0,012346*	*0,666667*	*0,041152*
5	*1*	*0,004115*	*1*	*0,004115*
				1

$$\mu = n * p = 5 * 0.33 = 1{,}67; \ \sigma = \sqrt{n * p * q} = \sqrt{5 * 0{,}33 * 0{,}67} = 1{,}05$$

Kā redzams varbūtību sadalījumā, tad atsevišķu notikumu varbūtība samazinās, vidējā gaidāmā vērtība (matemātiskā cerība) palielinās, tāpat arī sadalījuma izkliedes plašums palielinās.

4.4. tabulā ir doti varbūtību sadalījumi 10 mēģinājumiem.

4.4. tabula

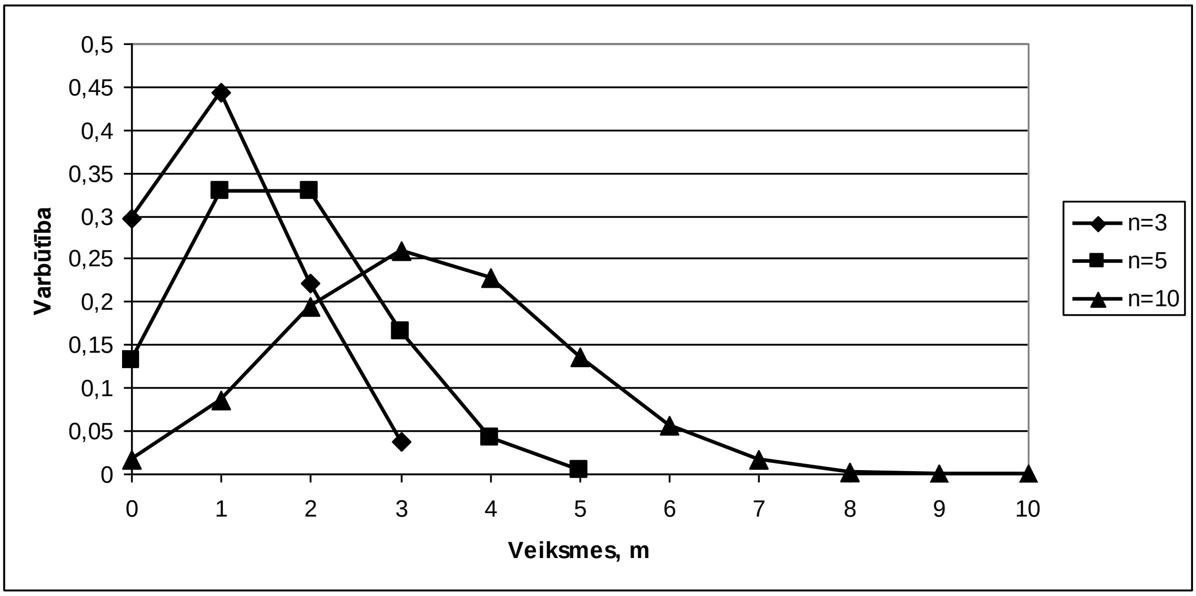
Binomiālā sadalījuma varbūtību aprēķins 4.4. piemēram (n=10)

Veiksmju skaits, m	*Kombināciju skaits,* C_n^m	p^m *= 0,33m*	q^{n-m} *= 0,67n-m*	$P_n(m)$
0	*1*	*1*	*0,017342*	*0,017342*
1	*10*	*0,333333*	*0,026012*	*0,086708*
2	*45*	*0,111111*	*0,039018*	*0,195092*
3	*120*	*0,037037*	*0,058528*	*0,260123*
4	*210*	*0,012346*	*0,087791*	*0,227608*
5	*252*	*0,004115*	*0,131687*	*0,136565*
6	*210*	*0,001372*	*0,197531*	*0,056902*
7	*120*	*0,000457*	*0,296296*	*0,016258*
8	*45*	*0,000152*	*0,444444*	*0,003048*
9	*10*	*0,000051*	*0,666667*	*0,000339*
10	*1*	*0,000017*	*1*	*0,000017*
				1

$\mu = n * p = 10 * 0.33 = 3{,}33$; $\sigma = \sqrt{n * p * q} = \sqrt{10 * 0{,}33 * 0{,}67} = 1{,}49$

Kā redzams tabulā un tālākajos aprēķinos, vidējā vērtība un datu izkliede palielinās, bet atsevišķu notikumu notikšanas varbūtības samazinās. Varbūtība vinnēt visos 10 mēģinājumos ir ļoti niecīga, tāpat arī varbūtība vinnēt 9, 8 vai 7 pudeles ir ļoti maza.

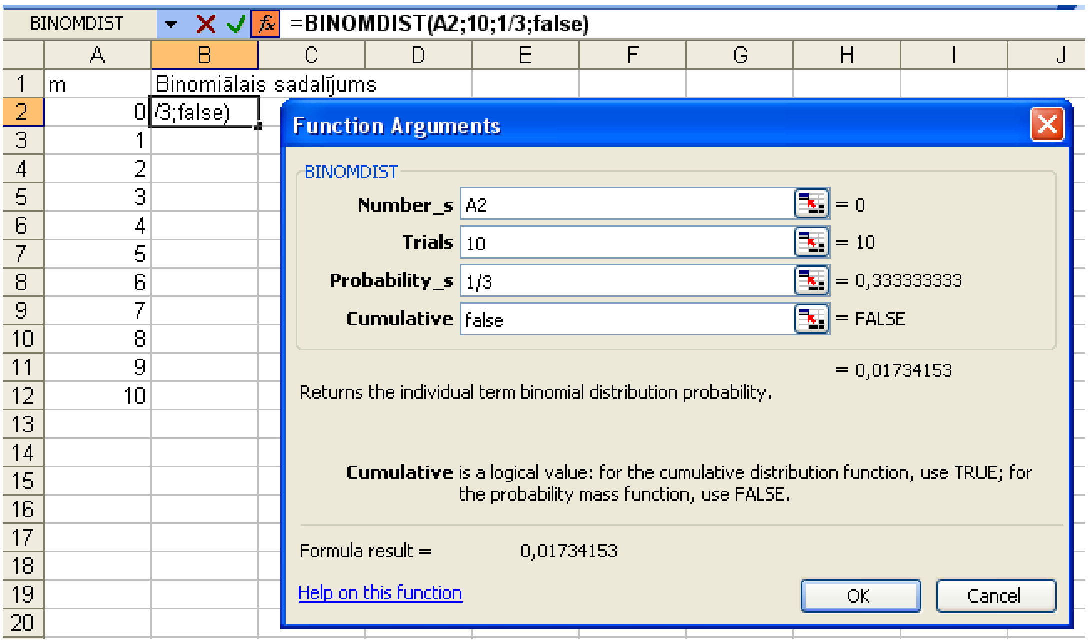
4. 6. attēlā ir parādīti visi trīs varbūtību sadalījumu poligoni.



4.6. attēls. **Binomiālie sadalījumi 4.4. piemēram**

4.6. attēlā redzama iepriekšējo secinājumu vizualizācija – palielinoties mēģinājumu skaitam, sadalījums kļūst zemāks (atsevišķu notikumu notikšanas varbūtība samazinās), plašāks (dispersija palielinās) un nolaidenāks.

Binomiālā sadalījuma varbūtības var aprēķināt ar programmas *Excel* statistisko funkciju "BINOMDIST". 4.7. attēlā ir parādīts *Excel* darba lapas fragments ar dialoga logu binomiālā sadalījuma varbūtību aprēķināšanai.



4.7. attēls. **Programmas *Excel* darba lapas fragments ar binomiālā sadalījuma varbūtību aprēķināšanas dialoga logu**

Ja kopējais vienību skaits kopā nav liels, izņemtās vienības ietekmē atlikušo varbūtību sadalījumu, tad lieto **hiperģeometrisko sadalījumu**. Hiperģeometriskā sadalījuma varbūtības var aprēķināt ar 4.25. formulu:

$$P_n(m) = \frac{\dfrac{M!}{m! \times (M-m)!} \times \dfrac{(N-M)!}{(n-m)! \times ((N-M)-(n-m))!}}{\dfrac{N!}{n! \times (N-n)!}}, \qquad (4.25.)$$

kur

$P_n(m)$ – varbūtība, ka būs m veiksmes n mēģinājumos;

M – kopējais veiksmju skaits ģenerālkopā;

N – ģenerālkopas vienību skaits;

m – veiksmju skaits izlasē;

n – izlases apjoms, vienības.[1]

Tā kā varbūtību sadalījuma izpratnei pietiks ar binomiālo sadalījumu, tad hiperģeometriskais sadalījums sīkāk netiks analizēts.

Daudzu mēģinājumu gadījumā binomiālo sadalījumu aproksimē (tuvināti pielīdzina – aizvieto) ar normālo sadalījumu.

[1] Дэвид М. Левин [и др.]; [Пер. с англ. Д. А. Клюшина]. (2005). *Статистика для менеджеров с использованием Microsoft Excel.* (4-е изд.) М. [и др.]: Вильямс, (ГПП Печ. Двор). 316 с.

Ja p vai q vērtības ir nelielas (<0,1), tad pie liela novērojumu skaita to aproksimē ar Puasona sadalījumu.

4.2.3. Puasona sadalījums

Puasona sadalījums ir reto notikumu sadalījums. To izmanto, ja interesē, cik reizes sagaidāms kāds notikums noteiktā iespējamo rezultātu apgabalā. Parasti tas ir ierobežots ar laiku un teritoriju.

Puasona sadalījumam ir šādas īpašības:

1) interesējošā iznākuma varbūtība ir vienāda visiem iespējamo iznākumu apgabaliem;
2) viena apgabala iznākums neietekmē cita iespējamo rezultātu apgabalu iznākumu;
3) ja iespējamo apgabalu iznākumu apgabalu samazina, tad varbūtība, ka interesējošais rezultāts iestāsies vairāk nekā vienu reizi, tiecas uz nulli.

Puasona sadalījumam ir viens parametrs, kurš reizē ir gan matemātiskā cerība, gan dispersija, un to apzīmē ar grieķu burtu λ (lambda).[1,2]

4.5. piemērs. Latvijā no 2003. gada līdz 2012. gadam vidēji ir dzimuši 3,2 trīnīši gadā. Jāaprēķina varbūtību sadalījums, ka gadā nepiedzims nevieni trīnīši, būs vieni, divi, trīs, četri, pieci un vairāk trīnīšu gadā.

Formula Puasona sadalījuma varbūtību aprēķināšanai ir:

$$P(m) = \frac{e^{-\lambda} * \lambda_m}{m!} , \qquad (4.26.)$$

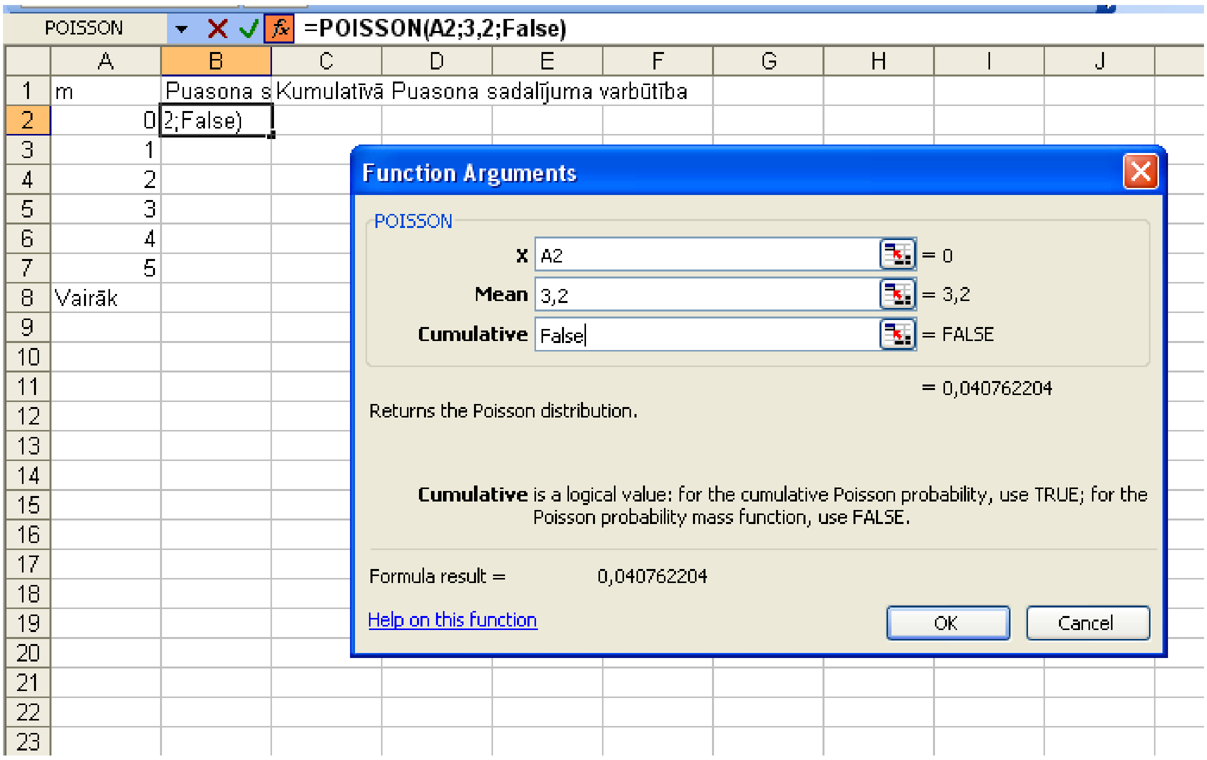
kur *P(m)* – varbūtība, ka notiks m interesējošie notikumi;
e – naturālo logaritmu bāze (2,718);
λ – Puasona sadalījuma parametrs, parasti izmanto empīrisko novērojumu vidējo vērtību.

Puasona sadalījuma vērtības var atrast ar statistisko funkciju POISSON. 4.8. attēlā ir parādīts *Excel* darba lapas fragments ar dialoga logu 4.5. piemēra risināšanai. 4.5. tabulā ir aprēķinātas Puasona sadalījuma varbūtības.

Tā kā Puasona sadalījums ir neierobežots, tad visiem iespējamajiem rezultātiem aprēķināt varbūtības nav iespējams, un, jo lielāks m, jo niecīgāka ir šī notikuma iestāšanās varbūtība. Zinot to, ka pilnas notikumu kopas varbūtību summa ir 1, tad atlikušo notikumu kopējās notikšanas varbūtību aprēķina, no 1 atņemot aprēķināto gadījumu iestāšanās varbūtības.

---

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis*. Rīga: Datorzinību Centrs. 72. lpp.
[2] Дэвид М. Левин [и др.]; [Пер. с англ. Д. А. Клюшина]. (2005). *Статистика для менеджеров с использованием Microsoft Excel.* (4-е изд.) М. [и др.]: Вильямс, (ГПП Печ. Двор). 320 с.



4.8. attēls. ***Excel*** **darba lapas fragments ar dialoga logu Puasona sadalījuma varbūtības aprēķināšanai**

Diferenciālā funkcija ir kāda notikuma notikšanas varbūtība, bet **integrālā funkcija** parāda varbūtību, ka pazīmes vērtība būs mazāka nekā uzdota. Integrālā funkcija pēc savas būtības atbilst empīriskā sadalījuma kumulatīvajiem biežumiem.

4.5. tabula

Puasona sadalījuma diferenciālās un integrālās (kumulatīvās) varbūtības

m	Puasona sadalījuma varbūtība	Kumulatīvā Puasona sadalījuma varbūtība
0	0,040762	0,040762
1	0,130439	0,171201
2	0,208702	0,379904
3	0,222616	0,60252
4	0,178093	0,780613
5	0,113979	0,894592
Vairāk	0,105408	1

Ja salīdzina iegūtās varbūtības ar faktisko sadalījumu, tad redzams, ka rezultāti ir atšķirīgi, bet jāatgādina, ka statistiskās likumsakarības izpaužas tikai masu novērojumos. Ja spēļu kauliņš tiks ripināts 10 reizes, tad arī empīriskie novērojami var būtiski atšķirties no sagaidāmajām varbūtībām.

Puasona sadalījumu var lietot daudzās situācijās, ja samazina iespējamo iznākumu apgabalu, piemēram, analizē, cik klientu atnāk uz iestādi vienas minūtes laikā. Tie būs tikai daži (0, 1, 2 utt.), bet atkārtojumu būs tik daudz, ka Puasona sadalījuma varbūtības var pārbaudīt ar empīrisko novērojumu relatīvajām frekvencēm.

Tomēr dabas un sociāli ekonomiskajos pētījumos dominē nepārtraukti sadalījumi, un no tiem visbiežāk sastopams ir normālais sadalījums.

4.2.4. Normālais sadalījums

Ar normalitāti saprot biežāk sastopamās vērtības, un šo apzīmējumu piešķir pēc pieredzes. Katram cilvēkam normalitātes diapazons var atšķirties, bet kopējais ir tas, ka gan mazākas, gan lielākas varianšu vērtības ir sastopamas retāk par to, kas tiek uzskatīts par normu. Piemēram, cilvēki ar vidējo svaru ir sastopami biežāk nekā kārnie un resnie. Turklāt, jo tālāk no vidējās vērtības, jo retāk ir sastopami šādi novērojumi.

Normālā sadalījuma īpašības:

1) grafiks ir simetrisks ar zvanveida formu;
2) matemātiskā cerība (vidējā vērtība), mediāna un moda ir identiska;
3) vairums novērojumu ir 4/3 standartnoviržu diapazonā – 2/3 pa kreisi no vidējās vērtības un 2/3 pa labi, tā ir starpkvartiļu amplitūda (Q_3–Q_1);
4) normālā sadalījuma vērtības nav ierobežotas (no $-\infty$līdz $+\infty$)

Normālo sadalījumu ir neierobežoti daudz, un tos nosaka to parametri (μ – matemātiskā cerība un σ – standartnovirze).

Sadalījuma raksturošanai izmanto funkcijas – diferenciālo jeb blīvuma funkciju (4.27. formula) un integrālo funkciju (4.30. formula), kura parāda varbūtību, ka variantes vērtība būs mazāka nekā uzdotā.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (4.27.)$$

kur μ – sadalījuma matemātiskā cerība (vidējā vērtība);
σ – sadalījuma standartnovirze;
π – konstante, 3,14…;
e – konstante 2,718…

Tā kā normālo sadalījumu ir bezgala daudz, tad parasti izmanto standartizēta normālā sadalījuma funkcijas, kur μ = 0, bet σ = 1.

Standartizēta normālā sadalījuma diferenciālā funkcija izskatās šādi:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}, \qquad (4.28.)$$

kur z – x vērtības attālums no μ (sadalījuma matemātiskās cerības), izteikts standartnovirzēs.

Standartizācijas formula:

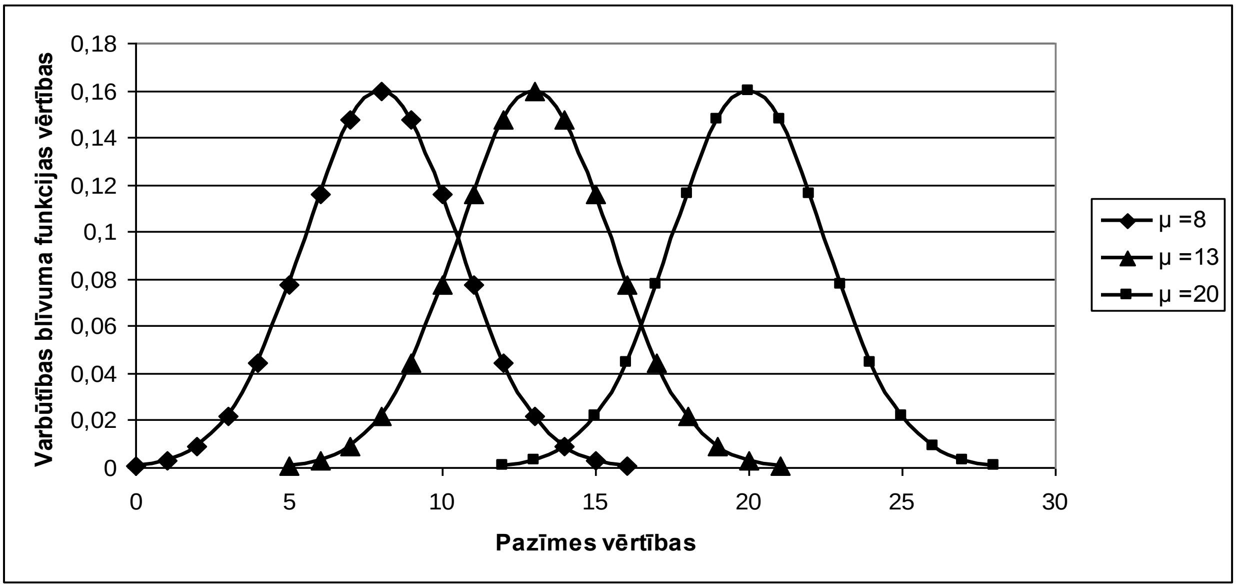
$$z = \frac{x - \mu}{\sigma} \tag{4.29.}$$

Standartizēta integrālā funkcija ir:

$$F(z) = \frac{1}{2\pi} * \int_{-\infty}^{z} e^{-\frac{z^2}{2}} \tag{4.30.}$$

Standartizētās normālā sadalījuma funkcijas ir tabulētas, praktiskajā darbā vairāk izmanto integrālās funkcijas vērtības, jo normālā sadalījuma varbūtības var saistīt tikai ar vērtību intervāliem, nevis ar atsevišķām vērtībām, jo tās tiecas uz nulli, pie nosacījuma, ka iespējamo varianšu skaits tiecas uz bezgalību.[1,2,3]

4.9. attēlā ir parādīti vairāki normālie sadalījumi, kuriem ir vienāda dispersija, bet atšķirīgi vidējie (μ – matemātiskā cerība).



4.9. attēls. **Normālie sadalījumi ar σ = 2,5 un atšķirīgiem μ**

Kā redzams, līknes ir identiskas, bet nobīdītas pa x asi.

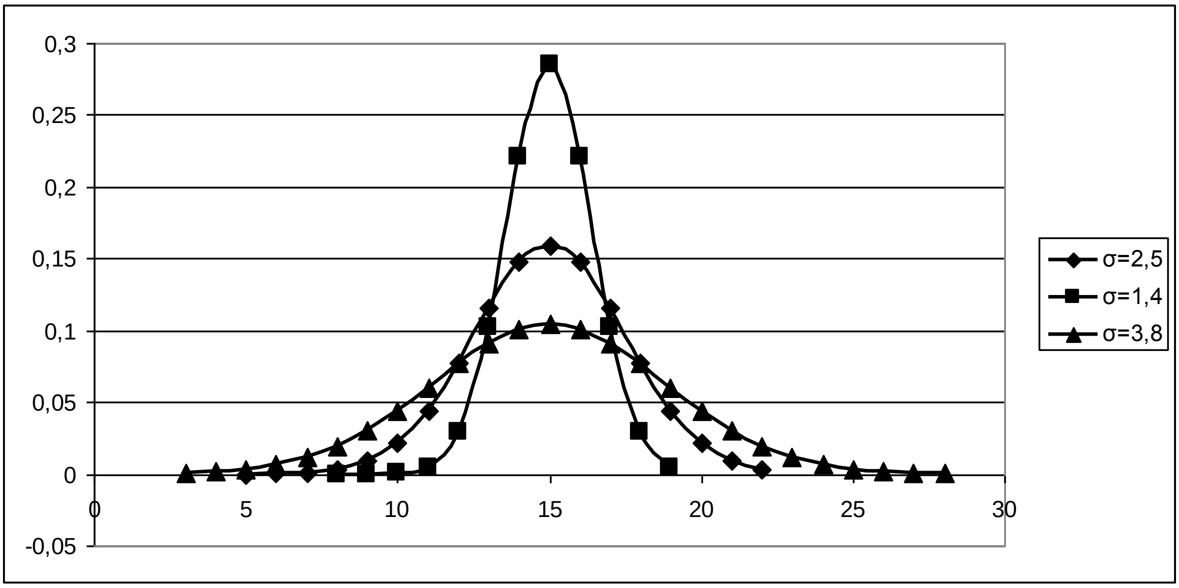
4.10. attēlā ir parādīti normālie sadalījumi, ja μ (aritmētiskie vidējie) ir vienādi, bet atšķiras dispersijas.

---

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 78. lpp.

[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 264.

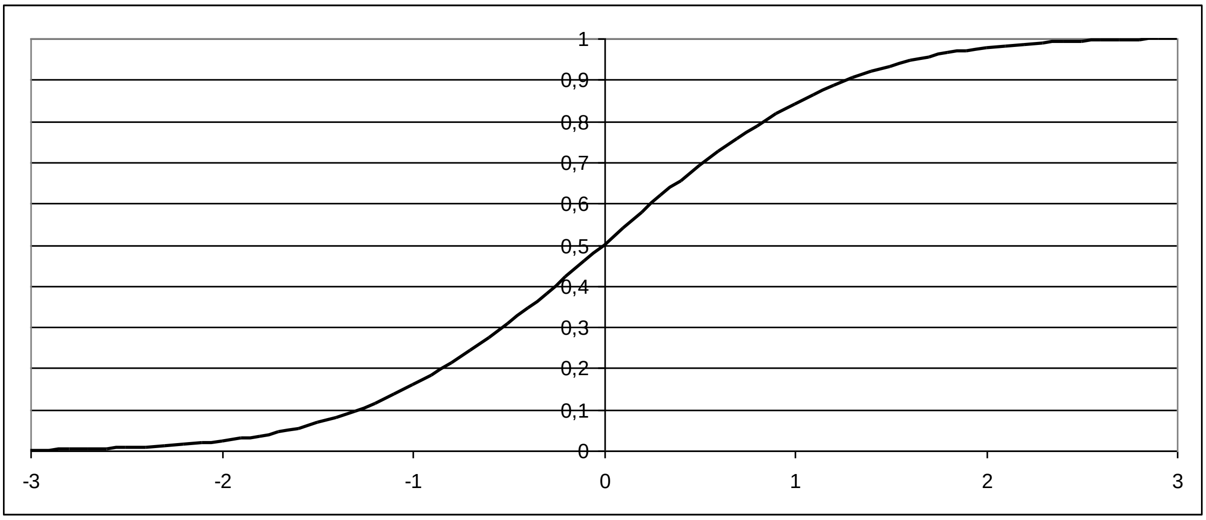
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 95. lpp.



4.10. attēls. **Normālie sadalījumi ar μ = 15 un atšķirīgām standartnovirzēm**

Līknes forma ir vienāda, bet, samazinoties σ, tā tiek "saspiesta", palielinoties – "izplesta". Rēķinot līkņu pārliekuma punkti standartnovirzēs no aritmētiskā vidējā ir vienādā attālumā.

4.11. attēlā ir parādīta standartizēta normālā sadalījuma ($\mu = 0$, $\sigma = 1$) integrālā funkcija.



4.11. attēls. **Standartizēta normālā sadalījuma integrālās funkcijas grafiks**

Jebkuram empīriskajam sadalījumam var aprēķināt atbilstošu normālo sadalījumu (teorētisko).

4.6. piemērs. Aprēķināt teorētiskā sadalījuma frekvences tirdzniecības aģentu apgrozījuma 2.1.piemēram, grupējuma informācija dota 2.2. tabulā.

Gradācijas klases, tūkstošos €	Kopā
6 – 14	9
14 – 22	22
22 – 30	12
30 – 38	2
38 – 46	2
46 – 54	5
Kopā	52

Aprēķinātais vidējais apgrozījums ir 23,077 tūkst. € (3.3. tabula) un standartnovirze 11,425 tūkst. € (3.8. tabula).

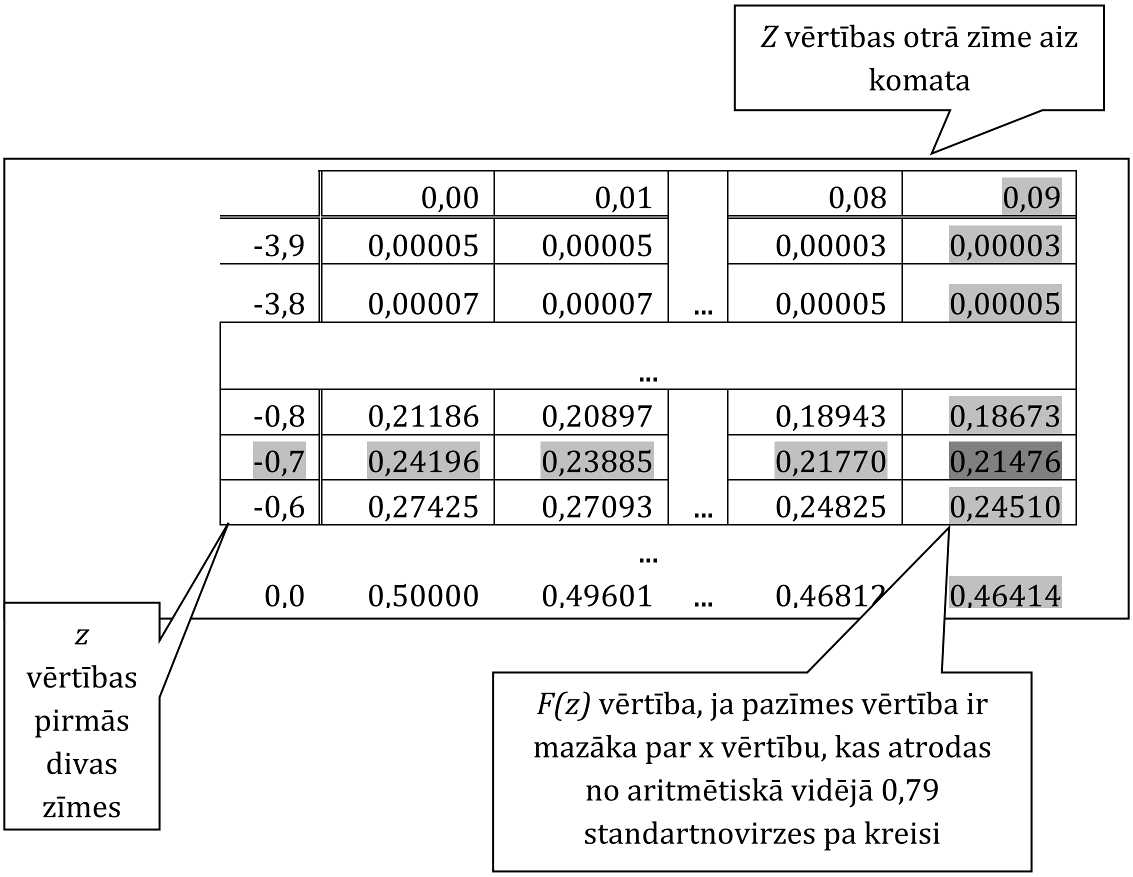
Uzdevuma izpildes kārtība[1]

1. Tabulā ieraksta intervālu robežas, nenorādot pirmā intervāla apakšējo robežu un pēdējā intervāla augšējo robežu (normālais sadalījums ir neierobežots), šīs robežas nestandartizē, un integrālās funkcijas atbilstošās vērtības nemeklē tabulā.
2. Standartizē intervālu robežas (4.29. formula) – aprēķina, cik standartnoviržu attālumā tās atrodas no aritmētiskā vidējā (*z* vērtība pirmā intervāla augšējai robežai ir šāda:

$$z = \frac{x - \mu}{\sigma} = \frac{14 - 23{,}077}{11{,}425} = -0{,}79\,.$$

3. Normālā sadalījuma integrālās funkcijas tabulā (pielikums) atrod atbilstošās vērtības. Šo vērtību noteikšana ir parādīta 4.12. attēlā. Integrālās funkcijas vērtība rāda normālā sadalījuma varbūtību, ka pazīmes vērtība būs mazāka par uzdoto vērtību (kumulatīvā varbūtība).
4. Aprēķina normālā sadalījuma intervāla varbūtību, no intervāla augšējās robežas integrālās funkcijas vērtības atņemot attiecīgā intervāla zemākās robežas integrālās funkcijas vērtību.
5. Destandartizē iegūtās varbūtības, pārrēķinot atbilstošā empīriskā sadalījuma absolūtajās vienībās – intervāla normālā sadalījuma varbūtību sareizina ar attiecīgā empīriskā sadalījuma kopējo novērojumu skaitu. 4.6. tabulā ir dota aizpildīta normālā sadalījuma aprēķināšanas tabula.

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 94. lpp.



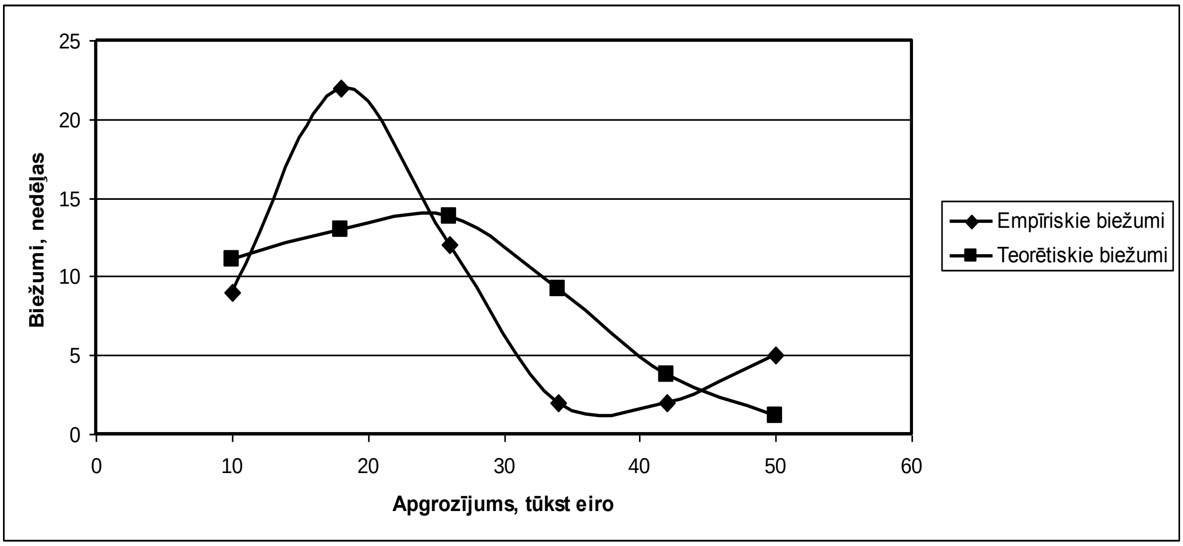
4.12. attēls. **Integrālās funkcijas vērtības noteikšana pēc vērtību tabulas**

5.5. tabula

Normālā sadalījuma biežumu aprēķināšana tirdzniecības aģentu apgrozījuma 4.6.piemēram (sākotnējie dati 2.1. piemērā)

Intervālu robežas		Intervālu robežu z vērtības		Intervālu robežu F(z) vērtības		Normālā sadalījuma biežumi	Destandartizētās teorētiskā intervāla vērtības
zemākā	augstākā	z_z	z_a	$F(z_z)$	$F(z_a)$	$f = F(z_a) - F(z_z)$	$f = f^* \Sigma f_i$
-∞	14	-∞	-0,79	0	0,21346	0,21346	11,10
14	22	-0,79	-0,09	0,21346	0,46245	0,24899	12,95
22	30	-0,09	0,61	0,46245	0,72773	0,26528	13,79
30	38	0,61	1,31	0,72773	0,90425	0,17653	9,18
38	46	1,31	2,01	0,90425	0,97759	0,07334	3,81
46	∞	2,01	∞	0,97759	1	0,02241	1,17
						1	52

6. Grafiski attēlo empīrisko un tam atbilstošo teorētisko sadalījumu, vizuāli novērtē to līdzību.



4.13. attēls. **Empīriskā un tam atbilstošā teorētiskā sadalījuma līkne tirdzniecības aģentu apgrozījuma piemēram** (4.6. piemēra dati, sākotnējā informācija dota 2.1. piemērā)

Vizuāli izskatās, ka attēli ir pietiekami atšķirīgi, taču vienu un to pašu informāciju dažādi pētnieki vizuāli var novērtēt atšķirīgi. Pārliecību ar noteiktu varbūtību par to, ka empīriskais sadalījums atbilst normālajam sadalījumam, var dot hipotēžu pārbaude, kas tiks analizēta nākamajā nodaļā.

Normālā sadalījuma integrālās funkcijas vērtības izmanto divu uzdevumu tipu risināšanai:

1) jānosaka varbūtība, ka pazīmes vērtības būs mazākas par kādu noteiktu robežvērtību (lielākas, atradīsies noteiktā intervālā vai atradīsies ārpus tā). Aprēķinu metodika ir tāda pati kā, aprēķinot varbūtības 4.6. piemērā;
2) ar noteiktu varbūtību jāatrod pazīmes robežvērtības.[1,2]

4.2.5. Logaritmiski normālais sadalījums

Ekonomikā ir sastopami sadalījumi ar labo asimetriju. Tāds sadalījums ir raksturīgs iedzīvotāju ienākumu sadalījumam. Lai cilvēks varētu eksistēt, viņam ir nepieciešams kāds ieņēmumu minimums, kuru valsts garantē ar minimālās darba algas noteikšanu, garantēto minimālo ienākumu (GMI), dažādiem pabalstiem, tādēļ sadalījuma kreisais zars apraujas strauji. Vairums cilvēku, kuri strādā, saņem nelielas algas, vidējās algas saņēmēju jau ir mazāk, vidēji lielas, lielas un ļoti lielas algas (vai dividendes, citus ienākumus) saņem arvien mazāk cilvēku.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 100. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 269.

Statistikā ir labi izstrādāta metodika, kas balstās uz teorētisko normālo sadalījumu. Daudzos gadījumos, kad empīriskais sadalījums neatbilst normālajam sadalījumam, to normalizē, tas ir, veic matemātiskas manipulācijas, kuru rezultātā sākotnējie dati tiek pārveidoti datos, kas variē atbilstoši normālā sadalījuma likumam. Šajā gadījumā sadalījuma nosaukums jau ietver normalizācijas principu – sākotnējos datus logaritmējot, tiek iegūts sadalījums, kas vairs nav asimetrisks.[1,2,3]

Logaritmiski normālā sadalījuma varbūtības var aprēķināt ar *Excel* funkciju **„LOGNORMDIST"**.

4.2.6. Eksponenciālais sadalījums

Eksponenciālais sadalījums arī ir pozitīvi asimetrisks sadalījums, bet atšķirībā no logaritmiski normālā sadalījuma sākas no nulles. Ar šādu sadalījumu var aprakstīt gaidīšanas laiku dažādās iestādēs (restorānos u.tml.). Atnākušo klientu parasti apkalpo uzreiz, ja klientu ir vairāk, var nākties pagaidīt. Jo vairāk klientu un ilgāks to gaidīšanas laiks, jo mazāka ir varbūtība, ka tas notiks.

Eksponenciālajam sadalījumam ir viens parametrs, kuru apzīmē ar λ (lambda), un tas ir vidējais klientu skaits kādā laika periodā, piemēram, minūtē.

$1/\lambda$ ir vidējais laiks starp diviem pasūtījumiem.

Eksponenciālā sadalījuma varbūtību aprēķina pēc 4.31. formulas:

$$P(x < X) = 1 - e^{-\lambda x}, \qquad (4.31.)$$

kur $P(x < X)$ – varbūtība, ka pasūtījums būs ātrāk nekā *X* laikā;
e – naturālo logaritmu bāze, konstante (2,718);
λ – vidējais pasūtījumu skaits laika vienībā.[4]

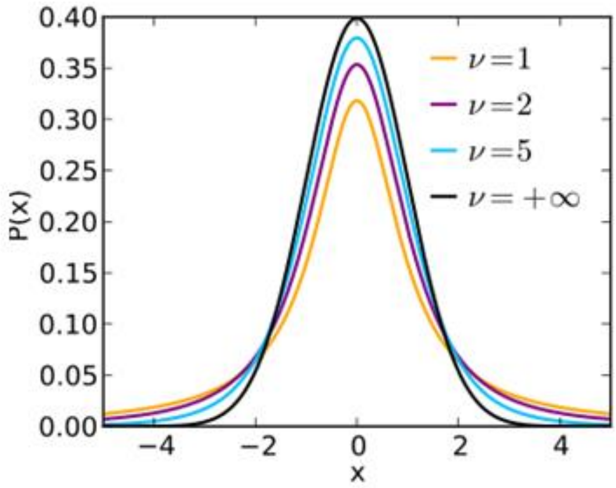
Eksponenciālā sadalījuma varbūtības var aprēķināt ar *Excel* funkciju **„EXPONDIST"**.

4.2.7. Stjūdenta sadalījums

Angļu matemātiķis un kvalitātes kontrolieris Ginesa alus darītavā Gosets ievēroja, ka izlase, kura veidota no normāli sadalītas ģenerālkopas, arī veido

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 113. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 90. lpp.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 96. lpp.
[4] Дэвид М. Левин [и др.] ; [Пер. с англ. Д. А. Клюшина]. (2005). *Статистика для менеджеров с использованием Microsoft Excel.* (4-е изд.) М. [и др.]: Вильямс, (ГПП Печ. Двор). 382 с.

normālajam līdzīgu sadalījumu, kurš tomēr atšķiras ar varbūtību sadalījumu, jo vidējām vērtībām ir lielāka varbūtība nokļūt izlasē. Viņam par godu šo sadalījumu nosauca viņa pseidonīma vārdā par Stjūdenta (sauc arī par *t* sadalījumu) sadalījumu. Šī sadalījuma varbūtības plaši izmanto statistisko hipotēžu pārbaudē. Īpaši tas ir svarīgs nelielām izlasēm (n<30). Ja novērojumu ir vairāk, tad Stjūdenta sadalījuma varbūtības ir aptuveni vienādas ar normālā sadalījuma varbūtībām. 4.14. attēlā ir parādīta normālā un vairāku Stjūdenta sadalījumu proporcionālās atšķirības.



4.14. attēls. **Normālais sadalījums υ (brīvības pakāpes) ir bezgalība un Stjūdenta sadalījumi, ja υ ir 1, 2 un 5 brīvības pakāpes**

Brīvības pakāpes ir novērojumu skaits mīnus saistošo nosacījumu skaits. Statistikā tās lieto, kad ir darīšana ar izlases novērojumiem.

Jo vairāk novērojumu, jo tuvāk normālajam sadalījumam atrodas atbilstošais Stjūdenta sadalījums.[1]

Teorētiskie sadalījumi ir pamats secinājumiem secinošajā statistikā, kas balstās uz izlases novērojumu vispārināšanu un attiecināšanu uz visu ģenerālkopu.

4.3. Izlases metode

Visas interesējošās kopas vienības sauc par **ģenerālkopu**.[2] Tām ir raksturīgs vienveidīgu pazīmju kopums, pēc kurām identificē šo kopu, piemēram, Rēzeknes Tehnoloģiju akadēmijas studenti, rūpnīcā saražotās zāģripas, Latvijas graudaugu sējumu platības u.tml. Visi iepriekš nosauktie piemēri ir atsevišķas ģenerālkopas. Ja apseko (pēta, mēra) visas kopas vienības, tad to sauc par **pilno statistisko novērošanu**. Taču ļoti bieži secinājumu

[1] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 413.
[2] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem*. Rīga: Zvaigzne. 9. lpp.

izdarīšanai netiek izpētītas visas kopas vienības, bet tikai kāda daļa, tad to sauc par **nepilno statistisko novērošanu**. Ja secinājumus izdara pēc tām kopas vienībām, kuras ir vieglāk pieejamas, un neveic pētāmo kopas vienību mērķtiecīgu, zinātniski pamatotu atlasi, tad to sauc par **daļējo statistisko novērošanu.**[1,2,3] Atlasītās kopas vienības var būt ar kādām kopējām iezīmēm un labi nepārstāvēt (nereprezentēt) visu ģenerālkopu. Piemēram, studenti saņēmuši zinātniski pētniecisko uzdevumu noskaidrot RTA studentu brīvā laika pavadīšanas paradumus. Informāciju nolemts iegūt aptaujā. Tā kā anketētāji labprāt apmeklē naktsklubus, tad viņi aptaujā naktsklubā sastaptos studentus. Iegūtie rezultāti slikti pārstāvēs visu ģenerālkopu, jo novērojamajās vienībās nebūs iekļuvuši tie studenti, kuri brīvo laiku pavada mājās pie televizora, datora, darinot rokdarbus, lasot grāmatas u.tml. Daļējo statistisko novērošanu izmanto praktiskajā uzņēmējdarbības vadīšanā (nepretendējot uz pētījuma zinātniskumu). Ja loģiskā analīze liecina, ka pētāmajā kopā iekļuvušās vienības pietiekami labi reprezentē visu ģenerālkopu, tad šādus rezultātus var izmantot.

Zinātniski pamatoti atlasītu pētāmo kopu sauc par **izlasi** (paraugkopu). Galvenais princips, veidojot izlasi, ir nodrošināt vienādu iespēju nokļūt izlasē jebkurai kopas vienībai. Tas nozīmē, ka izlasē vienības nokļūst nejauši, piemēram, izlozējot.

4.3.1. Izlases metodes priekšrocības

1. Pētījumu var veikt lētāk. Par pilno statistisko novērošanu var uzskatīt vēlēšanas. To sarīkošana ir dārga, bet iedzīvotāju politiskās simpātijas un to izmaiņas regulāri nosaka, veicot izlases pētījumus – socioloģiskās aptaujas.
2. Pilno statistisko novērošanu dažreiz nav iespējams veikt, piemēram, kvalitātes pētījumos, kuru rezultātā pētāmās vienības tiek sabojātas. Piemēram, novērtējot ķieģeļu izturības rādītājus, tos speciālās iekārtās pakļauj slodzēm, kas tos sabojā un reģistrē to slodzes līmeni, kuru ķieģelis vairs neiztur. Ja šādi pārbaudītu visus ķieģeļus, tad nebūtu, ko pārdot. Vienīgā iespēja ir pārbaudīt reprezentablu izlasi un iegūtos rezultātus attiecināt uz visu ģenerālkopu.
3. Pētījumu var veikt ātrāk, jo jāapseko mazāk vienību. Tas ir ļoti būtiski uzņēmējdarbības lēmumu pieņemšanai, kad ir svarīgi, lai nepieciešamā informācija ir pēc iespējas ātrāk.
4. Tā kā pētījums ir lētāks un ātrāks, var paplašināt pētījumu programmu.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 116. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 9.
[3] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 14. lpp.

5. Var piesaistīt kvalificētāku darbaspēku, kā rezultātā var samazināt kopējo pētījuma kļūdu.[1,2,3]

Pēdējā priekšrocībā ir iekļauta pretruna, jo izlase ir daļa no visas ģenerālkopas un vienmēr būs kāda neatbilstība starp šīm kopām. Šo neatbilstību sauc par reprezentācijas kļūdu, un tās nav pilnajā statistiskajā novērošanā. Taču statistiskajā pētījumā var būt arī citi kļūdu veidi.

4.3.2. Kļūdu veidi statistiskajā novērošanā

1. **Novērošanas kļūdas**. Ekonomiskajos pētījumos bieži izmanto aptaujas, ko var uzskatīt par cilvēku mērīšanu. Cilvēki bieži atbild nevis, kā viņi domā vai rīkotos attiecīgajā situācijā, bet gan tā, kā uzskata, kā viņiem būtu jāatbild. Novērošanas kļūdas var būt arī, ja ir darbības ar nedzīviem priekšmetiem, piemēram, mēraparāti kvalitātes pētījumam nav verificēti, vai ir bijuši salūzuši un pēc remonta tie nav atkārtoti pārbaudīti. Varbūt cilvēks, kurš strādā ar mēraparātiem, kaut ko neapzināti vai apzināti nedara pareizi.
2. **Reģistrācijas kļūdas**. Ja rezultātus reģistrē pierakstot, tad ir jācenšas skaitļi pierakstīt tā, lai nevar sajaukt 0 ar 6 vai 4 ar 11. Dažreiz cipari piereģistrējot tiek sajaukti vietām, decimālais komats ielikts nevietā, skaitļa beigās tiek pierakstīts par daudz vai par maz nuļļu. Ekstrēmās variantes (lielākās un mazākās) analizē atsevišķi ar hipotēzi par to izcelšanos. Ja ir pamats domāt, ka tās ir reģistrācijas kļūdas, tad aprēķinos cenšas izvairīties no šādām kļūdām, izslēdzot no pētījuma vienu līdz trim variantēm no ranžētās variācijas rindas abiem galiem. Ja novērojumu ir nedaudz, tad atmet pa vienai vērtībai no katra gala, ja daudz, tad var izslēgt vairāk novērojumu.
3. **Apkopošanas kļūdas.** Mūsdienās datu apstrāde tiek veikta ar datoru, līdz ar to tiek izslēgta aritmētisko kļūdu iespējamība, kam tika pievērsta īpaša uzmanība pirms datoru laikmetā. Visbiežāk pētnieki mūsdienās kļūdās, izvēloties konkrētajam pētījumam neatbilstošu metodiku, un bieži vien iegūst pilnībā neinterpretējamu rezultātu.
4. **Reprezentācijas kļūdas**. Izlase precīzi neatspoguļo ģenerālkopu, tajā var nokļūt vienības, kuras vidēji ir ar labākiem vai sliktākiem rādītājiem nekā visā kopumā. Reprezentācijas kļūda ir arī daļējai statistiskajai novērošanai, bet to nevar aprēķināt, tur ir svarīga loģiskā analīze, vērtējums, vai pētāmajā kopā nokļuvušās vienības plaši pārstāv visu ģenerālkopu, vai

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 117. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 115.
[3] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 213. lpp.

lielāka varbūtība tur nokļūt ir kādām noteiktām vērtībām. Izlasei reprezentācijas kļūdu var aprēķināt.[1,2,3]

Piesaistot kvalificētu darbaspēku, ir iespējams samazināt pirmos trīs kļūdu veidus un tad, neskatoties uz to, ka izlasei ir reprezentācijas kļūda, kopējā novērojuma precizitāte palielinās.

Izlases metode paredz:
1) pētāmo vienību atlasi tā, lai tās plaši reprezentētu visu ģenerālkopu;
2) izlases statistisko rādītāju aprēķināšanu;
3) izlases reprezentācijas kļūdas aprēķināšanu.[4]

4.3.3. Pētāmo vienību atlases veidi[5,6,7]

1. **Īsti nejaušā jeb gadījumizlase.** Klasiskais risinājums būtu izlasē iekļaujamās vienības noteikt izlozes veidā. Praksē tas ir sarežģīti vai pat neiespējami veikt, ja ģenerālkopa ir ļoti apjomīga. Ja kopas vienības ir numurētas, piemēram, uzņēmumiem ir reģistrācijas numuri, tad pētījumam atlasāmās vienības var noteikt, veicot izlozi ar *Excel* matemātisko funkciju **„RANDBETWEEN".** Gadījumizlase var būt atkārtota, tas nozīmē, ka vienreiz izlozētā vienība izlasē var nokļūt atkārtoti, un neatkārtotā, kad vienreiz iekļautās vienības, pat ja tās tiek atkārtoti ielozētas izlasē, vēlreiz netiek iekļautas. Ja grib būt ļoti korekts savos secinājumos, tad reprezentācijas kļūdas aprēķina atšķirīgi, bet starpība nav liela (ja ģenerālkopa ir liela), tāpēc *Excel* datu analīzē ir iekļauta tikai atkārtotas izlases reprezentācijas kļūdas aprēķinu formula.

2. **Mehāniskā izlase**. Pētāmās vienības ir sakārtotas sarakstā pēc nebūtiskas pazīmes. Tāda nebūtiska pazīme varētu būt cilvēku uzvārda pirmie burti. Saraksts, kas sakārtots alfabētiskā secībā pēc uzvārdiem, vai uzņēmumu reģistra numuri, kuri ir piešķirti hronoloģiskā secībā, varētu būt piemērs sarakstam, kas sakārtots pēc nebūtiskas pazīmes. Vispirms nosaka soli, ar kādu tiks atlasītas pētāmās vienības. Piemēram, ja sarakstā ir 600 vienības un tiek plānots atlasīt 30 vienības izlasei, tad solis būs 600/30 – katrs divdesmitais. Pirmajā solī izlozē sākumvienību (var aizvērtām acīm pagriezt

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 118. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 109.lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 125.
[4] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 215. lpp.
[5] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 119. lpp.
[6] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 125.
[7] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 103. lpp.

pildspalvu un iebakstīt saraksta sākumā – pret kuru numuru trāpa, no tā sāk, var izlozēt ar iepriekšminēto *Excel* funkciju „RANDBETWEEN”). Ja ir „trāpīts” 13 vienībai, tad tālākās atlasa ar soli 20 – 33, 53, 73 utt. līdz saraksta beigām. Reprezentācijas kļūdu nosaka tāpat kā gadījumizlasē.

3. **Ligzdveida izlase.** Dažreiz racionāli ir atlasīt pētāmo vienību grupas – „ligzdas”, un tajās apsekot visas vienības, piemēram, veicot kvalitātes kontroli, mums ir jāpārbauda 100 izstrādājumi, kuri ir iepakoti kastēs pa 20. Nav jāatplēš 100 kastes un no katras nav jāpaņem pa vienam izstrādājumam, bet nejauši izvēlas 5 kastes un tajās pārbauda visus izstrādājumus pēc kārtas. Līdzīgi varam rīkoties, veicot socioloģisko aptauju par daudzdzīvokļu māju iedzīvotāju apmierinātību ar komunālajiem pakalpojumiem – izlozē apsekojamās kāpņu telpas un tajās nointervē visu dzīvokļu iemītniekus pēc kārtas.

4. **Stratificētā, jeb tipoloģiskā izlase.** Ja visu kopu varam sagrupēt apakšgrupās pēc pētījumam būtiskas pazīmes, tad visprecīzākos pētījuma rezultātus dod tieši šī metode. Piemēram, uzņēmumus var sagrupēt pēc to pamatdarbības veida, juridiskās formas, apgrozījuma lieluma u.tml. Tad no katras apakšgrupas, proporcionāli tās lielumam, atlasa pētāmās vienības ar gadījumizlasi vai mehānisko izlasi. Šādos pētījumos gadās, ka grupas pēc vienību skaita ir ļoti atšķirīgas, bet obligāti ir jābūt pārstāvētām visām grupām. Varbūtība nokļūt izlase mazāk pārstāvēto grupu vienībām ir lielāka nekā to grupu vienībām, kurās vienību ir daudz vairāk. Aprēķinot raksturotājus (vidējo, standartnovirzi u. c.), ir jāņem vērā varbūtība, ar kādu vienība nokļūst izlasē, vidējam to var izdarīt, rēķinot ar 4.32. formulu:

$$\mu = \frac{\sum \frac{x_i}{P_i}}{\sum \frac{1}{P_i}}, \qquad (4.32.)$$

kur P_i – varbūtība, ar kādu kopas vienība nonāk izlasē (n_i/N_i = no apakšgrupas atlasīto vienību skaits / kopējais vienību skaits apakšgrupā).

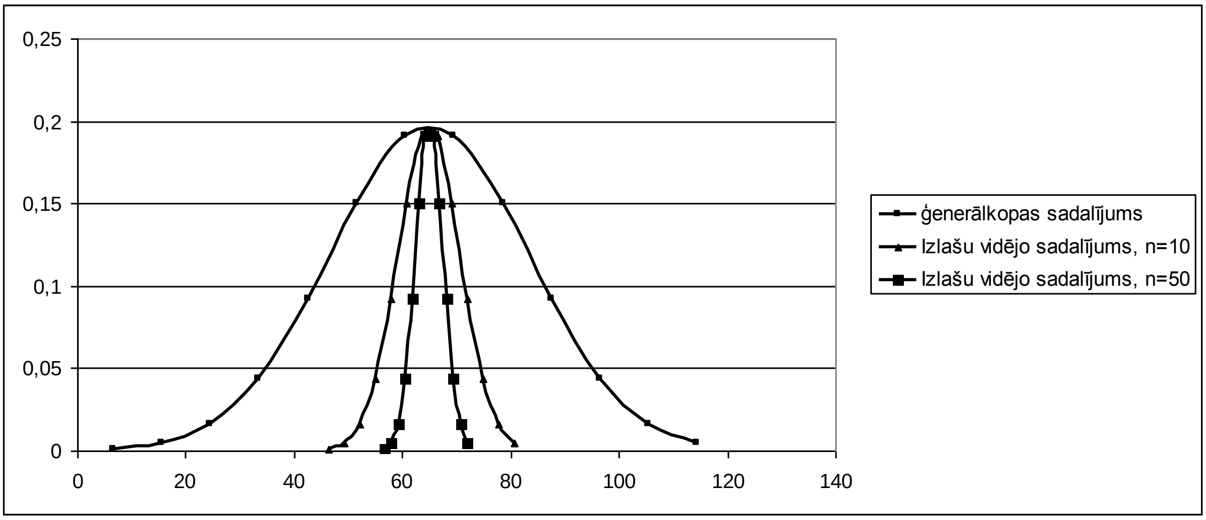
Ja lieto apgriezto lielumu (N_i/n_i), tad iegūst parasto vidējā svērtā aritmētiskā formulu, kur par frekvencēm lieto vienas izlasē nokļuvušās vienības pārstāvēto vienību skaitu. Līdzīgi aprēķini arī jāveic, rēķinot standartnovirzi vai kādu citu statistisko rādītāju.

Ja vienības tiek atlasītas proporcionāli, tad šī formula nav jālieto, var izmantot vienkāršo aritmētisko vidējo.

4.3.4. Ģenerālkopas parametru vērtēšana

Parasti nezināmos ģenerālkopas parametrus (apzīmē ar grieķu alfabēta burtiem, līdzīgi kā teorētisko sadalījumu parametrus) novērtē kā aptuveni vienādus izlases raksturotājiem. Šādu vērtējumu sauc par punktveida, un tas atbilst četrām pamatprasībām – nenobīdīts, konverģējošs, pietiekams un efektīvs. Tas nozīmē, ka vērtējumam nav sistemātiskās kļūdas un bezgala daudzu izlašu vidējo summārās kļūdas tiecas uz nulli.[1,2]

Dažreiz nezināmos ģenerālkopas parametrus novērtē ar intervālu, to dēvē par ticamības intervālu. Ja no ģenerālkopas, kas sadalīta normāli, veido ļoti daudzas izlases ar n novērojumiem, tad izlašu vidējie arī veido normālo sadalījumu ap ģenerālkopas vidējo. Ja izlases apjoms (*n*) ir mazs, tad izlašu vidējie ir plašāk izkliedēti nekā izlasēm ar lielāku novērojumu skaitu.[3,4] 4.15. attēlā ir parādīts ģenerālkopas sadalījums un attiecīgi izveidotu izlašu vidējo sadalījums pie atšķirīgiem n.



4.15. attēls. **Ģenerālkopas (μ=65 un σ=18)un no tās veidoto izlašu sadalījumi, ja izlases apjoms ir 10 un 50 vienības**

Izlašu vidējie veido normālo sadalījumu, un to izkliedi raksturo standartkļūda, apzīmē ar s un indeksu, kurā paskaidro, kādam rādītājam to rēķina. Aritmētiskā vidējā standartkļūdu apzīmē ar $s_{\bar{x}}$.

Gadījumizlasei standartkļūdu aprēķina pēc 4.33. formulas:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}\ , \qquad (4.33.)$$

kur *s* – standartnovirze;
n – novērojumu skaits.[5]

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 133. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 313.
[3] Turpat.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 135. lpp.
[5] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 226. lpp.

Kā redzams, tad standartkļūdas lielumu ietekmē novērotās kopas datu izkliede un izlases lielums. Ja vēlas iegūt precīzākus datus ar mazāku reprezentācijas kļūdu, tad vajag palielināt novērojumu skaitu.

Pētījumā parasti veido vienu izlasi, un pastāv varbūtība, ka tiks atlasītas vienības, kas vairāk pārstāv vai nu lielo, vai mazo vērtību vienības. Rekomendētu plānoto izlases apjomu mehāniski sadalīt divās uz pusi mazākās izlasēs. Ja aprēķinātie rādītāji ir tuvi viens otram, izlases var apvienot un turpināt darbu ar vienu lielāku izlasi. Pieņemot, ka vienas izlases vienības nav saistītas ar otras puses izlases vienībām, tad būtiski samazinās varbūtība, ka abās reizēs tiks atlasītas vienības, kas ģenerālkopā ir sastopamas retāk. Ja starp abām izlasēm ir būtiska atšķirība, tad veido papildus izlasi, lai noskaidrotu patieso ģenerālkopas sadalījumu un parametrus.

$\pm s_{\bar{x}}$, līdzīgi kā standartnovirze visam novērojumam, apraksta apmēram 68% no visu izlašu vidējo sadalījuma varbūtībām.

Parasti vēlas iegūt lielāku pārliecību par datiem. Ekonomikā lieto 95% ticamības intervālus. Tas nozīmē, ka pastāv 5 % varbūtība, ka ģenerālkopas vidējais (matemātiskā cerība) atrodas ārpus noteiktā intervāla. Jāatgādina, ka, metot divus spēļu kauliņus, varbūtība uzmest divus vieniniekus ir mazāka (1/36 jeb aptuveni 2,8 %), bet, ja tā notiek, metot kauliņu vienreiz, šī varbūtība netiek apstrīdēta. Tāpēc jebkurš statistiskais lēmums, kas balstīts uz izlases novērošanu, varbūt kļūdains.

Robežkļūda parāda iespējamo ģenerālkopas vidējā (matemātiskās cerības) intervālu ar noteiktu ticamības līmeni. Robežkļūdu aprēķina pēc 4.34. formulas:

$$\Delta_{\bar{x}} = t_{\alpha,\upsilon} * s_{\bar{x}}, \tag{4.34.}$$

kur $\Delta_{\bar{x}}$ – robežkļūda;

$t_{\alpha,\upsilon}$ – Stjūdenta koeficienta kritiskā vērtība, ko nosaka nozīmības līmenis (α) un brīvības pakāpes (υ).[1,2,3] Nolasa tabulās pielikumā vai atrod ar *Excel* funkciju "TINV".

4.7. piemērs. Ir reģistrētas atzīmes divās studentu grupās, un iegūti šādi rezultāti: $\bar{x}_1 = 6.11$, s_1=1,8, n_1=97 un $\bar{x}_2 = 6{,}82$, s_2=2,2, n_2=39.
Jāaprēķina matemātiskās cerības ticamības intervāls ar nozīmības līmeni 0,05. Jāpārbauda, vai starp uzrādītajiem rezultātiem ir būtiska atšķirība.

[1] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 320. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 135. lpp.
[3] Goša, Z. (2003). *Statistika.* Rīga: SIA "Izglītības soļi". 227. lpp.

Vispirms abām grupām aprēķina aritmētiskā vidējā standartkļūdas un robežkļūdas: $s_{\bar{x}_1} = \frac{1{,}8}{\sqrt{97}} = 0{,}183$ *un* $\Delta_{\bar{x}_1} = 1{,}985 * 0{,}183 = 0{,}363$.

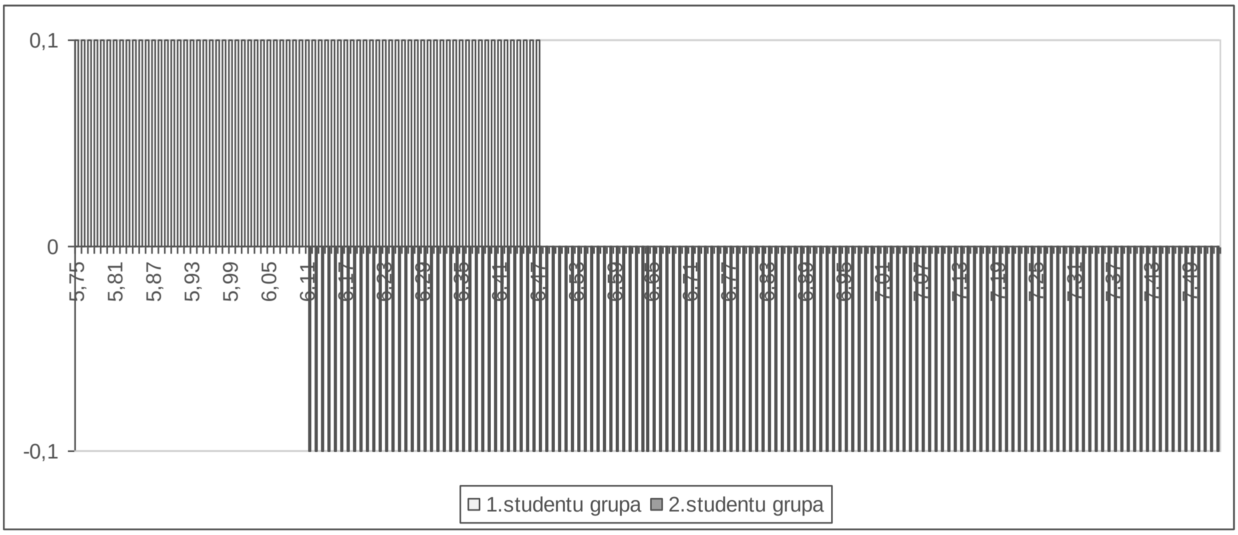
Un otrajai grupai: $s_{\bar{x}_2} = \frac{2{,}2}{\sqrt{39}} = 0{,}352$ *un* $\Delta_{\bar{x}_2} = 2{,}024 * 0{,}352 = 0{,}713$

Ticamības intervālu aprēķina šādi:

$$\bar{x} - \Delta_{\bar{x}} \le \mu \le \bar{x} + \Delta_{\bar{x}}$$

$$6{,}11 - 0{,}363 = 5{,}75 \le \mu_1 \le 6{,}11 + 0{,}363 = 6{,}47$$

Labākai izpratnei ticamības intervālus var atzīmēt uz skaitļu ass (4.16. attēls).



4.16. attēls. **Divu studentu grupu vidējās atzīmes ticamības intervāli**

Kā redzams, intervāli daļēji pārklājas, un ar 95 % pārliecību nevar apgalvot, ka otrās grupas studentiem ir labākas zināšanas par pirmās grupas studentiem. Šādi vērtē arī tad, ja ir apsekotas visas kopas vienības, bet datus vēlas attiecināt uz nākotni – rezultātus vispārina kā studentu zināšanas, nevis konkrētā eksāmena vērtējumus.

Pēdējais piemērs ieskicē jau secinošās statistikas idejas, par kurām plašāk tiks izklāstīts nākamajās nodaļās.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
1	*2*	*3*	*4*
Alternatīvi notikumi	Alternative events	Альтернатив-ные (противо-положные) случаи	Pilnu notikumu kopu veido divi savstarpēji izslēdzoši notikumi
Binomiāls sadalījums	Binomial distribution	Биномиальное распределение	Teorētisks sadalījums, kur ir tikai divas savstarpēji izslēdzošas pazīmes vērtības – „veiksme", „neveiksme".
Daļējā statistiskā novērošana	Determinate sampling		Pētījumā iekļautas vienības, kas pētniekam ir vieglāk pieejamas, izveidotā kopa var labi nereprezentēt visu kopu
Diskrēts sadalījums	Dicrete distribution	Дискретное распределение	Teorētiskais sadalījums, kur pazīmes vērtības var pieņemt tikai konkrētas vērtības, parasti veseli skaitļi
Eksponenciā-lais sadalījums	Exponential distribution	Экспоненциальное распределение	Sadalījums ar labo asimetriju, sākas no nulles, izmanto gaidīšanas rindu sadalījumu aprakstīšanā
Gadījumizlase	Random sample	Случайная выборка	Izlasē iekļaujamās vienības atlasa ar izlozes palīdzību (nejaušo skaitļu tabulas)
Ģenerālkopa	Population	Генеральная совокупность	Visas interesējošās kopas vienības
Izlase (paraugkopa)	Sample	Выборка	Pētījumā iekļautās vienības, daļa no visām ģenerālkopas vienībām
Izmēģinājums	Trial	Опыт (испытание)	Mērķtiecīga apstākļu kompleksa radīšana, kā rezultātā var iestāties notikums
Klasiskā varbūtība	A priori probability	Класическая вероятность	Nosaka, pretstatot notikumam labvēlīgo iespēju skaitu pret visām iespējām, nebalstās uz pētījumiem, bet loģisko analīzi
Ligzdveida (sērijveida) izlase	Claster sample	Серийная выборка	Izlasē atlasa nevis atsevišķas vienības, bet vienību kopas, kurās apseko visas vienības
Logaritmiski normālais sadalījums	Logarithmic normal distribution	Логаритми-чески нормальное распределение	Sadalījums ar labo asimetriju, var aprakstīt iedzīvotāju ienākumu sadalījumu
Nesavienojami notikumi	Mutually exclusive events	Несовмести-мые случаи	Nevar vienlaikus notikt viena novērojuma vai izmēģinājuma rezultātā
Normālais sadalījums	Normal distribution	Нормальное распределение	Visbiežāk dabas un sociāli ekonomiskajos pētījumos sastopamais sadalījumu veids, raksturīgs tas, ka vidējās vērtības ir pārstāvētas vairāk nekā lielās un mazas, apraksta ar zvanveida sadalījuma līkni

1	2	3	4
Notikums	Event	Случай	Jebkurš fakts, kuru var konstatēt novērojuma vai izmēģinājuma rezultātā
Novērojums	Observation	Наблюдение	Noteikta apstākļu kompleksa realizācija kā rezultātā var iestāties notikums, pētnieks nerada apstākļu kompleksu
Parametrs	Parameter	Параметр	Ģenerālkopu raksturojošs rādītājs (vidējais – matemātiskā cerība, standartnovirze u. c.)
Puasona sadalījums	Poisson distribution	Распределение Пуасона	Reto notikumu sadalījums, apraksta diskrēta sadalījuma varbūtības, ja notikumam labvēlīgā varbūtība (p) ir mazs skaitlis
Raksturotājs	Statistic	Статистик	Izlasi raksturojošs rādītājs (vidējais, standartnovirze u. c.)
Savienojami notikumi	Nonmutu-ally exclusive events	Совместимые случаи	Var vienlaikus notikt viena novērojuma vai izmēģinājuma rezultātā
Standartkļūda	Standard error	Стандарная ошибка	Izlašu vidējo izkliedi ap ģenerālkopas vidējo (matemātisko cerību) raksturojošs lielums
Statistiskā varbūtība	Posteriori probability	Статистическа я вероятность	Notikuma relatīvais biežums, ap kuru tas svārstās atsevišķās novērojumu vai izmēģinājumu sērijās
Stjūdenta sadalījums	Student's distribution (t-distribu-tion)	Распределение Стюдента	Izlases novērojumu sadalījums, kas iegūts no normāli sadalītas ģenerālkopas, līdzīgs normālajam sadalījumam, jo mazāk ir izlasē novērojumu, jo tas ir plakanāks
Stratificētā (tipoloģiskā) izlase	Stratified sample	Типологичес-кая выборкв	Ģenerālkopu sadala apakškopās pēc pētījumam būtiskas pazīmes. No izdalītajām apakškopām proporcionāli to lielumam atlasa pētāmās vienības
Ticamības (reprezentāci-jas) intervāls	Confidence interval	Доверитель-ный интервал	Intervāls, kurā ar noteiktu varbūtību atradīsies nezināmais ģenerālkopas parametrs

SECINOŠĀ STATISTIKA

Iepriekšējās nodaļās tika apgūta prasme aprakstīt masveida objektus un parādības, iegūtos datus grupējot, zīmējot histogrammu un aprēķinot lokācijas rādītājus. Tas viss ir aprakstošā statistika. Neapšaubāmi, ir svarīgi saprast, kas ir pētāmais objekts. Piemēram, zinot veikala klientu raksturojumu, var labāk apmierināt klientu vēlmes. Lūk, iepriekšējā teikuma otrā daļa satur pamatojumu secinošajai statistikai. Statistiskie pētījumi uzņēmējdarbībā nekad nav pašmērķis, bet gan līdzeklis lēmumu pieņemšanai. No statistiskās informācijas ir jāizdara secinājumi, kas ir labāks, piemērotāks, kas jādara, lai iegūtu vēlamo rezultātu.

Secinošā statistika aptver vairākas analīzes metodes:

- hipotēžu pārbaudi, vai pastāv būtiskas atšķirības starp divām empīriskajām kopām vai arī empīrisko un teorētisko sadalījumu;
- dispersijas analīzi, kas parāda, vai grupēšanas pamats (grupēšanai izvēlētā pazīme) veiksmīgi atklāj variācijas cēloni;
- korelācijas analīzi, kas parāda sakarības ciešumu starp neatkarīgo (faktoriālo) pazīmi un atkarīgo (rezultatīvo) pazīmi;
- regresijas analīzi, kas atklāj sakarības matemātisko formu starp faktoriālo (X) un rezultējošo (Y) pazīmi – $y = f(x)$. Šīs formulas pieraksts nozīmē to, ka, zinot x vērtību, y vērtību var aprēķināt pēc kādas konkrētas formulas.

5. HIPOTĒŽU PĀRBAUDE

Pēc nodaļas apgūšanas studentiem:

- *jāzina statistisko hipotēžu pārbaudes soļi;*
- *jāprot definēt nulles un alternatīvo hipotēzi, pēc dotas formulas aprēķināt empīrisko kritērija vērtību un interpretēt iegūtos rezultātus;*
- *jāprot veikt hipotēžu pārbaudi ar Microsoft Excel datu analīzes rīku;*
- *jāzina nodaļā apskatītie hipotēžu pārbaudes (testu) veidi un to pielietošanas sfēras.*

5.1. Ievads

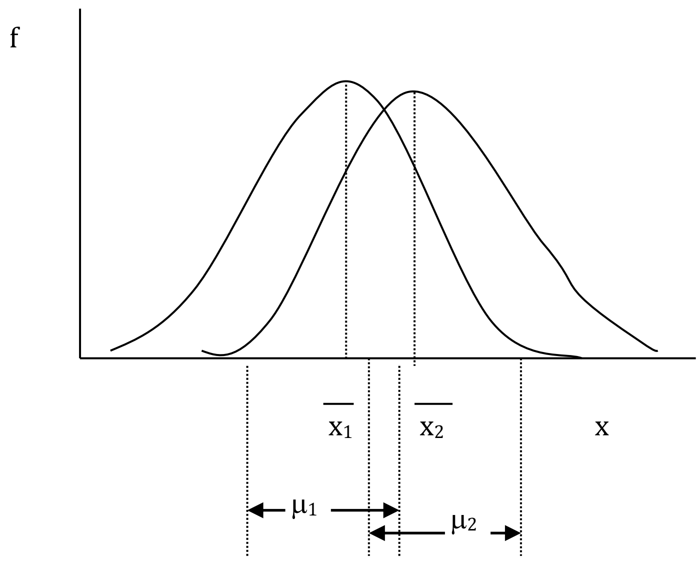
Tā kā pētot uzņēmējdarbības statistikā gandrīz vienmēr izmanto nepilno statistisko novērošanu (netiek apsekotas visas pētāmās kopas vienības, bet tikai daļa, un pēc apsekotajām vienībām tiek izdarīti secinājumi par visu kopu), tad iegūtie izlases raksturotāji precīzi neatbilst ģenerālkopas parametriem (ir reprezentācijas kļūda). Līdz ar to, salīdzinot divas kopas (izlases), kuru raksturotāji, protams, atšķiras, nevar droši teikt, ka šīs atšķirības ir statistiski nozīmīgas. Piemēram, tiek pārbaudītas divu specialitāšu studentu zināšanas ekonomiskajā statistikā. Vienas specialitātes studenti vidēji uzrāda nedaudz labākas zināšanas nekā otras specialitātes studenti. Bet, lai droši apgalvotu, ka pirmās specialitātes studenti ir spējīgāki statistikā, ir jāveic papildus aprēķini, jo kā pirmajā, tā otrajā grupā ir studenti, kuri uzrāda izcilas, labas un vājas sekmes. Jautājums par statistisko nozīmīgumu ir par to, vai, atkārtojot novērojumu (izmēģinājumu), iepriekšējie secinājumi paliek spēkā. Ja salīdzina ģenerālkopas, tad iegūtā starpība, lai cik tā arī nebūtu niecīga, vienmēr ir būtiska.

Veicot izlases pētījumu, ir zināmi tikai izlases raksturotāji ($\bar{x}$, s, A un E), bet ģenerālkopas parametri (μ, σ, A un E) – nē. Tāpēc pieņem, ka nezināmie ģenerālkopas parametri ir aptuveni vienādi ar izlases raksturotājiem.[1,2] Problēmas grafiskā interpretācija ir parādīta 5.1. attēlā.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 151. lpp.

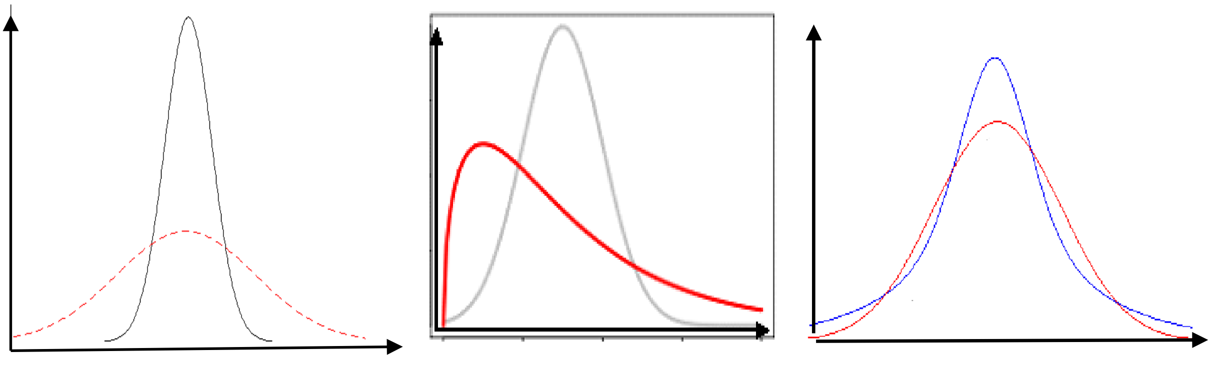
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 118. lpp.



5.1. attēls. **Divu paraugkopu salīdzināšanas grafiskā interpretācija**

5.1. attēlā ir parādīti divu kopu sadalījumi – X_1 kopai ir novērotas nedaudz mazākas pazīmju vērtības nekā X_2 kopai, bet tomēr ievērojamu daļu novēroto pazīmju lielumi "pārklājas". Zinot to, ka ģenerālkopas sadalījums atšķiras no izlases sadalījuma, tad ir iespējams, ka starp ģenerālkopām nav atšķirības ($\mu_1 = \mu_2$). Tas nozīmē, ka pazīme, pēc kuras ir izdalītas šīs kopas, ir nebūtiska datu variācijas izskaidrošanai, iegūtās paraugkopas (izlases) var apvienot, uzskatot, ka tās ir vienas kopas divas izlases. 5.1. attēlā ir parādītas kopas, kuras faktiski atšķiras ar vidējām vērtībām, dispersija, asimetrija un ekscess faktiski ir vienādi. Kopas var atšķirties ar jebkuru no pieminētajiem parametriem, kaut arī parasti tiek pārbaudīts galvenais kopu raksturojošais parametrs – vidējais.

5.2. attēlā ir parādīta citu parametru grafiskā interpretācija.



a b c

5.2. attēls. **Grafiskās interpretācijas, ja kopas atšķiras a) pēc standartnovirzes; b) pēc asimetrijas rādītāja; c) pēc ekscesa rādītāja**

Ja kopas tiek salīdzinātas, balstoties uz parametriem (vidējie lielumi, standartnovirze, asimetrijas un ekscesa rādītāji), tad salīdzināšanas metodes dēvē par parametriskajām metodēm.[1]

Parametriskās metodes lieto, ja ir zināma empīriskā sadalījuma atbilstība kādam teorētiskajam sadalījumam (normālais, logaritmiski normālais, binomiālais, Puasona u. c.). Visbiežāk sastopamais sadalījuma veids ir normālais (variantes ar vērtībām, kas tuvas vidējam, ir pārstāvētas plašāk nekā lielākās vai mazākās vērtības). Ļoti bieži empīriskā sadalījuma atbilstību normālajam (teorētiskā sadalījuma veids) nemaz nepārbauda, tikai vizuāli novērtē histogrammu vai grupējuma rezultātus tabulā un lieto parametriskās salīdzināšanas metodes normālajam sadalījumam. Tomēr, ja ir šaubas par sadalījuma raksturu vai arī vēlas iegūt viennozīmīgi traktējamu rezultātu (divi pētnieki par vienu attēlu var izdarīt atšķirīgus subjektīvus vērtējumus), tad ir jāsalīdzina iegūtais empīriskais sadalījums ar teorētisko sadalījumu.

Ja nav zināma sadalījuma atbilstība kādam teorētiskajam sadalījumam, empīriskais sadalījums neatbilst nevienam no pārbaudītajiem teorētiskajiem sadalījumiem, t. i., pārbauda arī atributīvu sadalījumu (grupēšanas pamats ir atributīva pazīme), tad ir jālieto neparametriskās sadalījumu salīdzināšanas metodes.

Nākamajā apakšnodaļā tiks apskatīts, kā pārbaudīt empīriskā sadalījuma atbilstību teorētiskajam sadalījumam.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 378. lpp.

5.2. Empīriskā sadalījuma atbilstības teorētiskajam sadalījumam pārbaude

Iepriekš jau tika minēts, ka sadalījuma atbilstību teorētiskajam sadalījumam var novērtēt vizuāli, bet, ja vēlas lielāku precizitāti vizuālajam novērtējumam, tad var izmantot grafiskās pārbaudes:

- lineāro metodi;
- sadalījuma līkņu metodi.

Vizuālais novērtējums tomēr ir pakļauts vērtētāja subjektivitātei, tādēļ labāk izmantot analītiskās metodes, kas dod viennozīmīgu vērtējumu. Visbiežāk lietotā ir χ^2 (hī kvadrāta) metode. Pārbaudes var veikt arī pēc Kolmogorova - Smirnova λ (lambda) metodes, vai arī novērtējot asimetrijas un ekscesa rādītājus.[1]

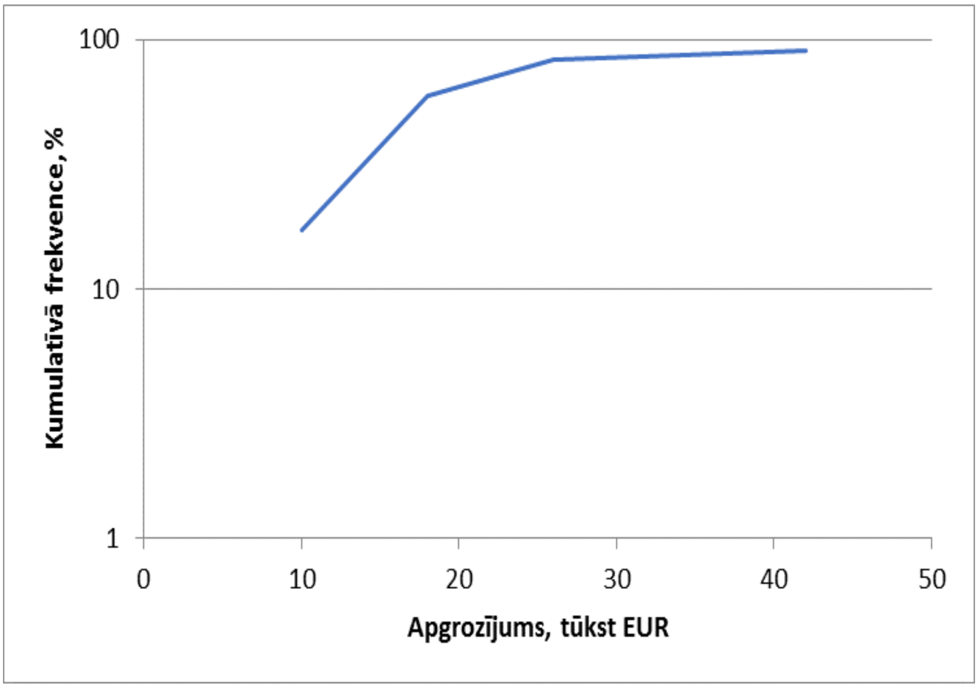
5.2.1. Grafiski vizuālā metode

Empīriskā un normālā sadalījuma atbilstības grafisko pārbaudi pēc lineārās metodes veic šādi:

1) aprēķina sadalījuma relatīvās kumulatīvās frekvences;
2) zīmē grafiku, uz abscisu (*X*) ass atliek pazīmes vērtības parastajā skalā, bet uz ordinātu (*Y*) ass attēlo kumulatīvās relatīvās frekvences logaritmiskajā skalā;
3) ja atliktie punkti uz grafika veido apmēram taisnu līniju, tad sadalījums atbilst teorētiskajam sadalījumam. Punktu izvietojumu novērtē intervālā no 10 līdz 90 %. Ārpus šī intervāla esošos punktus grafikā neatzīmē.

5.3. attēlā ir parādīts metodes pielietojums 2.1. piemēra datiem.

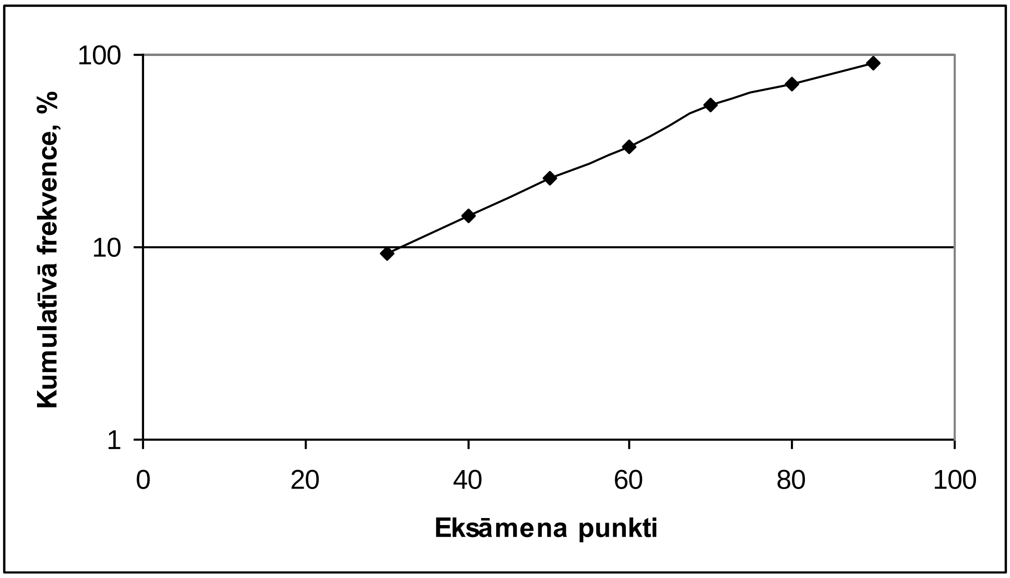
[1] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 70. lpp.



5.3. attēls. **Grafiskā pārbaude – lineārā metode tirdzniecības aģentu apgrozījuma piemēram**

Vizuāli izskatās, ka līnija nav taisne, tas nozīmē, ka tirdzniecības aģentu apgrozījums neatbilst normālajam sadalījumam. Līdzīgu pārbaudi var veikt arī eksāmena punktu uzdevumam (3.1. piemērs).

Grafiku iespējams konstruēt ar programmu *Excel* no datiem, kas ir iegūti ar rīku ”Histogram”. No grupējumu tabulas pārraksta klašu labās robežas un atbilstošās kumulatīvās frekvences (robežās no 10 līdz 90 % kumulatīvajiem biežumiem). Skaitļus labāk pierakstīt parastajā formātā bez procentu zīmes. Tad izvēlas grafika tipu – *XY Scatter*, izpilda grafika konstruēšanai nepieciešamās darbības dialoga logā. Veic izvēlētās formatēšanas darbības, piemēram, nomaina fona krāsu u.tml. *Y* asij ir jānomaina lineārais mērogs uz logaritmisko. Lai to izdarītu, uzklikšķina uz *Y* ass, parādās dialoga logs “Format Axis”, izvēlas opciju “Scale” (mērogs) un ieliek ķeksīti lodziņā pretī “Logaritmic scale”. Grafiks ir gatavs, un to var novērtēt.



5.4. attēls. **Grafiskā pārbaude – lineārā metode eksāmena punktu piemēram**

Šķiet, ka šī piemēra sadalījums atbilst normālajam sadalījumam.

Grafiskās pārbaudes pēc sadalījuma līkņu salīdzināšanas metodes būtība ir empīriskā sadalījuma poligona (var arī histogrammu) salīdzināšana ar šim konkrētajam sadalījumam atbilstošo teorētisko sadalījumu. Teorētiskā sadalījuma biežuma aprēķināšana ir šāda:

1) aprēķina intervālu robežu standartizētās vērtības. normālajam sadalījumam μ = 0, bet σ = 1, tādēļ intervālu robežas (x vērtības) izsaka standartnovirzēs pēc formulas:

$$z = \frac{x_i - \mu}{\sigma} \quad , \qquad (5.1.)$$

kur *z* (dažreiz apzīmē ar t) – intervālu robežas standartizētā vērtība;

x_i – intervāla robežas nestandartizētā vērtība;

μ ($\bar{x}$) – vidējais;

σ (s) – standartnovirze;

2) atbilstoši aprēķinātajām z vērtībām normālā sadalījuma integrālās funkcijas vērtību tabulā nolasa atbilstošos normālā sadalījuma lielumus;
3) aprēķina intervāliem atbilstošās diferenciālās funkcijas vērtības (pēc savas būtības tie ir teorētiskā sadalījuma relatīvie biežumi);
4) destandartizē iegūtās relatīvās vērtības, sareizinot normālā sadalījuma relatīvās frekvences ar empīrisko novērojumu skaitu;
5) attēlo empīrisko un atbilstošo teorētisko sadalījumu līniju grafikā;
6) vizuāli novērtēja līniju līdzību.[1]

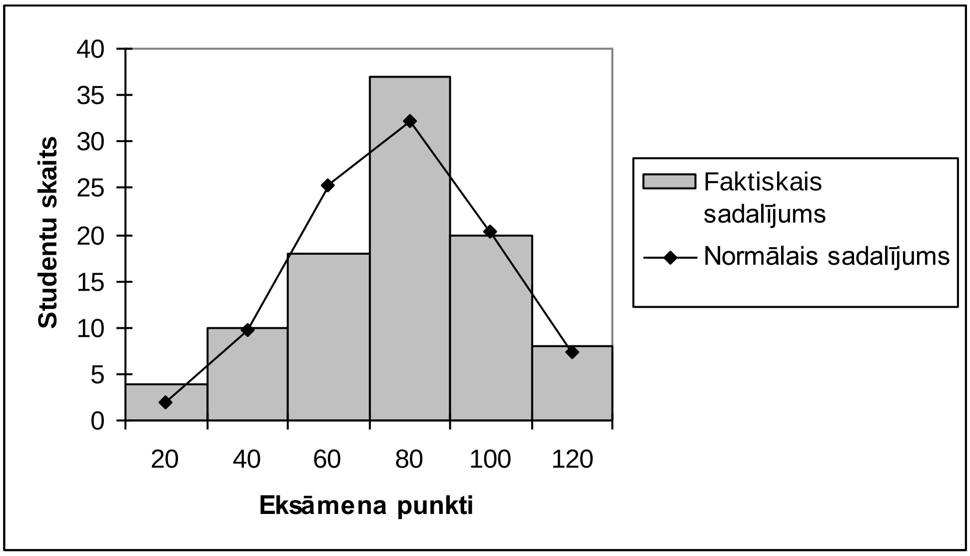
[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 94.lpp.

5.1. tabula

Normālā sadalījuma biežumu aprēķināšanas tabula eksāmena punktu piemēram
(3.1. piemērs)

Intervāla zemākā robeža	Intervāla augstākā robeža	Novērojumu biežums	Standartizētā intervāla zemāka robeža	Standartizētā intervāla augstākā robeža	Normālā sadalījuma integrālās funkcijas vērtība zemākajai robežai	Normālā sadalījuma funkcijas vērtība augstākajai robežai	Intervāla normālā sadalījuma varbūtība	Destandartizēts normālā sadalījuma biežums intervālam
-∞	20	4	-∞	-2,038	0	0,0208	0,0208	2,02
20	40	10	-2,038	-1,170	0,0208	0,1209	0,1000	9,71
40	60	18	-1,170	-0,303	0,1209	0,3808	0,2599	25,21
60	80	37	-0,303	0,564	0,3808	0,7135	0,3327	32,27
80	100	20	0,564	1,431	0,7135	0,9237	0,21023	20,39
100	+∞	8	1,431	+∞	0,9237	1	0,0763	7,40
Kopā		97					1	97

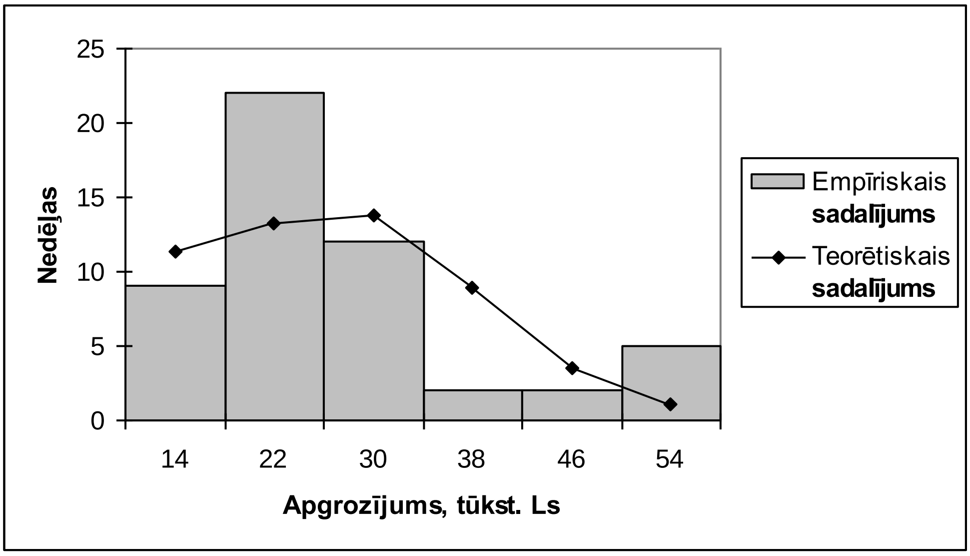
Iegūtos rezultātus atzīmē grafikā. Lai labāk saprastu atšķirības starp empīrisko un tam atbilstošo teorētisko sadalījumu, empīrisko sadalījumu attēlo ar histogrammu, bet teorētisko sadalījumu – ar poligonu. 3.1. piemēra un 5.1. tabulā aprēķinātie dati ir parādīti 5.5. attēlā.



5.5. attēls. **Empīriskā sadalījuma salīdzināšana ar atbilstošo teorētisko, izmantojot līkņu salīdzināšanas metodi eksāmena punktu piemēram**

Empīriskajam sadalījumam ir dažas nesakritības ar teorētisko sadalījumu: ir mazāks novērojumu skaits gradācijas klasē no 40 līdz 60 punktiem, toties lielāks tas ir nākamajā klasē, salīdzinot ar teorētiskā sadalījuma biežumiem.

5.6. attēlā ir parādīts grafiks tirdzniecības aģentu apgrozījuma piemēram (2.1. piemērs).



5.6. attēls. **Empīriskā sadalījuma salīdzināšana ar atbilstošo teorētisko, izmantojot līkņu salīdzināšanas metodi tirdzniecības aģentu apgrozījuma piemēram**

Tirdzniecības aģentu apgrozījuma piemēram nesakritības šķiet lielākas nekā eksāmena punktu piemēram.

Empīriskais sadalījums vienmēr (kaut nedaudz) atšķirsies no teorētiskā, un grafiskās analīzes metodes nedos viennozīmīgu vērtējumu, vai apskatāmais empīriskais sadalījums atbilst, vai neatbilst teorētiskajam sadalījumam. Analītiskās metodes, kuras balstās uz aprēķinātiem kritērijiem, dod viennozīmīgu atbildi, tādēļ, vērtējot empīriskā sadalījuma atbilstību teorētiskajam sadalījumam, šiem kritērijiem būtu dodama priekšroka.

5.2.2. Kritēriju metodes empīriskā sadalījuma atbilstības teorētiskajam sadalījumam pārbaudei

Plašāk lietotais ir χ^2 **(hī kvadrāta) kritērijs**.[1,2,3] Metodi 1900. gadā ieteica lietot K. Pīrsons. Kā daudzas kritēriju metodes, tā arī šī paredz aprēķināt empīrisko kritērija vērtību, kuru vēlāk salīdzina ar tabulās doto (datora

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 187. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 384. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 685.

ģenerētu) kritisko vērtību. Ja empīriskā (pēc pētījuma datiem aprēķinātā) kritērija vērtība pārsniedz kritisko vērtību (tabulā), tad nulles hipotēzi ar izvēlēto varbūtību noraida. Nulles hipotēze nozīmē, ka starp pētāmajiem sadalījumiem (empīrisko un teorētisko) nav būtiskas atšķirības. Ja nulles hipotēze tiek noraidīta, tad ir konstatētas būtiskas atšķirības un dotais empīriskais sadalījums neatbilst izvēlētajam teorētiskajam sadalījumam. Šādā situācijā var pārbaudīt sava sadalījuma atbilstību kādam citam teorētiskam sadalījumam. Ja neatrod nevienu piemērotu teorētisko sadalījumu, tad turpmākajā analīzē lieto neparametriskās metodes.[1]

Empīrisko χ^2 (hī kvadrāta) kritēriju aprēķina pēc 5.2. formulas:

$$\chi^2 = \sum_i \frac{\left(f_i - f_{ti}\right)^2}{f_{ti}^{`}}, \qquad (5.2.)$$

kur f_i – i klases frekvence (novērojumu skaits grupā);
f_{ti} – i klasei atbilstošā teorētiskā frekvence (cik novērojumiem būtu jābūt šajā grupā, ja sadalījums būtu normālais vai kāds cits teorētiskais sadalījums.

Kā jau iepriekš tika apskatīts, formulas, kas satur summēšanas zīmi, parasti aprēķina tabulas veidā. Ir jāturpina analizēt tirdzniecības aģentu apgrozījuma 2.1. piemēru, jo iepriekš lietotās grafiskās metodes nedeva pietiekamu pamatu lēmuma par sadalījuma atbilstību normālajam izdarīšanai.

5.2. tabula

Empīriskā un normālā sadalījuma atbilstības pārbaude pēc χ^2 metodes tirdzniecības aģentu apgrozījuma piemēram (sākotnējie dati doti 2.1. piemērā, teorētiskās frekvences aprēķinātas 4.6. tabulā)

x_i	f_i	f_{ti}	f_i-f_{ti}	$(f_i-f_{ti})^2$	$(f_i-f_{ti})^2/f_{ti}$
10	9	11	-2,170	4,707	0,421426
18	22	13	9,036	81,657	6,298908
26	12	14	-1,780	3,168	0,229927
34	2	9	-7,142	51,002	5,57916
42	7	5	2,000	4,000	0,8
50	0	0			
Σ	52	52			13,32942

|d_{max}| - maksimālā starpība starp teorētisko un empīrisko frekvenci Kolmogorova - Smirnova kritērija aprēķināšanai

$$\chi^2 = \sum_i \frac{\left(f_i - f_{ti}\right)^2}{f_{ti}^{`}}$$

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 378. lpp.

χ^2_{krit} – *hī* kvadrāta kritērija kritiskā vērtība ir atkarīga no izvēlētā nozīmības līmeņa α (alfa), kas ir pieļaujamās kļūdas varbūtība, un brīvības pakāpju skaita ν (*nī*). Brīvības pakāpju skaitu nosaka pēc formulas:

$$\nu = k - 1 - r,$$

kur k – gradācijas klašu skaits;

r – teorētiskā sadalījuma parametru skaits, kas ir zināms katram teorētiskajam sadalījumam. Normālajam sadalījumam ir divi parametri – vidējā vērtība un standartnovirze.

Skaitlis "1" formulā nozīmē paraugkopas apjoma ierobežojumu.

α ekonomiskajos pētījumos parasti izvēlas 0,05. Tas nozīmē, ka ir pieļaujama 5 % varbūtība, ka izdarītie secinājumi ir kļūdaini. Lielāku precizitāti neļauj izvēlēties ekonomisko pētījumu dati (parasti pētījuma apjoms ir pārāk mazs datu iegūšanas izmaksu un pētījuma neatkārtojamības dēļ, lai kādus secinājumus varētu balstīt augstākā ticamības līmenī (1 – α)*100 %).

Pielikumā tabulā (vai datorā ar funkciju „CHINV") atrastā kritiskā hī kvadrāta (χ^2_{krit}) vērtība ir

$\chi^2_{\alpha=0{,}05;\ \nu=5-1-2=2} = 5{,}99$

$\chi^2_{emp} > \chi^2_{krit}$ *(13,33 > 5,99), tātad nulles hipotēze tiek noraidīta. Tirdzniecības aģentu apgrozījuma sadalījums neatbilst normālajam sadalījumam.*

Iespējamie risinājuma ceļi ir vairāki:

- pārbauda empīriskā sadalījuma atbilstību citam teorētiskajam sadalījumam. Šajā gadījumā pārbaudīt būtu vērts, vai tirdzniecības aģentu apgrozījuma sadalījums atbilst logaritmiski normālajam sadalījumam, jo empīriskais sadalījums ir ar pozitīvu asimetriju;

- veic papildus informācijas iegūšanu (uzskaita tirdzniecības aģentu apgrozījumu par vēl vienu periodu). Tad pārbauda, vai sadalījums pēc būtības saglabā savu raksturu. Lielākā pētījuma apjomā nejaušās datu svārstības izlīdzinās, saglabājas likumsakarības. Ar paplašināto datu apjomu veic atkārtotas pārbaudes atbilstībai kādam no teorētiskajiem sadalījumiem;

- ja saglabājas empīriskā sadalījuma tendence ar divām virsotnēm, tad tā ir pazīme, ka vienā sadalījumā ir ietvertas divas atšķirīgas kopas. Kopu identifikācija nav statistikas uzdevums, to veic pētnieks ar ekonomiski loģiskās analīzes palīdzību. Konkrētajā piemērā varētu izvirzīt hipotēzi par to, ka dažiem aģentiem ir atšķirīgi apstākļi – palīdz ģimenes locekļi, ir zināms kāds noieta kanāls, kas nav zināms citiem aģentiem.

Bez minētās χ^2 metodes empīrisko sadalījumu pārbaudei var lietot arī **Kolmogorova - Smirnova λ metodi**.[1,2] Tā arī ir kritēriju metode. Vispirms aprēķina empīrisko kritērija vērtību, kuru salīdzina ar tabulās doto kritisko vērtību. Iegūtos rezultātus interpretē tāpat kā citiem kritērijiem. Ja $\lambda_{emp} > \lambda_{krit}$, tad nulles hipotēze tiek noraidīta, tas nozīmē, ka pētāmais sadalījums neatbilst izvēlētajam teorētiskajam sadalījumam. Savukārt, ja $\lambda_{emp} < \lambda_{krit}$, tad nulles hipotēze paliek spēkā, un tas nozīmē, ka var veikt tālāko analīzi ar parametriskajām metodēm, atšķirības, kas ir starp empīrisko un teorētisko sadalījumu nosaka nejauši faktori. λ_{emp} aprēķina pēc 5.3.formulas:

$$\lambda_{emp} = \frac{|d_{\max}|}{\sqrt{n}}, \qquad (5.3.)$$

kur $|d_{\max}|$ – maksimālā absolūtā vērtība starpībai starp empīriskā un teorētiskā sadalījuma klašu frekvencēm;
n – izlases (paraugkopas) apjoms.

Tirdzniecības aģentu apgrozījuma uzdevumam (5.3. tabula) λ_{emp} ir:

$$\lambda_{emp} = \frac{9{,}036}{\sqrt{52}} = 1{,}253$$

Kritisko λ vērtību nosaka tikai izvēlētais nozīmības līmenis, un to var viegli aprēķināt pēc 5.4.formulas:

$$\lambda_{\alpha} = \sqrt{\frac{1}{2}\ln\frac{2}{\alpha}}, \qquad (5.4.)$$

kur α – izvēlētais nozīmības līmenis.

Ja izvēlas α= 0,05, tad λ_{α} = 1,36.

$\lambda_{emp} < \lambda_{krit}$ (1,253<1,36). Pēc šī kritērija sanāk, ka nulles hipotēzi nevar noraidīt.

Kolmagorova - Smirnova kritērijs ir piemērots liela apjoma paraugkopām. Ja empīriskā kritērija vērtība tuvojas kritiskajai, tad tā ir pazīme, ka kļūdīšanās varbūtība ir liela.

Kā redzams no iepriekšējām divām metodēm, tad arī analītiskā metode neļauj viennozīmīgi traktēt iegūtos rezultātus. Spriežot pēc χ^2 metodes, tirdzniecības aģentu apgrozījuma sadalījums neatbilst normālajam sadalījumam, bet pēc λ metodes – ir atbilstošs normālajam sadalījumam.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 378. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 194. lpp.

Pārbaudīt, vai pētāmais empīriskais sadalījums atbilst normālajam, var arī, pārbaudot atbilstību pēc asimetrijas un ekscesa rādītājiem.[1]

Protams, empīriskajam sadalījumam šie rādītāji kaut nedaudz atšķirsies no normālā sadalījuma parametriem (abi koeficienti normālajam sadalījumam ir nulle). Vai šī atšķirība uzskatāma par statistiski nozīmīgu, vai arī ir radusies nejaušību rezultātā, pārbauda ar kritēriju palīdzību. Kritērijus aprēķina ar 5.5. un 5.6. formulu:

$$t_A = \frac{|A|}{s_A}, \tag{5.5.}$$

kur $|A|$ – asimetrijas koeficienta modulis;
S_A – asimetrijas koeficienta reprezentācijas kļūda;

$$t_E = \frac{|E|}{s_E}, \tag{5.6.}$$

kur $|E|$ – ekscesa koeficienta modulis;
S_E – ekscesa koeficienta reprezentācijas kļūda.

Abiem koeficientiem kritiskā vērtība ir 3.

Asimetrijas un ekscesa koeficientu aprēķina formulas ir dotas 3.nodaļā, bet praktiski tās iegūst ar aprakstošās statistikas (*descriptive statistics*) rīku.

Reprezentācijas kļūdu var aprēķināt ar 5.7. –5.10. formulu:

$$S_A = \sqrt{\frac{6}{n}} \tag{5.7.}$$

jeb precīzāku formulu:

$$S_A = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}, \tag{5.8.}$$

$$S_E = 2\sqrt{\frac{6}{n}} \tag{5.9.}$$

vai vēl precīzāku:

$$S_E = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \tag{5.10.}$$

Tirdzniecības aģentu apgrozījuma piemēram aprēķina asimetrijas un ekscesa koeficientu reprezentācijas kļūdas.

Pēc 5.7 .formulas: $S_A = \sqrt{\frac{6}{52}} = 0{,}3397$

[1]Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 78. lpp.

Pēc 5.8.formulas: $S_A = \sqrt{\frac{6*52*(52-1)}{(52-2)*(52+1)*(52+3)}} = 0{,}3304$

Pēc 5.9.formulas: $S_E = 2\sqrt{\frac{6}{52}} = 0{,}6794$

Pēc 5.10.formulas: $S_E = \sqrt{\frac{24*52*(52-1)^2}{(52-3)*(52-2)*(52+3)*(52+5)}} = 0{,}6501$

Kā redzams, īpaši lielas starpības starp precīzāko un aptuvenāko formulu dotajiem rezultātiem nav.

A (skewness) ir 1,159747, bet E (kurtosis) ir 0,77457. Aprēķinus skatīt nodaļā par lokācijas rādītājiem (3.11. tabulā).

$t_A = \frac{|1{,}159747|}{0{,}33074} = 3{,}51$*, ja reprezentācijas kļūdu aprēķina pēc 5.7. formulas, tad* t_A*= 3,41*

t_A *= 3,51 > 3, tātad sadalījums pēc asimetrijas rādītāja nav normālais.*

$$t_E = \frac{0{,}77457}{0{,}6501} = 1{,}19$$

t_E *= 1,19 < 3, tātad pēc ekscesa rādītāja nevar noraidīt nulles hipotēzi, un tas varētu būt arī atbilstošs normālajam sadalījumam, bet pietiek tikai ar to, ka viens no šiem rādītājiem neatbilst normālajam sadalījumam, lai atzītu, ka dotais empīriskais sadalījums neatbilst normālajam sadalījumam.*

5.2.3. Empīriskā un logaritmiski normālā sadalījuma atbilstības pārbaude

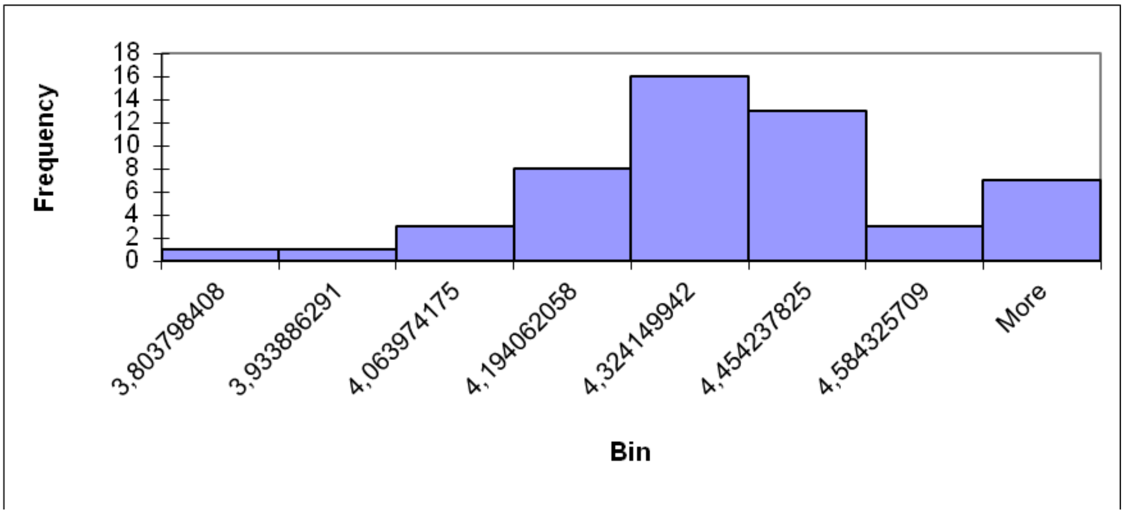
Logaritmiski normālā sadalījuma būtība ir aprakstīta 4.2.5. apakšnodaļā.

Iedzīvotāju ienākumu sadalījums atbilst šim teorētiskajam sadalījuma veidam. Mazākos ienākumus un atbilstoši tiem arī tēriņus nosaka izdzīvošanai nepieciešamais minimālais līdzekļu apjoms, to regulē minimālā alga un sociālie pabalsti, bet lielākie ienākumi nav ierobežoti. Nabagie un salīdzinoši mazturīgi iedzīvotāji ir dominējošā sabiedrības daļa, bet bagāto un ļoti bagāto iedzīvotāju ir nedaudz.

Viens no veidiem, kā pielietot normālā sadalījuma metodes sadalījumam, kurš neatbilst normālajam sadalījumam, ir šo sadalījumu normalizēt. Tas nozīmē, ar sākotnējiem datiem jāveic matemātiskas manipulācijas, kuru rezultātā pārrēķinātie dati veidotu normālajam sadalījumam atbilstošu līkni.

Normāli logaritmisko sadalījumu normalizē, sākotnējos datus logaritmējot (no šai normalizācijai nepieciešamās darbības arī ir iegūts sadalījuma nosaukums).

Pārbaudīt, vai tirdzniecības aģentu apgrozījuma sadalījums ir logaritmiski normālais sadalījums.



5.7. attēls. **Pēc logaritmētiem 2.1. piemēra datiem izveidotā histogramma**

Skatoties vizuāli, kā arī novērtējot logaritmēto datu asimetrijas (Skewness ir -0,0572) un ekscesa (Kurtosis ir 0,181848) rādītājus, var secināt, ka sadalījums ir normalizēts. Analītiskās metodes pielieto analoģiski, kā tas tika darīts ar normālo sadalījumu. Protams, joprojām ir jautājums par palielināto novērojumu skaitu lielākajā gradācijas klasē – vai tas ir nejaušības rezultāts, vai arī likumsakarība?! Šo atbildi var iegūt, veicot nākamos novērojumus un aprēķinus.

5.2.4. Empīriskā un citu teorētisko sadalījumu atbilstības pārbaude

Pārbaudot atbilstību kādam citam teorētiskajam sadalījumam, izmanto χ^2 un λ metodi, bet nevērtē pēc A un E koeficientiem, jo šīs divas metodes ir specifiskas normālajam sadalījumam. Atšķirīga ir tikai teorētisko frekvenču aprēķināšanas kārtība.

Puasona sadalījums ir reto notikumu sadalījums. Šim sadalījumam ir viens parametrs, kuru apzīmē ar grieķu alfabēta burtu λ (lambda).[1] Tieši tāpat apzīmē arī Kolmogorova - Smirnova kritēriju, bet šie rādītāji nav viens un tas pats. Vienādi apzīmējumi var maldināt, tāpēc pirms aprēķinu veikšanas ir jāpārliecinās par aprēķināmā rādītāja būtību un vai izvēlētā formula ir īstā.

---

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 72. lpp.

Puasona sadalījuma vidējo vērtību aprēķina pēc 5.11. formulas:

$$\lambda = n*p, \qquad (5.11.)$$

kur n – novērojumu skaits;

p – labvēlīgā notikuma iestāšanās varbūtība (mazāka par 0,1, ja tā ir lielāka, tad lieto vai nu binomiālo sadalījumu, vai arī normālo sadalījumu).

Ja p nav zināms, tad par λ pieņem empīrisko novērojumu vidējo vērtību.

Puasona sadalījumu varētu lietot tehnoloģiskajos pētījumos, novērtējot defektu rašanās biežumu (pārsvarā tiek ražota produkcija, kas atbilst standartam, bet ik pa laikam rodas brāķis). Tāpat Puasona sadalījums varētu būt noziegumu, slimību, ceļu negadījumu statistikā.

Pārbaudīt 5.1. piemērā minētā empīriskā sadalījuma atbilstību Puasona sadalījumam ar χ^2 metodi.

5.1. piemērs. Neliela uzņēmuma īpašnieks ir reģistrējis darbinieku skaitu, kas neierodas darbā.

Strādnieku skaits, kas nav darbā, x_i	Kavējumu reizes	Varbūtība, P(x)	x_i*P_i
0	125	0,50	0
1	58	0,23	0,23
2	30	0,12	0,24
3	25	0,10	0,3
4	5	0,02	0,08
5	5	0,02	0,1
6	2	0,01	0,06
	250		1,01 = $\Sigma\, x_i*P_i$

Vizuāli novērtējot datus, var izvirzīt hipotēzi, ka sadalījums atbilst Puasona sadalījumam, cilvēki pārsvarā iet uz darbu, tomēr vidēji katru dienu darbā neierodas nedaudz vairāk kā 1 cilvēks, konkrēti 1,01. Vidējais rādītājs Puasona sadalījumā ir λ.

Teorētisko varbūtību aprēķina pēc 5.12. formulas:

$$P_{m,n,p} = \frac{\lambda^m}{m!} e^{-\lambda}, \qquad (5.12.)$$

kur $P_{m,n,p}$ – varbūtība, ka notikums iestāsies m reizes no n mēģinājumiem, ja iestāšanās varbūtība ir p.

5.12. formulas manuāls aprēķins prasa ievērojamas matemātikas zināšanas un ir darbietilpīgs. Tāpēc teorētiskās varbūtības Puasona sadalījumam vieglāk un ērtāk atrast ar *Excel* funkciju *Poison*. 4.8. attēlā ir parādīta šīs funkcijas lietošana.

x ir notikumu skaits, ko var ievadīt manuāli vai norādīt ar šūnas adresi.

Mean – aritmētiskais vidējais, aprēķināts pēc empīriskajiem datiem, var ievadīt ar roku vai norādīt ar šūnas adresi.

Cumulative – vai ir nepieciešama kumulatīvā vērtība (integrālās funkcijas vērtība) – *True* (patiesi), ja vēlas iegūt integrālās funkcijas vērtību, *False* (kļūdaini), ja nepieciešama diferenciālās funkcijas vērtība.

5.3. tabula

Puasona sadalījuma varbūtību aprēķins

Kavētāju skaits	Faktiskais absolūtais biežums	Puasona sadalījuma varbūtība	Puasona sadalījums darba kavējumiem
0	125	0,364219	91
1	58	0,367861	92
2	30	0,18577	46
3	25	0,062543	16
4	5	0,015792	4
5	5	0,00319	1
6	2	0,000537	0

A66 šūna

Excel lietotā formula =POISSON(A66;1,01;FALSE)

Puasona sadalījuma varbūtība (0,364) x kopējais novērojumu skaits (250)

Tālāk aprēķinus veic līdzīgi kā citiem teorētiskajiem sadalījumiem.

5.4. tabula

χ^2 kritērija aprēķināšana 5.1. piemēram

Strādnieku skaits, kas nav darbā, x_i	*Faktiski*	*Teorētiski*	f_i-f_{ti}	$(f_i-f_{ti})^2$	$(f_i-f_{ti})^2/f_{ti}$
0	125	91	34	1156	12,7
1	58	92	-34	1156	12,6
2	30	46	-16	256	5,6
3	25	16	9	81	5,1
4 un vairāk	12	5	7	49	9,8
Kopā	250	250			45,7

χ^2_{emp} = 45,69

χ^2_{krit} = nosaka α (nozīmības līmenis) un ν (brīvības pakāpju skaits), kas ir klašu skaits mīnus divi, jo šim sadalījumam ir viens parametrs un skaitlis 1 brīvības pakāpju aprēķina formulā nozīmē paraugkopas apjoma ierobežojumu.

Pēc klašu apvienošanas ir izveidojušās 5 klases, tas nozīmē, ka ν = 3, bet χ^2_{krit} = 7,81 (nolasa pielikuma tabulā).

Tā kā χ^2_{emp} = 45,69 > χ^2_{krit} = 7,81, tad var izdarīt secinājumu, ka darba kavējumi neatbilst Puasona sadalījumam, tāpēc turpmākajā analīzē ir jālieto neparametriskās metodes.

Empīriskā sadalījuma atbilstības teorētiskajam sadalījumam praksē (uzņēmējdarbības vadībā, studentu diplomdarbos, bieži arī zinātniskos pētījumos) pārbaudi parasti izlaiž. Vairumā gadījumu *apriori* (bez pieredzes, tas ir, bez aprēķiniem) pieņemot, ka pētāmais objekts ir ar sadalījumu, kas atbilst normālā sadalījuma likumam, lieto parametriskās salīdzināšanas metodes. Šāda empīriskā sadalījuma atbilstības teorētiskajam sadalījumam pārbaude tiks veikta, ja būs nopietns zinātnisks pētījums, kur ir svarīga pētījuma precizitāte, vai sadalījuma vizuālais vērtējums radīs šaubas, vai šis sadalījums tiešām atbilst normālajam sadalījumam.

Daudz vairāk no praktiskā viedokļa interesē divu paraugkopu salīdzinājums – vai tās ir atšķirīgas, piemēram, tehnoloģiju ietekme uz rezultātiem u.tml. Šis jautājums tiks pētīts nākamajā apakšnodaļā.

5.3. Divu empīrisko kopu salīdzināšana

Hipotēze ir spriedums, kas ir loģiski ticams, bet prasa papildus pierādījumus. Kad nepieciešamie pierādījumi, kas apstiprina hipotēzi ir savākti, hipotēze kļūst par zinātnisku teoriju.[1,2] Šīm teorijām ir ļoti atšķirīga nozīmība, visbiežāk tādas teorijas ir tīri praktiskas ievirzes secinājumi uzņēmuma vadīšanā.

Statistiskā hipotēze izsaka slēdzienus par divu kopu līdzību vai atšķirību.[3,4] Lai izvirzītu statistisko hipotēzi, ir jābūt oriģinālās problēmas hipotēzei. Piemēram, ir "aizdomas" (hipotēze), ka vienas specialitātes studenti matemātiskos priekšmetus apgūst labāk nekā kādas citas specialitātes studenti. Kā starp vienas, tā otras specialitātes studentiem ir studenti ar izcilām, labām, viduvējām un vājām zināšanām. Ja secinājumus izdara par galīgām kopām ar pilno novērošanu (tieši šī kursa visi studenti), tad spriedums ir viennozīmīgs neatkarīgi no starpības lieluma. Ja kopas ir hipotētiskas, secinājumus var attiecināt arī uz atkārtotiem līdzīgiem pētījumiem (uz nākamo periodu peļņas vai ražošanas apjomiem u.tml.), vai arī galīgas kopas nav novērotas ar pilnu novērošanu, bet tās pārstāv izlases, tad starpībai ir jābūt pietiekami lielai, lai varētu droši apgalvot, ka kopas ir atšķirīgas.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 150. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 117. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 150. lpp.
[4] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 117. lpp.

Ir jānovērtē, vai atšķirība starp kopām ir būtiska. Datus parasti iegūst ar izlases metodi, tas nozīmē, ka nezināmās ģenerālkopas parametri nedaudz atšķirsies no izlases raksturotājiem. Problēmas grafiskā interpretācija ir parādīta 5.1.attēlā. Vispirms der apskatīt, kā salīdzināt divas kopas, kuru sadalījumi atbilst normālajam sadalījumam (visbiežāk sastopamā situācija).

Pārbaudot statistiskās hipotēzes, ir iespējamas divu veidu kļūdas:

1) ja noraida pareizu nulles hipotēzi (atzīst, ka starp kopām ir būtiska atšķirība, kaut arī patiesībā tās nav), šo kļūdas varbūtību apzīmē ar α, un tas ir tas pats nozīmības līmenis, par kuru tika runāts jau iepriekš;
2) ja nenoraida nepareizu nulles hipotēzi. Šo kļūdu apzīmē ar β, to nevar izmērīt, tādēļ to cenšas aizvietot ar α, palielinot tās varbūtību.[1]

Hipotēžu testēšanā ir 5 soļi:

1) definē nulles un alternatīvo hipotēzi;
2) atrod formulu, pēc kuras aprēķina empīrisko kritērija vērtību;
3) nosaka būtiskuma līmeni (α), ar kuru tiks noraidīta nulles hipotēze;
4) aprēķina empīrisko kritērija vērtību, salīdzina to ar kritisko kritērija vērtību, ko nolasa tabulās vai aprēķina ar *Excel* funkcijām. Izdara secinājumus par nulles hipotēzi. Neatkarīgi no izvēlētā kritērija veida nulles hipotēzi noraida, ja empīriskā kritērija vērtība ir lielāka par kritisko, atšķirība starp pārbaudāmajiem lielumiem ir statistiski nozīmīga. Ja kritiskā vērtība ir lielāka par empīrisko, tad nulles hipotēzi noraidīt nevar, atšķirība nav pierādīta (nav teikts, ka tā nepastāv);
5) "pārtulko" secinājums pētāmajā problēmā – nosaka, vai starp pētāmajām kopām ir statistiski nozīmīga atšķirība, kura kopa uzrāda labākus rezultātus.[2,3,4]

Galvenais kopu raksturojošais lielums ir aritmētiskais vidējais. Tādēļ parasti pārbaude tiek veikta, izvērtējot, vai ģenerālkopas vidējie (matemātiskās cerības) atšķiras, vai nē.

Otrs parametrs, kuru parasti izmanto kopu raksturošanai ir standartnovirze vai arī dispersija, kas ir standartnovirzes kvadrāts – datu izkliedes mērs. Kā individuālu kopu atšķirību raksturojošu rādītāju dispersiju lieto reti (varētu būt kvalitātes pētījumos, kad svarīgs ir ne tikai vidējais izmērs, bet arī atsevišķu mērījumu precizitāte), bet novērtējumu par dispersiju atšķirību izmanto,

[1] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 107. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 152. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 358.
[4] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 126. lpp.

izvēloties aritmētisko vidējo salīdzināšanas formulas. Šajā sakarā kā pirmā tiks apskatīta divu izlašu dispersijas salīdzināšanas metodika.

5.3.1. Hipotēžu pārbaude par divu izlašu dispersijām

Vispirms formulē nulles hipotēzi. H_0: $\sigma_1^2 - \sigma_2^2 = 0$. Nulles hipotēzi dispersijām pārbauda trīs gadījumos:[1]

1) lai izvēlētos formulas divu aritmētisko vidējo salīdzināšanai. Tas ir iemesls, kāpēc šis jautājums tiek skatīts tieši tagad. Praksē ļoti bieži aprēķinus aizvieto eksperta pieņēmums. Pētnieks izdara secinājumus par kopu raksturu – to atbilstību teorētiskajam, parasti normālajam sadalījumam, par dispersiju līdzību pārbaudāmajām kopām u.tml. bez īpašiem aprēķiniem;
2) atsevišķos gadījumos var interesēt tieši dispersiju atšķirība, piemēram, pārbaudot šķirojamās mašīnas precizitāti. Šai gadījumā interesē tieši datu izkliede;
3) lai salīdzinātu iekšgrupu un starpgrupu dispersijas, noskaidro, vai grupējums atklāj kādas būtiskas likumsakarības. Šis jautājums tiks apskatīts nākamajā tēmā par dispersijas analīzi.

Lai noteiktu, vai pastāv statistiski nozīmīga atšķirība starp dispersijām, ir jāizpilda vairāki soļi.

1. Jāaprēķina abas salīdzināmās empīriskās dispersijas pēc formulas, kura ņem vērā brīvības pakāpju zudumu:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} \tag{5.13.}$$

No formulas, kura tika apskatīta empīrisko sadalījumu analīzes nodaļā, tā atšķiras ar to, ka saucējā ir nevis novērojumu skaits, bet brīvības pakāpes, kas ir novērojumu skaits mīnus viens saistošais nosacījums. Šādai atšķirībai ir nozīme, ja izlases apjoms ir neliels ($n \leq 30$). Ja dispersija ir jau aprēķināta pēc 3.16.formulas, tad var veikt korekciju pēc 5.14.formulas:

$$\hat{s}^2 = s^2 \frac{n}{n-1}, \tag{5.14.}$$

kur $\hat{s}^2$ – uz brīvības pakāpēm koriģētais dispersijas lielums (nenobīdītais dispersijas novērtējums);
s^2 – pēc 3.16., 3.17., 3.21., 3.22. formulas aprēķinātā dispersija;
n – novērojumu skaits izlasē.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 157. lpp.

2. Aprēķina empīrisko Fišera kritēriju (F_{emp}):

$$F = \frac{s_{liel}^2}{s_{maz}^2} \qquad (5.15.)$$

Kritērija aprēķināšanai lielāko dispersiju noteikti dala ar mazāko (lielāko skaitli ar mazāko).

3. Nosaka brīvības pakāpju skaitu un atrodam kritisko Fišera kritērija vērtību. Brīvības pakāpes aprēķina, atņemot no katras izlases novērojuma skaita vieninieku ($\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$). Otrs faktors, kas nosaka kritiskās vērtības lielumu, ir α (nozīmības līmenis), kas ir pieļaujamās kļūdas varbūtība, parasti $\alpha = 0{,}05$.
4. Salīdzina F_{emp} ar F_{krit} un izdara secinājumu līdzīgi kā tas ir visās kritēriju metodēs. Ja $F_{emp} > F_{krit}$, nulles hipotēzi tad noraida un atzīst, ka vienas kopas datu izkliede ir būtiski lielāka nekā otras kopas datu izkliede. Ja $F_{emp} < F_{krit}$, tad nulles hipotēze paliek spēkā un tiek uzskatīts, ka nav būtiskas atšķirības datu izkliedē salīdzināmajās kopās.

Pārbaudīt nulles hipotēzi par divu studentu grupu eksāmenu rezultātu salīdzināšanu. Ir apsekotas visas vienības (pilnā statistiskā novērošana) – iegūta informācija par visiem abu grupu studentiem, taču, ja eksāmena rezultātus vēlas vispārināt kā studentu zināšanas attiecīgajā priekšmetā, tad eksāmena rezultātus var uztvert kā izlasi. Atkārtojot pārbaudījumu ar līdzīgiem jautājumiem un uzdevumiem, rezultāti nedaudz atšķirtos. Sākotnējie dati ir doti 2.3. tabulā.

$$x_1 = 6{,}11;\ x_2 = 6{,}38;\ s_1^2 = 2{,}57;\ s_2^2 = 2{,}19$$

Kā redzams, 1. studentu grupai vidējais vērtējums ir nedaudz zemāks, bet datu izkliede ir nedaudz lielāka nekā 2. studentu grupai.

Dispersijas ir aprēķinātas pēc 3.22. formulas. Ja aprēķinus vēlas veikt ļoti precīzi, tad ir jāaprēķina dispersiju nenobīdītie vērtējumi pēc 5.14. formulas. Šajā konkrētajā gadījumā sagaidāms, ka korekcija dos nelielu efektu, jo 1. kopā ir 97 studenti, bet 2. kopā – 39 studenti (pietiekami nozīmīgu efektu brīvības pakāpju zudums dod mazajām izlasēm, kur n < 30).

Aprēķinot dispersiju nenobīdītos vērtējumus, iegūst:

$$\hat{s}^2{}_1 = 2{,}57 * \frac{97}{96} = 2{,}6$$

$$\hat{s}^2{}_2 = 2{,}185 * \frac{39}{38} = 2{,}24$$

F_{emp} = 2,6/2,24 = 1,16, bet, ja lieto sākotnējos dispersiju lielumus, tad F_{emp} = 2,57/2,19 = 1,17. Šī atšķirība tiešām nav liela un nevar ietekmēt rezultātu vērtējumu.

Tabulā atrod F_{krit}, $\nu_1 = 96$; $\nu_2 = 38$; $\alpha = 0{,}05$.

Tabulā nav tik precīzu skaitļu, tādēļ kritisko vērtību novērtē aptuveni ar interpolācijas palīdzību. **Interpolācija** ir nezināma skaitļa meklēšana zināmā skaitļu apgabalā.

5.8. attēlā ir parādīts, kā tabulās sameklē vajadzīgo lielumu.

Meklē brīvības pakāpi, υ_1= 96. Tā kā tabulās nav precīzas vērtības, tad paņem divas ailes, starp kurām ir interesējošā vērtība

$\upsilon_2 \backslash \upsilon_1$	1	2	3	.. ..	80	90	100	120	250	∞
1	161,45	199,50	215,71		252,72	252,90	253,04	253,25	253,80	254,31
2	18,51	19,00	19,16		19,48	19,48	19,49	19,49	19,49	19,50
3	10,13	9,55	9,28		8,56	8,56	8,55	8,55	8,54	8,53
....										
30	4,17	3,32	2,92		1,71	1,70	1,70	1,68	1,65	1,62
35	4,12	3,27	2,87		1,65	**1,64**	**1,63**	1,62	1,59	1,56
40	4,08	3,23	2,84		1,61	**1,60**	**1,59**	1,58	1,54	1,51
....										
100	3,94	3,09	2,70		1,41	[illegible]	[illegible]	1,38	1,33	1,28
120	3,92	3,07	2,68		1,39	[illegible]	[illegible]	1,35	1,30	1,25
250	[illegible]	3,03	2,64							

Brīvības pakāpe υ_2=38. Tā kā ir 35 un 40, tādēļ paņem divas rindas, starp kurām ir interesējošais lielums

Fišera kritērija kritisko vērtību nolasa ailes un rindas krustpunktā. Tā kā šajā gadījumā ir divas ailes un divas rindas, tad kritiskā vērtība atrodas kaut kur starp 1,64 un 1,59. 96 ir tuvāk 100 nekā 90, un 38 ir tuvāk 40 nekā 35, tad nezināmā kritērija vērtība būs nedaudz tuvāk 1,59 nekā 1,64, tā varētu būt aptuveni 1,61

5.8. attēls. **Fišera kritērija kritisko vērtību tabulas fragments**

Kā redzams no datu apgabala, kas attiecas uz doto situāciju, $F_{krit} \approx 1,6$. Līdz ar to var izdarīt secinājumu, ka nulles hipotēze paliek spēkā, starp pētāmo kopu dispersijām nav būtiskas atšķirības, tātad eksāmena rezultātu izkliede abās studentu grupās ir aptuveni vienāda.

F_{krit} vērtību $F_{\alpha;\, v1;\, v2}$ var aprēķināt ar MS Excel funkciju FINV.

Divas <u>negrupētas</u> izlases var salīdzināt ar datu analīzes rīku F-test Two Samples for Variances.

Ja ir grupēti dati, tad aprēķinus vienkāršāk ir veikt manuāli. Lai ar F-test metodi salīdzinātu studentu grupas, tad 1. grupai pēc kārtas ir jāievada visas 97, bet 2. grupai – 39 atzīmes, un tas ir darbietilpīgāk nekā, veicot aprēķinus manuāli.

5.3.2. Nulles hipotēze par aritmētisko vidējo starpību

Aritmētisko vidējo starpības pārbaudei ir izšķirami vairāki gadījumi:

1) divu neatkarīgu kopu vidējo salīdzināšana[1,2,3], piemēram, izvērtējot divas tehnoloģijas, darbu metodikas u. tml., lai izdarītu secinājumu par to, vai viena no tām ir labāka;
2) divu atkarīgu kopu vidējo salīdzināšana[4,5], piemēram, izvērtējot kāda pasākuma efektivitāti. Šajā situācijā faktiski tiek pētītas vienas kopas izmaiņas laikā. Atkarīgas kopas var būt arī divi mērījumi katram izstrādājumam kvalitātes pētījumos u. c. Lai aritmētisko vidējo salīdzināšanai izmantotu šo metodiku, svarīgi, ka katram vienas kopas elementam ir pakārtots otras kopas elements, faktiski nesalīdzina divas kopas, bet gan izmaiņas;
3) salīdzinot kopas vidējo ar kādu uzdoto, parasti standarta, rādītāju[6,7], piemēram, kvalitātes pētījumos.

Kā pirmā no minētajām situācijām tiks apskatīta 3. situāciju, kad izlases vidējo salīdzina ar uzdoto standartu. Situācija ir aprakstīta 5.2. piemērā.

Šeit ir hipotētiskā kopa – saražotās produkcijas apjoms nepārtraukti tiek papildināts. Taču visas pudeles pārbaudīt nevar, jo tā tiks sabojāta prece.

Pārbaudot izlasi, rodas lielāka vai mazāka reprezentācijas kļūda (izlase precīzi neatspoguļo ģenerālkopu).

Metodes būtība balstās uz atziņu: ja no ģenerālkopas, kurai ir normāls sadalījums, veido daudzas izlases, tad izlašu vidējie arī veido normālo sadalījumu ar vidējo vērtību μ – ģenerālkopas vidējo. Ja izlases vidējais atrodas tuvu vēlamajam (standarta) lielumam, tad droši vien arī nezināmais ģenerālkopas vidējais būs apmēram standarta vērtībā, tas nebūs būtiski atšķirīgs.

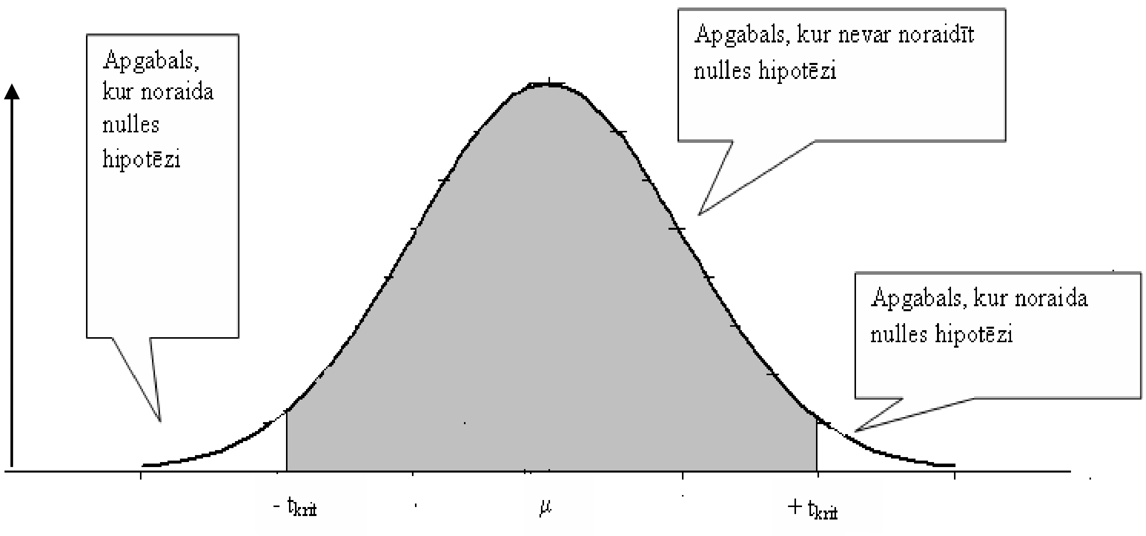
---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 152. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 122. lpp.
[3] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 438.
[4] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 157. lpp.
[5] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 125. lpp.
[6] Turpat, 119. lpp.
[7] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 412.

5.2. piemērs. Uzņēmumā tiek veikti kvalitātes pētījumi – tiek kontrolēta dozēšanas iekārta, kas pilda limonādi pudelēs. Reizi dienā no saražotās produkcijas apjoma tiek paņemtas 25 pudeles, kas tiek atkorķētas un pārmērītas precīzos mērcilindros. Lēmumu pieņemšanu par ražošanas pārtraukšanu un iekārtu regulēšanu pamato ar nulles hipotēzes pārbaudi. Ja nulles hipotēzi nevar noraidīt (saražotās produkcijas vidējais tilpums nav būtiski atšķirīgs no standarta tilpuma), tad ražošanu turpina. Ja nulles hipotēzi noraida, tad aptur ražošanu un regulē iekārtas. Pēdējā mērījuma dati mililitros ir šādi:

502,1	504,0	494,3	502,9	498,0
498,2	501,6	498,0	499,6	501,1
499,6	496,2	494,7	496,9	498,8
495,5	499,9	499,7	499,7	497,4
496,4	495,0	496,6	503,6	497,7

Pudeles standarta tilpums ir 500 ml.



5.9. attēls. **Grafiskā interpretācija nulles hipotēzes pārbaudei, salīdzinot izlases vidējo ar noteikto standartu**

t_{krit} ir Stjūdenta sadalījuma kritiskās vērtības.

Lai noskaidrotu, vai izlases vidējais atšķiras no standarta lieluma, jāveic šādi soļi:

1) formulē nulles un alternatīvo hipotēzi, piemērā $H_0: \mu = 500$ un $H_1: \mu \neq 500$;
2) aprēķina izlases standartkļūdu (pēc rakstura tā ir izlašu vidējo standartnovirze no ģenerālkopas vidējā). Standartkļūdas aprēķināšanas metodika tika apskatīta 4. nodaļā.

Vienkāršai gadījumizlasei aprēķināšanas formula ir:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{5.16.}$$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} \tag{5.17.}$$

Dispersija tiek aprēķināta kā nenobīdītais, konverģējošais vērtējums ģenerālkopas dispersijai (aprēķinot ņem vērā brīvības pakāpju zudumu), bet *n* – izlases apjoms;

3) aprēķina empīrisko Stjūdenta kritērija vērtību pēc 5.18. formulas:

$$t_{emp} = \frac{\left|\bar{x} - \mu\right|}{s_{\bar{x}}}, \tag{5.18.}$$

kur $\bar{x}$ – izlases vidējā vērtība;
μ – pārbaudāmais ģenerālkopas vidējais (standarta tilpums u. tml.);
$s_{\bar{x}}$ – standartkļūda;

4) izvēlas hipotēzes pārbaudes nozīmības līmeni α (parasti ekonomiskajos pētījumos tas ir 0,05). Kvalitātes pētījumos var iegūt statistiski labus datus. Ja ir svarīga tehnoloģiskā precizitāte, tad var izvēlēties arī lielāku precizitāti (α = 0,01 vai pat 0,001, piemēram, medikamentiem);
5) tabulās atrod t_{krit}, to nosaka iepriekš izvēlētais nozīmības līmenis α un brīvības pakāpju skaits $\nu = n - 1$;
6) izdara secinājumus par nulles hipotēzi.

Aprēķinus vieglāk veikt ir ar aprakstošās statistikas (*descriptive statistics*) datu analīzes rīku (tika apskatīts 3. nodaļā).

5.5. tabulā ir doti iegūtie rezultāti.

Aritmētiskais vidējais, standartnovirze (ņemot vērā brīvības pakāpju zudumus) un pat standartkļūda ir aprēķināta, atliek ievietot skaitļus 5.15. formulā un aprēķināt empīrisko Stjūdenta kritērija vērtību.

$$t_{emp} = \frac{\left|498{,}7 - 500\right|}{0{,}552} = 2{,}355$$

5.5. tabula

Ar aprakstošās statistikas datu analīzes rīku iegūtie rezultāti 5.2. piemēram

Column1	
Mean	498,7
Standard Error	0,552178
Median	498,2
Mode	499,6
Standard Deviation	2,760888
Sample Variance	7,6225
Kurtosis	-0,7029
Skewness	0,30951
Range	9,7
Minimum	494,3
Maximum	504
Sum	12467,5
Count	25

Stjūdenta kritisko vērtību tabulā vai ar Excel funkciju „TINV" atrod $t_{\alpha=0,05:\upsilon=24}=2,064$.

Tā kā $t_{emp}=2,355>t_{krit}=2,064$, *nulles hipotēzi var noraidīt. Pudelēs vidēji pildītais tilpums būtiski atšķiras no standarta tilpuma, ražošana ir jāaptur un jāveic iekārtu regulēšana.*

5.3.3. Divu neatkarīgu kopu vidējo salīdzināšana

Pirmkārt, salīdzinot divu kopu vidējos, novērtē, vai kopas ir savstarpēji nesaistītas. Ja kopas ir savstarpēji saistītas – katram vienas kopas elementam var pakārtot otras kopas atbilstošo elementu, tad lietderīgāk ir izmantot divu atkarīgu kopu salīdzināšanas metodiku, iegūtā informācija būs pilnīgāka. Ir situācijas, kad novērojumu informācija pēc būtības ir par saistītajām kopām, bet saistītās vienības nav sakārtotas viena otrai pretī, tad ir jāizvērtē iespēja veikt korekcijas datos un izmantot divu savstarpēji atkarīgu kopu salīdzināšanas metodiku. Piemēram, pētījums par iestāšanos eirozonā ietekmi uz uzņēmējdarbības rezultātiem. Pētījumā iegūta informācija par uzņēmumu rentabilitāti pirms valsts iestāšanās eirozonā un pēc tās, turklāt liela daļa informācijas ir par tiem pašiem uzņēmumiem pirms un pēc iestāšanās. Var izvērtēt un izslēgt no analīzes tos uzņēmumus, par kuriem ir tikai viena perioda informācija.

Otrkārt, ja kopas ir nesaistītas, tad ir jānovērtē, vai datu izkliede abās kopās ir statistiski vienāda. vai ir būtiskas atšķirības. Dispersiju salīdzināšana apskatīta iepriekš. Manuālajos aprēķinos tas, vai dispersijas ir vienādas, vai atšķirīgas, noteiks izmantojamās formulas veidu, bet datoraprēķinos – izmantojamo metodi.

Treškārt, noskaidro, vai abās kopās ir vienāds novērojumu skaits. Manuālajos aprēķinos tas nosaka izmantojamās formulas izvēli. Ja $n_1 = n_2$, tad izmantojamās formulas ir vienkāršākas.

Veicot aprēķinus manuāli, nepieciešams aprēķināt kopu vidējos un dispersiju nenobīdītos vērtējumus, tas ir, dispersiju aprēķināt, ņemot vērā brīvības pakāpju zudumu.

Vidējo salīdzināšanu veic ar šādu soļu palīdzību:

1) formulē nulles hipotēzi $H_0: \mu_1 = \mu_2$ vai ar savādāku pierakstu tā pati hipotēze $H_0: \mu_1 - \mu_2 = 0$ un alternatīvo hipotēzi $H_1: \mu_1 \neq \mu_2$, *jeb* $H_1: \mu_1 - \mu_2 \neq 0$. Iepriekšējais pieraksts attiecas uz abpusējo hipotēzi, kas paredz, ka var būt lielāks gan vienas, gan otras kopas vidējais. Bieži interesē, vai vienas kopas rādītāji nav labāki par otras kopas rādītājiem un tad alternatīvo hipotēzi mēs formulējam kā vienpusēju nevienādību – $H_1: \mu_1 > \mu_2$ jeb $H_1: \mu_1 - \mu_2 > 0$;

2) aprēķina vidējo lielumu starpības standartkļūdu pēc 5.19. (ja $n_1 = n_2$) vai 5.20. (ja $n_1 \neq n_2$) formulas:

$$s_{\bar{x}_2-\bar{x}_1} = \sqrt{\frac{s_2^{\,2} + s_1^{\,2}}{n}} \qquad (5.19.)$$

$$s_{\bar{x}_2-\bar{x}_1} = \sqrt{\frac{\sum\left(x_{1i} - \bar{x}_1\right)^2 + \sum\left(x_{2i} - \bar{x}_2\right)^2}{n_1 + n_2 - 2} * \frac{n_1 + n_2}{n_1 * n_2}} \qquad (5.20.)$$

3) aprēķina empīrisko Stjudenta koeficienta vērtību pēc 5.21. formulas:

$$t_{emp} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{s_{\bar{x}_1-\bar{x}_2}} \qquad (5.21.)$$

Empīriskā koeficienta aprēķina formulas skaitītājā ir modulis, tas nozīmē, ka nav svarīgi, kuras kopas vidējais ir lielāks, ņem vērā absolūto vērtību starpībai;

4) izvēlas nozīmības līmeni α (prasti 0,05) un aprēķina brīvības pakāpes. Vienkāršākajā gadījumā, ja dispersiju salīdzināšana (vai ekspertu

slēdziens) ir $\sigma_1^2 = \sigma_2^2$, tad $\nu = n_1 + n_2$. Ja ir konstatēts, ka dispersijas nav vienādas $\sigma_1^2 \neq \sigma_2^2$, tad brīvības pakāpju noteikšana kļūst stipri sarežģītāka. Formulas ir atkarīgas no tā, vai salīdzināmās izlases ir ar vienādiem apjomiem, vai atšķirīgiem. Ja $n_1 = n_2$, tad brīvības pakāpes aprēķina ar 5.22. formulu, ja $n_1 \neq n_2$, tad ar 5.23. formulu.

$$\upsilon = (n_1 + n_2 - 2) * \left(0,5 + \frac{s_1^2 * s_2^2}{s_1^4 + s_2^4}\right) \qquad (5.22.)$$

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \qquad (5.23.)$$

Mūsdienās šo sarežģīto formulu lietošana nav aktuāla. Pētnieka uzdevums ir izvēlēties pareizo metodi, bet dators pats veic aprēķinus ar ieprogrammētajām formulām;

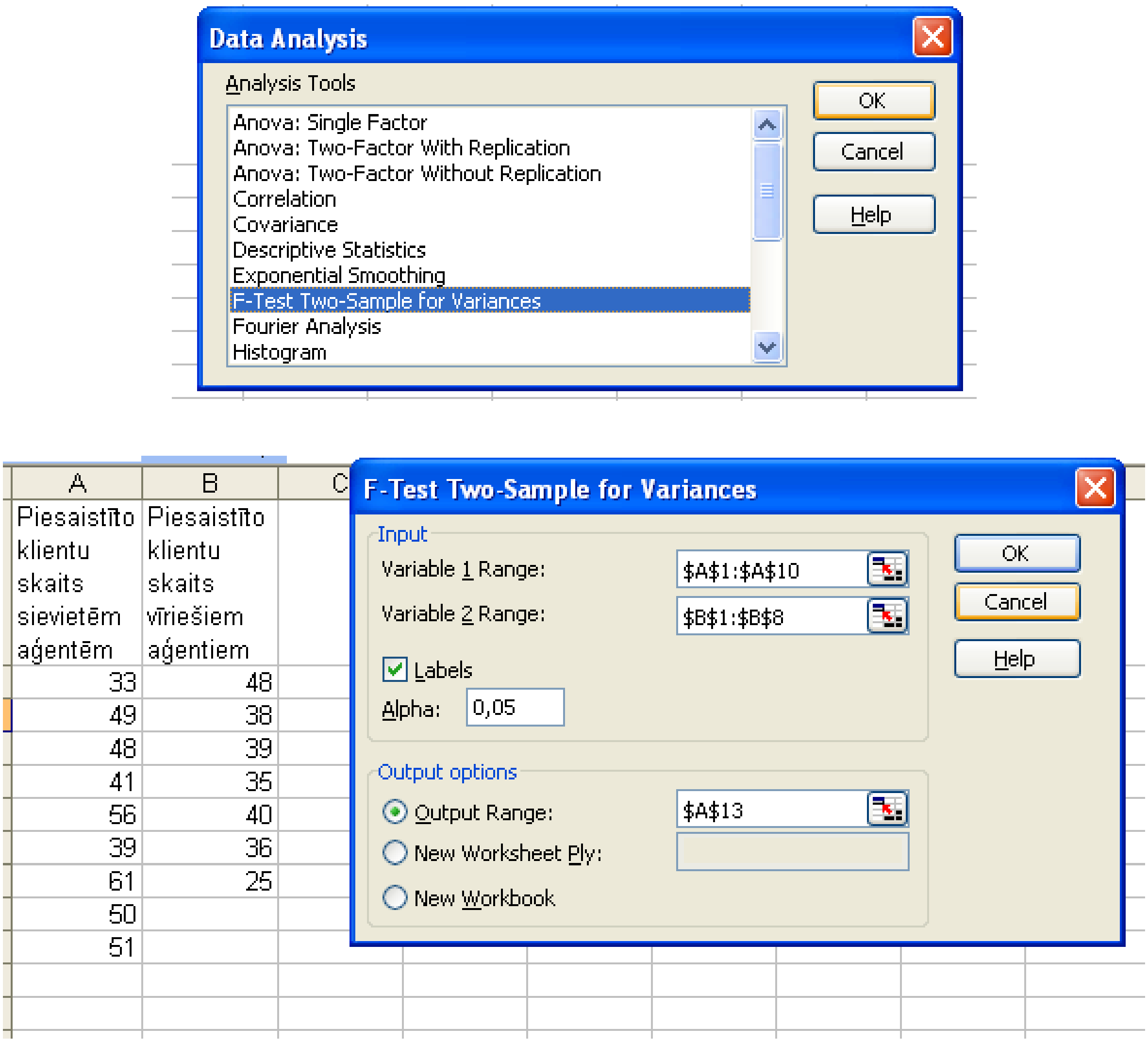
5) tabulās atrod kritisko Stjūdenta kritērija vērtību un interpretējam iegūtos rezultātus. t_{krit} nosaka izvēlētais nozīmības līmenis α un aprēķinātās brīvības pakāpes ν. Vienpusējai hipotēzei (ir jāpierāda, ka vienas ģenerālkopas vidējais ir lielāks nekā otras ģenerālkopas vidējais) α vērtību, kurai nolasa t_{krit}, tabulās, palielina 2 reizes. Kļūdas iespēja ir tikai vienā pusē atšķirībā no divpusējas hipotēzes, kur kļūdīšanās iespēja ir abās pusēs [1,2,3] (5.8. attēls).

5.3. piemērs. Apdrošināšanas sabiedrība vēlas pārbaudīt hipotēzi: sievietes ir veiksmīgākas klientu piesaistītājas nekā vīrieši. Pēdējā apmācību grupā bija 9 sievietes un 7 vīrieši, kas mēnesi nodarbojās ar jaunu klientu piesaisti. Piesaistīto klientu skaits:

Vīrieši	40	38	39	35	40	36	25		
Sievietes	33	49	48	41	56	39	61	60	51

[1] Moore, D. (2003). *The basic practice of statistics.* (3d ed.) New York: W.H.Freeman and Company. p. 438.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 152. lpp.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 122. lpp.

Aprēķinus var veikt, izmantojot *Excel* datu analīzes rīkus. Kopas ir neatkarīgas, tādēļ ir jāizvēlas tests vidējo salīdzināšanai, ja ir vienādas dispersijas vai tās ir atšķirīgas. To var veikt ar Fišera (*F-test two sample for variances*) testu. Dialoga logu aizpildīšanas secība ir parādīta 5.10. attēlā.



5.10. attēls. **Nulles hipotēzes pārbaude par divu kopu dispersiju līdzību**

Iegūtie rezultāti ir atspoguļoti 5.6. tabulā.

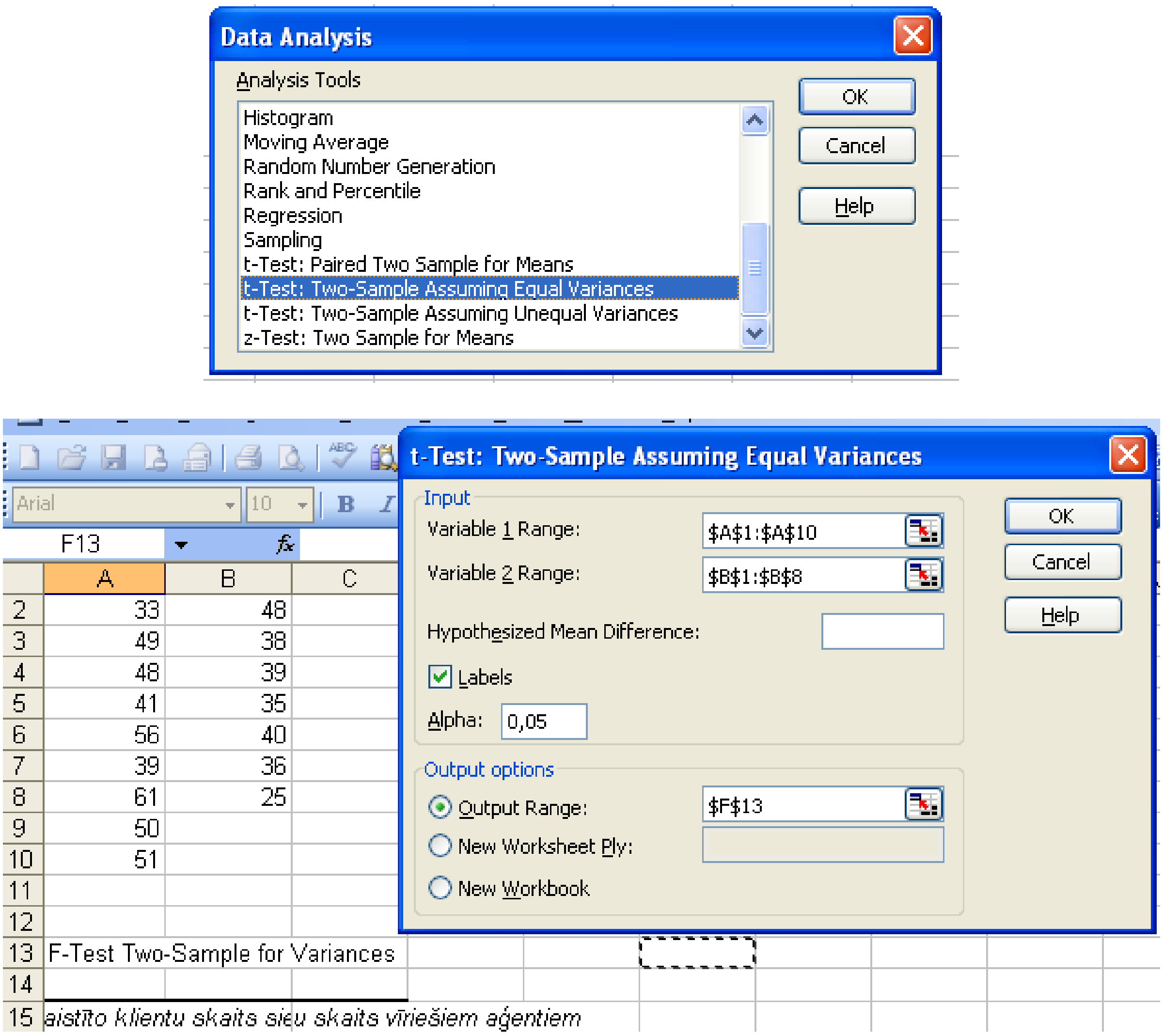
Var izdarīt secinājumu, ka nulles hipotēzi noraidīt nevar, kaut arī sieviešu grupā dispersija (75,0) ir lielāka nekā vīriešu grupā (47,2), jo F=1,59 < F_{krit}=4,15.

Šajā gadījumā šis tests tika veikts, lai izvēlētos piemērotāko testu aritmētisko vidējo salīdzināšanai. Nākamie soļi ir atspoguļoti 5.11. attēlā. (visās kritēriju metodēs)

Ar F-test iegūtais rezultāts

F-Test Two-Sample for Variances

	Piesaistīto klientu skaits sievietēm aģentēm	*Piesaistīto klientu skaits vīriešiem aģentiem*
Mean	*47,555556*	*37,28571*
Variance	*75,027778*	*47,2381*
Observations	*9*	*7*
df	*8*	*6*
F	*1,5882897*	
P(F<=f) one-tail	*0,2950039*	
F Critical one-tail	*4,1468042*	



5.11. attēls. **Darbības, kas jāveic, salīdzinot divu neatkarīgu kopu aritmētiskos vidējos**

Iegūtais rezultāts ir atspoguļots 5.7. tabulā.

5.7. tabula

Divu neatkarīgu kopu salīdzināšana ar t-testu, ja dispersijām nav būtisku atšķirību

t-Test: Two-Sample Assuming Equal Variances

	Piesaistīto klientu skaits aģentēm sievietēm	*Piesaistīto klientu skaits aģentiem vīriešiem*
Mean	*47,55556*	*37,28571*
Variance	*75,02778*	*47,2381*
Observations	*9*	*7*
Pooled Variance	*63,11791*	
Hypothesized Mean Difference	*0*	
df	*14*	
t Stat	*2,565061*	
P(T<=t) one-tail	*0,011226*	
t Critical one-tail	*1,76131*	
P(T<=t) two-tail	*0,022452*	
t Critical two-tail	*2,144787*	

Atkal salīdzina *t – Stat* $(t_{emp}) > t_{krit}$. Šajā gadījumā empīriskā kritērija vērtība ir lielāka gan par vienpusējo (*t Critical one-tail*), gan divpusējo (*t Critical two-tail*) kritērija teorētisko vērtību. Tādēļ nulles hipotēzi noraida, atšķirības starp kopām ir statistiski nozīmīgas. Kaut arī konkrētajā situācijā bija jāpierāda, ka sievietes vidēji piesaista vairāk klientu, pietiek ar vienpusēju hipotēzi un to ir vieglāk pierādīt, jo kļūdīšanās iespēja ir tikai vienā pusē.

Secinājums: sievietes (atbilstoši piemēra datiem) spēj apdrošināt (piesaistīt) vairāk klientu. Tīri no ekonomiski loģiskā viedokļa darbinieku atlasei būtisks kritērijs būtu dzimums, kaut arī tas ir pretrunā ar politkorektumu par dzimumu līdztiesību.

5.3.4. Pa pāriem saistīto novērojumu aritmētisko vidējo starpības hipotēzes pārbaude

Daudzos pētījumos ir situācijas, kad tiek analizēta kāda faktora (piemēram, pieņemta lēmuma) izraisītās sekas. Ir dota informācija par vienu kopu divos laika momentos – pirms pētāmā faktora ietekmes un pēc. Kopas vienību vērtības, protams, variē gan līdz, gan pēc faktora iedarbības. Turklāt variācija var būt tik liela, ka, lietojot divu neatkarīgu kopu vidējo salīdzināšanas metodi, nevar pierādīt, ka izlašu pārstāvēto ģenerālkopu vidējie atšķirtos pietiekami būtiski. Taču, ja skatās katru novērojumu pāri individuāli, tad redzams, ka izmaiņas pārsvarā ir notikušas vienā virzienā, piemēram, iestāšanās ES pozitīvi ietekmē lauku cilvēku labklājību. Ja skatās datu variāciju, tad gan pirms iestāšanās Eiropas Savienībā, gan pēc iestāšanās ir zemnieku saimniecības ar maziem, vidējiem un ir arī ar lieliem ienākumiem, bet kopumā izmaiņu dinamika ir pozitīva. Tāpēc, ja ir darīšana ar šāda tipa novērojumiem, salīdzina nevis divu kopu vidējos lielumus, bet gan pārbauda, vai izmaiņu vidējā vērtība ģenerālkopā nav nulle.

Hipotēzes pārbaudei ir šādi soļi:

1) formulē nulles hipotēzi $H_0: \mu(\Delta) = 0$ un alternatīvo hipotēzi $H_1: \mu(\Delta) \neq 0$ divpusējai hipotēzei, bet parasti interesē, vai pētāmais faktors ir devis pozitīvu rezultātu $H_1: \mu(\Delta) > 0$ vai negatīvu – $H_1: \mu(\Delta) < 0$;

2) aprēķina noviržu dispersiju, ņemot vērā brīvības pakāpju zudumu:

$$s_{\Delta}^2 = \frac{\sum\left(\Delta_i - \overline{\Delta}\right)^2}{n-1} \qquad (5.24.)$$

vai ar momentu formulu:

$$s_{\Delta}^2 = \left(\frac{\Delta_i^2}{n} - \overline{\Delta^2}\right)\frac{n}{n-1} \qquad (5.25.)$$

Šīs formulas pēc savas būtības ir analogas 3.16. un 3.21.formulai. Atšķirība ir tā, ka pazīmju vērtības vietā lieto pazīmes vērtību izmaiņas – $\Delta_i = x_i' - x_i$ un saucējā ir brīvības pakāpes, nevis novērojumu skaits;

3) aprēķina vidējās novirzes standartkļūdu ar formulu:

$$s_{\overline{\Delta}} = \sqrt{\frac{s_{\Delta}^2}{n}} \qquad (5.26.)$$

4) aprēķina empīrisko Stjūdenta kritērija vērtību:

$$t_{emp} = \frac{\overline{\Delta}}{s_{\overline{\Delta}}} \qquad (5.27.)$$

5) izvēlas hipotēzes pārbaudes nozīmības līmeni α un nosaka brīvības pakāpju skaitu $\nu = n - 1$, kur n ir novērojumu pāru skaits. Tabulās nolasa kritisko Stjūdenta kritērija vērtību – t_{krit};

6) izdara secinājumus par iegūtajiem rezultātiem. Ja $t_{emp} > t_{krit}$, tad nulles hipotēzi noraida, tas nozīmē, ka pētāmā faktora ietekme ir bijusi būtiska. Ja $t_{emp} < t_{krit}$, tad nulles hipotēzi noraidīt nevar un pētāmā faktora ietekme nav pierādīta. Situācijā, kad pētījuma hipotēze neapstiprinās, ir jānovērtē, vai pētījums ir bijis pietiekami reprezentabls. Ja izlase ir bijusi adekvāti liela, tad pētījumā ir iegūts statistiski nozīmīgs negatīvs rezultāts – pētāmais faktors nav būtisks. Ja nulles hipotēzi noraidīt nevar tāpēc, ka izlases apjoms ir nepietiekams, tad turpina novērojumus, atkārtoti pārbauda hipotēzi un vai nu gūst tai apstiprinājumu, vai arī galīgi noraida.[1]

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 125. lpp.

5.4. piemērs. Veselīga dzīvesveida grupa piesaista cilvēkus ar lieko svaru, informē tos par veselīgām diētām un organizē sporta nodarbības. Jaunu klientu piesaistei vēlas realizēt mārketinga aktivitāti, kas balstītos uz zinātnisku rezultātu izvērtējumu. Tālāk ir doti rezultāti par 10 programmas dalībnieku svara izmaiņām 4 mēnešu laikā.

Svars, sākot programmu, kg	Svars pēc 4 mēnešiem, kg
114,2	107,8
128,1	113,5
139,5	128,7
131,3	121,8
123,0	109,1
112,3	102,1
122,4	110,4
128,1	118,6
131,9	118,0
147,8	129,4

Uzdevums: noteikt, vai svara samazinājums ir statistiski nozīmīgs, vai tas nevarētu būt nejaušu faktoru rezultāts.

Vidējās vērtības un dispersijas aprēķināšanai manuāli izmantot 5.8. tabulu.

5.8. tabula

Noviržu aritmētiskā vidējā un dispersijas aprēķināšanas palīgtabula 5.4. piemēram

N.p.k.	x_i	x_i'	$\Delta_i = x_i' - x_i$	Δ_i^2
1	*114,2*	*107,8*	*-6,4*	*40,96*
2	*128,1*	*113,5*	*-14,6*	*213,16*
3	*139,5*	*128,7*	*-10,8*	*116,64*
4	*131,3*	*121,8*	*-9,5*	*90,25*
5	*123*	*109,1*	*-13,9*	*193,21*
6	*112,3*	*102,1*	*-10,2*	*104,04*
7	*122,4*	*110,4*	*-12*	*144*
8	*128,1*	*118,6*	*-9,5*	*90,25*
9	*131,9*	*118*	*-13,9*	*193,21*
10	*147,8*	*129,4*	*-18,4*	*338,56*
Σ			*-119,2*	*1524,28*

Aprēķina noviržu dispersiju pēc 5.24. formulas:

$$s_{\Delta}^2 = \left(\frac{1524,28}{10} - 11,92^2\right)\frac{10}{10-1} = 11,49,$$

vidējās novirzes standartkļūdu pēc 5.25. formulas:

$$s_{\bar{\Delta}} = \sqrt{\frac{11{,}49}{10}} = 1{,}07$$

un t_{emp} – pēc 5.26. formulas:

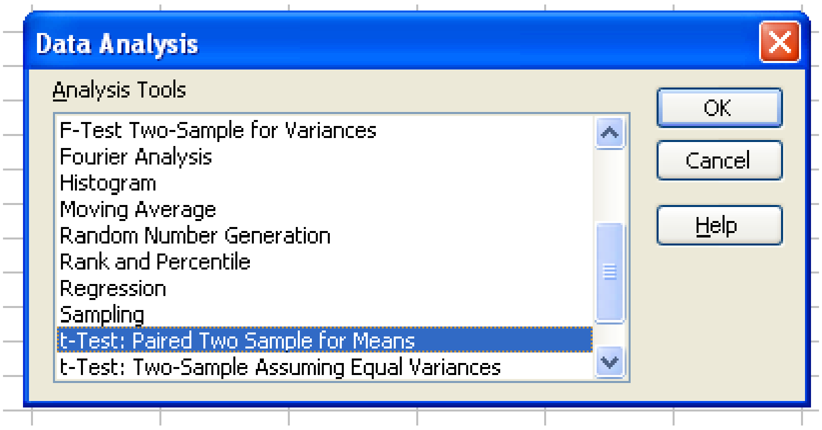
$$t_{emp} = \frac{11{,}92}{1{,}07} = 11{,}14$$

Stjūdenta kritērija kritisko vērtību t_{krit} nolasa tabulā.

υ=10–1=9; α parasti izvēlas 0,05, atbilstošais t_{krit}=2,262, un tas ir daudz mazāks par t_{emp}. Var pārbaudīt arī hipotēzi ar augstāku ticamības līmeni, tabulās dotais mazākais α =0,001, atbilstošā t_{krit} vērtība ir 4,781< t_{emp}=11,14. Tādēļ droši var apgalvot, ka svara sazināšanas metode ir efektīva.

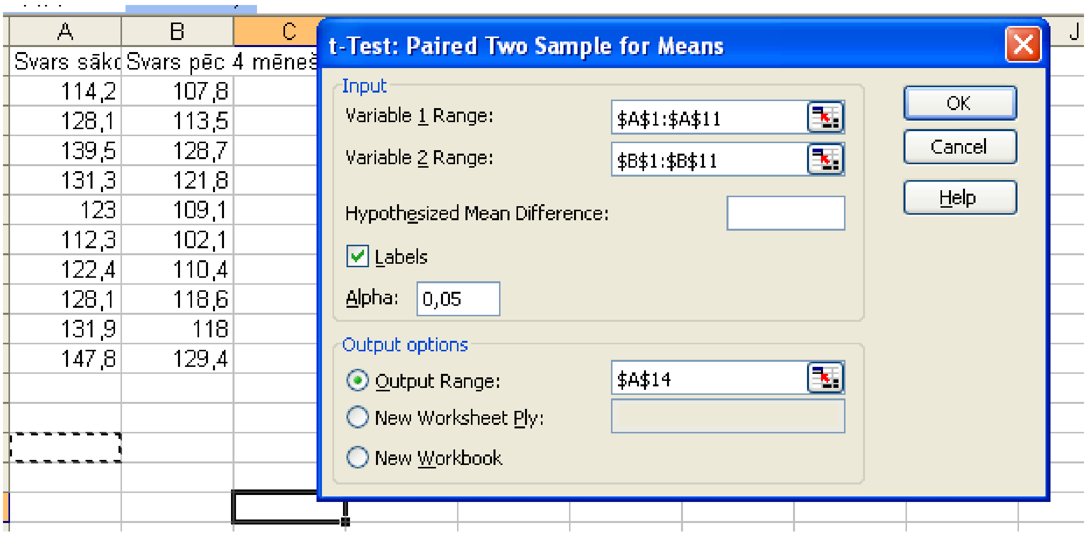
Tagad šis uzdevums tiks *veikts* ar *Excel*.

1. solī izvēlas datu analīzes rīku *„t-test Paired Two Sample for Means"* (5.12. attēls).



5.12. attēls. **Dialoga logs t-testa veikšanai pa pāriem saistītajiem novērojumiem**

Nākamajā 5.13. attēlā ir parādīts *Excel* darba lapas fragments ar aizpildītu dialoga logu.



5.13. attēls. **Nākamais solis t-testam pa pāriem saistītajiem novērojumiem**

5.9. tabula

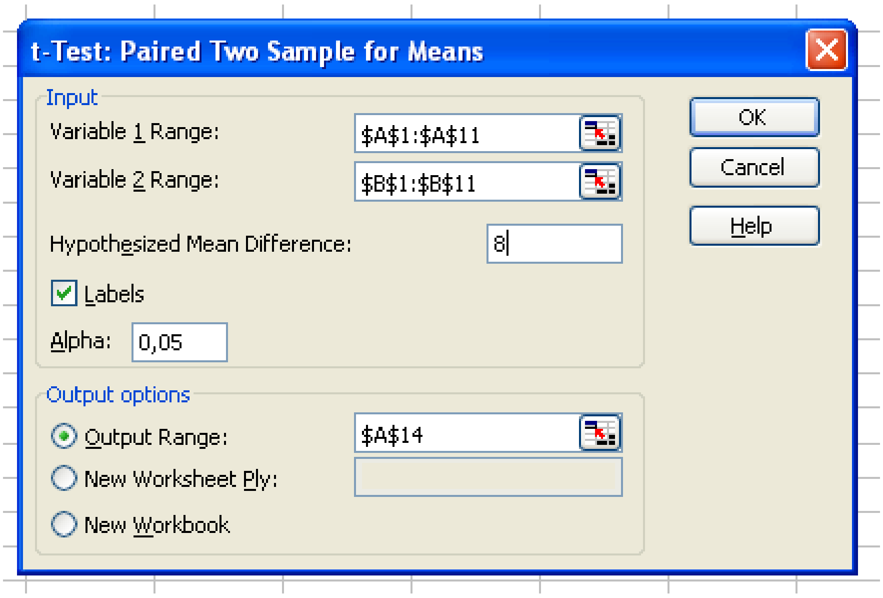
Ar *Excel* datu analīzes rīku iegūtā tabula (ar tulkojumiem un komentāriem)

t-Test: Paired Two Sample for Means

		Svars sākot programmu, kg	*Svars pēc 4 mēnešiem, kg*
Aritmētiskais vidējais	*Mean*	*127,86*	*115,94*
Dispersija	*Variance*	*115,9448889*	*81,20933*
Novērojumu skaits	*Observations*	*10*	*10*
Pīrsona korelācijas koeficients, skaidrojums būs nodaļā par korelāciju	*Pearson Correlation*	*0,956682634*	
Pārbaudāmo vidējo starpība	*Hypothesized Mean Difference*	*0*	
Brīvības pakāpes	*df*	*9*	
Empīriskā kritērija vērtība	*t Stat*	*11,11997068*	
Vienpusējas hipotēzes varbūtība, ka nulles hipotēze ir pareiza	*P(T<=t) one-tail*	*7,34553E-07*	
Vienpusējas hipotēzes kritērija kritiskā (teorētiskā) vērtība	*t Critical one-tail*	*1,833112923*	
Divpusējas hipotēzes varbūtība, ka nulles hipotēze ir pareiza	*P(T<=t) two-tail*	*1,46911E-06*	
Divpusējas hipotēzes kritērija kritiskā vērtība	*t Critical two-tail*	*2,262157158*	

Iepriekšējā tabulā var atrast manuālajos aprēķinos iegūtos skaitļus t_{emp} (*t Stat*)=11,12>t_{krit} (*t Critical two tail*)=2,26, tātad nulles hipotēzi var pārliecinoši noraidīt, svara samazināšanas metode ir efektīva. Ja nulles hipotēzi iespējams noraidīt divpusējai hipotēzei, tad var nedomāt, vai ir vienpusējā, vai divpusējā hipotēze. Šajā gadījumā pietiktu arī ar vienpusējo hipotēzi, jo ir doma pierādīt, ka eksperimenta dalībnieku svars pēc 4 mēnešiem ir samazinājies. Varbūtība, ka tas ir nevis likumsakarīgi, bet nejaušas izmaiņas un svara samazināšanas metodei nav būtiskas ietekmes, ir 7 gadījumi no 10 miljoniem mēģinājumu (7,34553E-07).

Ar *Excel* var pārbaudīt, vai metode garantē svara samazināšanos par 8 kilogramiem. Ir zināms, ka vidējais svara samazinājums grupā bija 11,92 kg. Kāda daļa būs nejaušas svārstības, bet daļu var pierādīt, kā svara nomešanas metodes radītu. Aprēķinus var atkārtot vairākkārt, līdz atrod maksimālo pierādāmo vērtību.



5.14. attēls. **Dialoga logs kādas iepriekš definētas starpības pierādīšanai**

Nākamajā 5.10. tabulā ir iegūtie rezultāti.

5.10. tabula

Excel ģenerētā rezultātu tabula, ja pārbauda vismaz 8 kg svara samazinājumu 5.4.piemēram

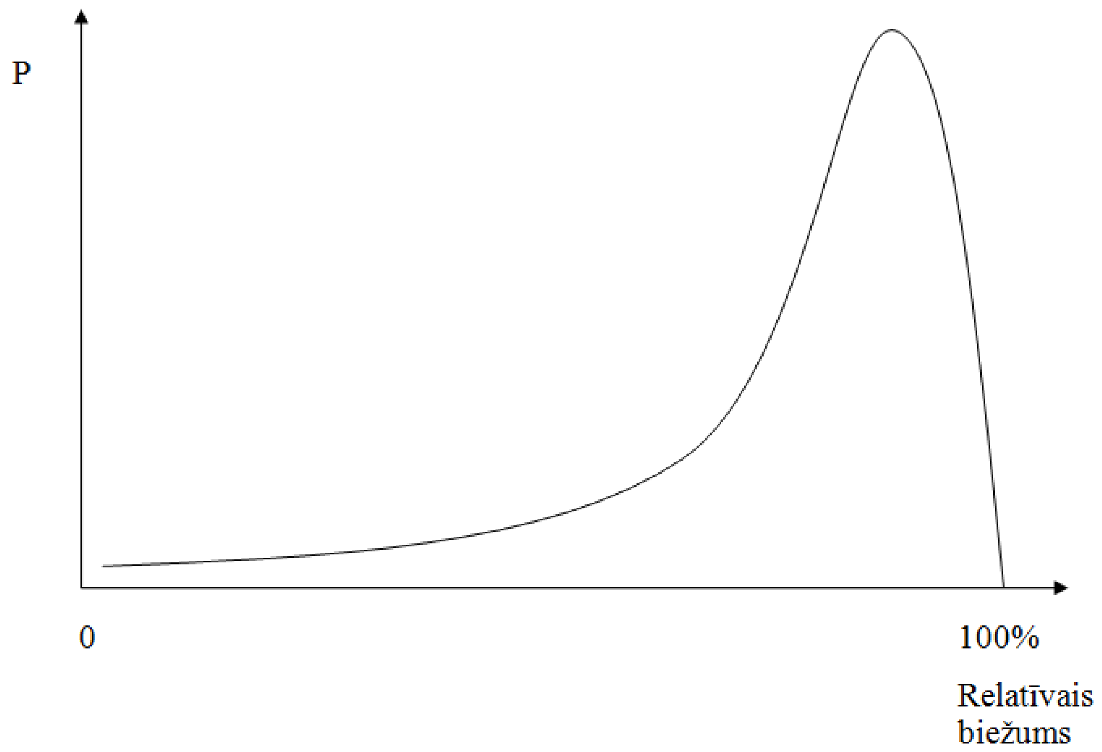
t-Test: Paired Two Sample for Means

	Svars, sākot programmu, kg	*Svars pēc 4 mēnešiem, kg*
Mean	*127,86*	*115,94*
Variance	*115,9448889*	*81,20933*
Observations	*10*	*10*
Pearson Correlation	*0,956682634*	
Hypothesized Mean Difference	*8*	
df	*9*	
t Stat	*3,65690311*	
P(T<=t) one-tail	*0,002630387*	
t Critical one-tail	*1,833112923*	
P(T<=t) two-tail	*0,005260774*	
t Critical two-tail	*2,262157158*	

Kā redzams, sākuma daļa ir tāda pati kā 5.8. tabulā. Atšķiras aprēķinātās empīriskā kritērija vērtības. Taču *t Stat* = 3,66 > *t Critical one-tail* = 1,83 un arī 8 kilogramu samazinājums ir droši pierādīts. Tas, ka atšķirības varētu būt nejaušas, varbūtība ir palielinājusies un ir 0,263 %, bet tā ir mazāka par iepriekš definēto kļūdas varbūtību α = 0,05.

5.3.5. Nulles hipotēzes pārbaude par divu relatīvo biežumu starpību

Relatīvie biežumi ir jāizmanto, kad salīdzina dažāda apjoma kopas. Piemēram, ja klientus apkalpojošs uzņēmums (banka, tūrisma birojs, sadzīves pakalpojumu uzņēmums u. tml.) salīdzina divus klientu apkalpošanas centrus, par salīdzināšanas kritēriju var izvēlēties apmierināto klientu īpatsvaru. Ja salīdzina skaitļu absolūtās vērtības, tad vismaz teorētiski tie var iegūt jebkuru vērtību. Relatīvais biežums ir ierobežots, un tas var svārstīties no 0 līdz 1, turklāt ļoti bieži salīdzināmā vērtība ir tuva kādai no šīm robežvērtībām (apmierināti ir lielākā daļa klientu, savukārt saražotā brāķa īpatsvars kopējā ražošanas apjomā ir relatīvi neliels). Šādā situācijā problēma ir tā, ka no vienas ģenerālkopas izveidoto izlašu vidējo biežumu sadalījums būs asimetrisks. Piemēram, ja ģenerālkopā apmierināto klientu īpatsvars ir 95 %, tad no šīs kopas izveidotajās izlasēs visbiežāk šāda vidējā vērtība arī parādīsies. Vidējās vērtības palielinājuma pusē ir tikai 5 %, kamēr samazinājuma pusē tā ir 95 %. Ļoti, ļoti niecīga ir varbūtība, ka izlasē būs tikai 5, 10 vai 15 % apmierināto klientu, taču tāda tomēr pastāv. 5.15. attēlā ir parādīta problēmas grafiskā interpretācija.



5.15. attēls. **Relatīvo biežumu sadalījums izlasēm, kas veidotas no ģenerālkopas ar augstu relatīvo biežumu**

Ja sadalījums nav normāls, to normalizē (veic sākotnējo datu matemātiskos pārveidojumus tā, ka pārveidotie dati veido aptuveni normālo sadalījumu) un tad izmanto metodes, kas paredzētas normālajam sadalījumam.

R. Fišers ir izstrādājis un piedāvā normalizācijas funkciju:

$$\varphi = 2 * \frac{\pi}{180} \arcsin \sqrt{p}\,, \qquad (5.27.)$$

kur φ *(fī)* – sākotnējo datu transformētā vērtība;
π – konstante – 3,14....;
p – relatīvais biežums (notikuma statistiskā varbūtība, kas iegūta empīrisko novērojumu rezultātā).

Salīdzinot divus relatīvos biežumus, veic šādus soļus:

1) formulē nulles un alternatīvo hipotēzi – H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$ vai H_1: $p_1 > p_2$ un pārbaudāmo hipotēzi – H_0: $\varphi_1 = \varphi_2$; H_1: $\varphi_1 \neq \varphi_2$ vai H_1: $\varphi_1 > \varphi_2$;
2) datiem veic Fišera transformāciju. Šos aprēķinus var izdarīt ar zinātnisko kalkulatoru vai *Excel* funkcijām vai atrast statistikas mācību grāmatas pielikumos, kur šīs vērtības ir tabulētas;
3) aprēķina transformēto vidējo vērtību starpības standartkļūdu:

$$s_{\varphi_1-\phi_2} = \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}\ , \tag{5.28.}$$

kur n_1 un n_2 ir pārbaudāmo izlašu apjomi vienībās;

4) aprēķina empīrisko Stjūdenta kritēriju:

$$t_{emp} = \frac{|\phi_1 - \phi_2|}{s_{\phi_1-\phi_2}} \tag{5.29.}$$

Aprēķinot t_{emp}, nav svarīgi, kuras izlases relatīvie biežumi un to transformētā vērtība ir lielāki, tāpēc formulā starpībai ir modulis, bet praksē vienkārši no lielākā skaitļa atņem mazāko;

5) izvēlas hipotēzes pārbaudes nozīmības līmeni α, nosaka brīvības pakāpju skaitu $\nu = n_1 + n_2 - 2$ un Stjūdenta kritērija tabulās nolasa t_{krit};

6) izdara secinājumus, līdzīgi kā citās kritēriju metodēs – ja $t_{emp} > t_{krit}$, tad nulles hipotēzi noraida, relatīvo biežumu starpība ir būtiska, ja $t_{emp} < t_{krit}$, tad nulles hipotēzi noraidīt nevar un atšķirība starp relatīvajiem biežumiem nav pierādīta.[1,2,3]

5.5. piemērs. Tiek salīdzinātas divas viesnīcas vienā uzņēmumā. Veikta klientu aptauja, vai viņi labprāt atgrieztos apmeklētajā viesnīcā vēlreiz. 1. viesnīcas 116 klienti no kopā aptaujātajiem 130 atbildēja apstiprinoši (p_1=116/130=0,89). 2. viesnīcā tika aptaujāti 180 klienti, no tiem atkārtoti atgrieztos 167 klienti (p_2 =0,93).

Ar nozīmības līmeni 0,05 jāpārbauda, vai otrās viesnīcas klienti ir apmierinātāki ar viesnīcas servisu un labprātāk atgrieztos šajā viesnīcā vēlreiz.

$p_1=0{,}89 \rightarrow \varphi_1=2{,}4655$ *un* $p_2=0{,}93 \rightarrow \varphi_2=2{,}6061$

$$s_{\varphi_1-\phi_2} = \sqrt{\frac{130+180}{130*180}} = 0{,}115,$$

$$t_{emp} = \frac{|2{,}4655 - 2{,}6061|}{0{,}115} = 1{,}222,$$

$t_{0,05;308} = 1{,}97$

Nulles hipotēzi noraidīt nevar, atšķirība klientu apmierinātības līmenī var būt nejaušība.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 136. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 391.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 130. lpp.

Statistiskās hipotēzes pārbauda divu kopu līdzību vai atšķirību. Bieži ir vairāk grupu, kuras ir izveidotas pēc kādas būtiskas pazīmes, kas ietekmē datu variāciju. Šādus uzdevumus risina ar dispersijas analīzi, par kuru tiks runāts nākamajā tēmā.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
Alternatīvā hipotēze	Research hypothesis	Альтернативная гипотиза	Pieņēmums, ka pētāmo kopu apskatāmie parametri ir atšķirīgi
Brīvības pakāpes	Degree of freedom	Степени свободы	Novērojumu skaits mīnus saistošo nosacījumu skaits
Divpusējā hipotēze	Two tailed hypothesis	Двухстороняя гипотиза	Pieņēmums, ka kopas var būt abpusēji atšķirīgas
F kritērijs	F test	Критерий Фишера	Divu kopu dispersiju salīdzināšanas metode
Hī kvadrāta kritērijs	Chi - squared test	Критерий *хи* квадрат	Metode, ar kuras palīdzību var novērtēt empīriskā sadalījuma atbilstību teorētiskajam, kā arī salīdzināt empīriskās kopas, ja tās nevar salīdzināt ar parametriskajām metodēm
Kolmogorova - Smirnova kritērijs	Kolmogoroff - Smirnow test	Критерии Колмогорова - Смирнова	Divu kopu salīdzināšanas metode, balstās uz lielāko frekvenču starpību
Nozīmības līmenis	Significance levei (alpha)	Уровень значимости	Maksimālā kļūdas varbūtība, ja noraida nulles hipotēzi, bet tā izrādās pareiza
Nulles hipotēze	Null hypothesis	Нулевая гипотиза	„*Status quo*" apliecinājums, atšķirība starp kopām nav statistiski nozīmīga, ir nejauša.
Nulles hipotēzes noraidīšana	Reject null hypothesis	Отклонение нулевой гипотизы	Noraidot nulles hipotēzi, pieņem alternatīvo – atšķirības starp pētāmajām kopām ir statistiski nozīmīgas
Standart-kļūda	Standard error	Стандартная ошибка	Potenciālo izlašu vidējo izkliedes rādītājs ap nezināmo ģenerālkopas vidējo, ļauj novērtēt ticamības intervālus
Starpības reprezentāci-jas kļūda	Standard error of difference	Ошибка репрезентации разности	Divu aritmētisko vidējo apvienotā standartkļūda
Stjūdenta kritērijs	Student's test(t- test)	Критерий Стюдента	Plaši izmantota parametriskā divu kopu salīdzināšanas metode, balstās uz Stjūdenta sadalījumu
Vienpusējā hipotēze	One tailed hypothesis	Одностороняя гипотиза	Pārbauda tikai vienu iespējamo atšķirību virzienu. Kuras izlases raksturotājs ir lielāks, tās kopas parametrs arī ir lielāks

6. DISPERSIJAS ANALĪZE

Pēc nodaļas apgūšanas studentiem:

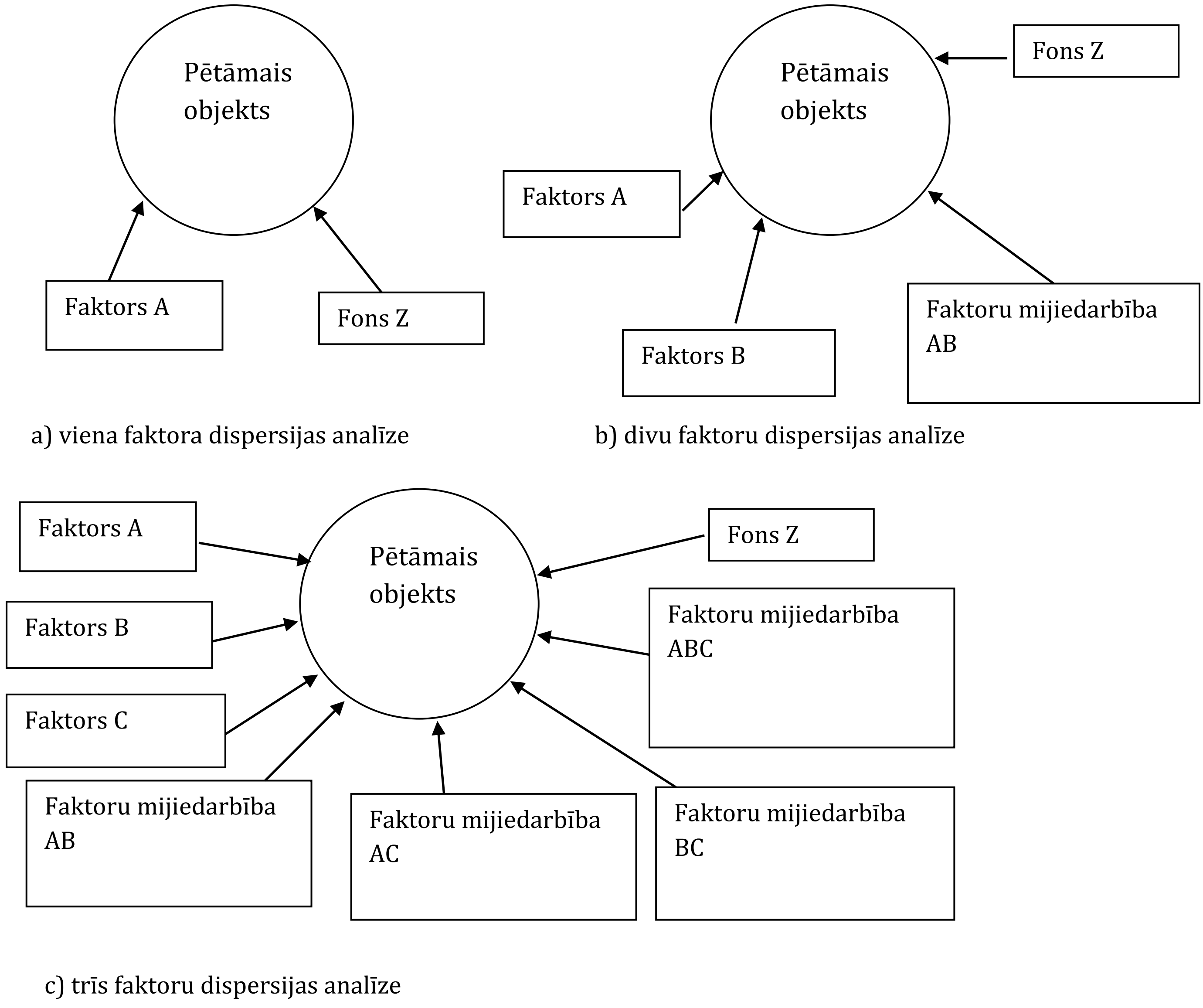
- *jāzina dispersijas analīzes būtība un lietošanas nosacījumi;*
- *jāzina dispersijas analīzes soļi;*
- *jāprot lietot dispersijas analīzes formulas un interpretēt iegūtos rezultātus;*
- *jāprot veikt dispersijas analīzi ar Microsoft Excel datu analīzes rīku.*

6.1. Ievads

Dispersijas analīzes metodi 20. gadsimta 20. gados izstrādāja angļu biologs un matemātiķis R. Fišers. Tās rašanos var saistīt ar zināmu tā laika lauksaimniecības "pasūtījumu". Attīstoties lauksaimniecībai, bija nobriedusi vajadzība izpētīt, kuri ir būtiskie ražu ietekmējošie faktori – agroklimatiskie apstākļi, kultūraugu šķirnes, mēslošanas kombinācijas u. tml. Protams, bija jāatrod optimālais resursu un citu kultūraugu ražību ietekmējošo faktoru kombināciju variants, bet, lai sekmētos meklēšana, bija jānoskaidro faktori, kas būtiski ietekmē rezultējošo pazīmi – lauksaimniecības kultūru ražas. Jaunā metodika izrādījās universāla, un ātri vien tā tika pielietota visdažādākajās jomās, tajā skaitā arī izvērtējot dažādas ekonomikas un uzņēmējdarbības problēmas. Ļoti plašas iespējas pielietot dispersijas analīzi ir dažādos tehnoloģiskos pētījumos (lauksaimniecībā, rūpniecībā, pakalpojumos). Piemēram, var izvērtēt izglītības nozīmi darba ražīgumā, lietoto tehnoloģiju, darba galdu, dažādu izejvielu piegādātāju ietekmi uz saražotās produkcijas apjomu un kvalitāti u. c.

"Dispersijas analīze" ir gadījums, kad nosaukums atklāj saturu. Dispersija ir datu izkliede. Kā jau statistikas kursa sākumā tika noskaidrots – statistikā pēta tikai variējošus datus. To, ka pētāmā objekta elementi ir ar dažādām vērtībām, nodrošina apkārtējās vides ietekme. No šīs vides analīzes rezultātiem var izdalīt dažas pazīmes, kuru variācija (kvalitatīva vai kvantitatīva) ietekmē pētāmā objekta elementu variēšanu. Piemēram, strādnieku darba ražīgums – saražotās produkcijas apjoms laika vienībā variē, daži strādnieki var izdarīt vairāk, citi mazāk, arī vienam strādniekam izstrāde variē pa dienām. Nekad nevar identificēt visus rezultējošo pazīmi ietekmējošos faktorus, bet cenšas izdalīt būtiskākos, lai vēlāk ar lēmumiem spētu ietekmēt un panākt vēlamo rezultātu. Var izvirzīt vairākas hipotēzes, piemēram, darba ražīgums ir atkarīgs no stāža vai arī no apmeklēto kvalifikācijas celšanas kursu ilguma u.tml. Ja izdalīts ir viens faktors, tad ir viena faktora dispersijas analīze. Ja identificēti vairāki faktori (divi, trīs), tad ir daudzfaktoru dispersijas analīze. Ar izdalīto faktoru variāciju saistīto rezultējošās pazīmes izkliedi sauc par faktoriālo vai izskaidroto dispersiju. Pārējo (neidentificēto) faktoru ietekmi apzīmē par fona ietekmi, un ar to saistīto rezultējošās pazīmes datu izkliedi sauc par neizskaidroto dispersiju.[1,2,3]

Faktoru un pētāmā objekta mijiedarbības grafiskā interpretācija ir parādīta 6.1. attēlā.



6.1. attēls. **Faktoru un pētāmā objekta mijiedarbība dispersijas analīzē**

Izdalītos apkārtējās vides faktorus apzīmē ar lielajiem latīņu alfabēta burtiem – “A”, “B” utt. Visu pārējo (neizskaidroto, neidentificēto) faktoru ietekmi

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 136. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 179. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 724.

sauc par fona ietekmi un apzīmē ar latīņu burtu Z. Divu faktoru dispersijas analīzē identificē un analizē divu faktoru (piemēram, agroklimatiskie apstākļi, šķirne) ietekmi uz rezultējošo pazīmi, turklāt izdalītajiem faktoriem arī ir mijiedarbība (piemēram, kāda šķirne ir piemērotāka sausiem, bet cita mitriem laika apstākļiem). Jo vairāk faktoru izdala, jo sarežģītāka kļūst analīze. Kā redzams 6.1. c) attēlā, tad trīs faktoru dispersijas analīzē ir trīs faktori (ABC), kas ietekmē rezultātu, turklāt trīs pirmā līmeņa (starp diviem faktoriem) mijiedarbības (AB, AC, BC) un viena otrā līmeņa (starp visiem trim faktoriem) mijiedarbība (ABC).

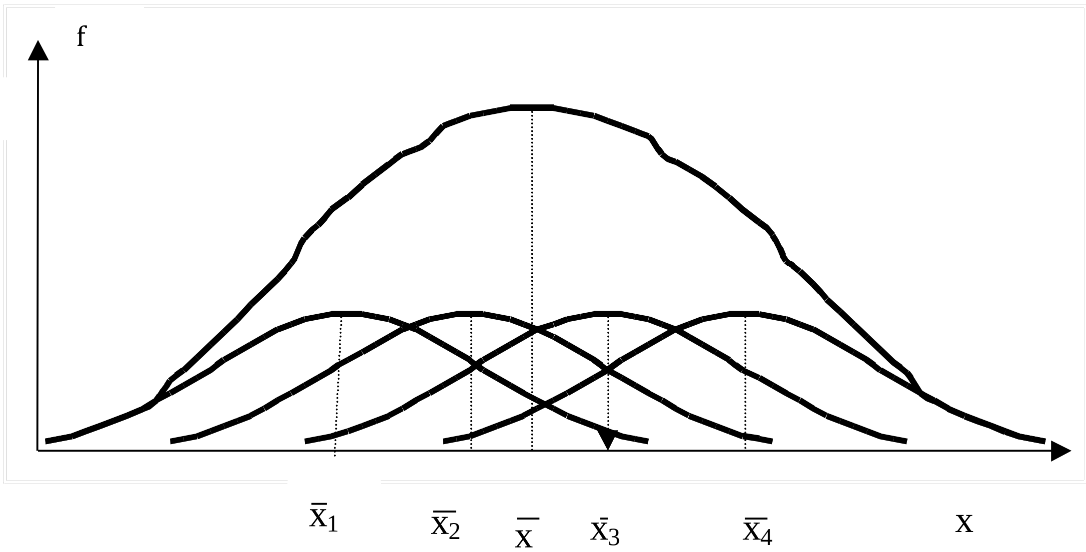
Pētniekam vilinoši ir izdalīt, identificēt pēc iespējas vairāk faktoru, tādējādi labāk izprotot pētāmā objekta variāciju. Taču pārāk daudz dažādu faktoru iekļaušana analīzē parasti neattaisnojas. Bez tā, ka aprēķini kļūst stipri sarežģīti, iegūtais modelis bieži vien nav stabils. Vienkāršākiem modeļiem (viena vai divu faktoru dispersijas analīze), atkārtojot novērojumus (veicot atkārtotu eksperimentu), rezultātu būtība saglabājas (ir tikai nelielas variācijas). Sarežģītiem modeļiem kāda faktora vai faktoru mijiedarbības ietekme atkārtotā pētījumā var būt pilnīgi pretēja sākotnējiem rezultātiem – izstrādātais modelis nav stabils.

Faktoru identificēšana un izdalīšana nav statistikas uzdevums, ar to nodarbojas zinātnes nozarē, kurā izmanto statistikas metodes.

Visbiežāk tiek veikta viena faktora dispersijas analīze.

6.2. Viena faktora dispersijas analīzes būtība un manuālie aprēķini

Dispersijas analīzes gaitā pētāmā objekta vērtības sagrupē pa faktora pazīmes grupām un tad aprēķinu ceļā novērtē, vai faktoriālās pazīmes variācija ir tā, kas būtiski ietekmē rezultējošās pazīmes variāciju, vai arī fona ietekme ir lielāka. Viena faktora dispersijas analīzes būtību var parādīt grafiski plaknē (6.2. attēls).



6.2. attēls. **Viena faktora dispersijas analīzes grafiskā interpretācija**

6.2. attēlā ir parādīts, kā pazīmes vērtība variē visas kopas ietvaros (lielā līkne) un izdalīto grupu ietvaros (mazās līknes). Kā redzams, tad pazīmes vērtības variē arī grupu ietvaros un ievērojama daļa no pazīmju vērtībām pārklājas. Apakšgrupa x_1 uzrāda mazākās vērtības, kaut arī grupas lielākās vērtības ir tādas kā x_4 grupas mazākās vērtības.

Datu izkliedē izdala trīs dispersijas:

- s^2 – kopējā dispersija, raksturo datu izkliedi ap kopējo vidējo;
- s_A^2 – izskaidrotā (faktoriālā) dispersija, kas ir grupu vidējo izkliede ap kopējo vidējo. Rezultējošās pazīmes izkliede tiek izskaidrota ar faktoriālās pazīmes variāciju;
- s_Z^2 – neizskaidrotā dispersija, raksturo datu izkliedi ap grupu vidējiem.[1,2]

Lai veiktu aprēķinus manuāli, empīriskos datus vispirms ieraksta tabulā, kuru sauc par **statistisko kompleksu**. Tabulas iekārtojums ir atkarīgs no vienlaicīgi analizējamo faktoru skaita.

Ja visās gradācijas klasēs ir vienāds varianšu skaits, tad ir **homogēns statistiskais komplekss**, bet, ja dažāds, tas ir **heterogēns statistiskais komplekss**.[3]

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 179. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 136. lpp.
[3] Liepa, I. (1974). Biometrija : mācību līdzeklis augstskolu studentiem. Rīga: Zvaigzne. 104. lpp.

Dispersijas analīzi veic šādos soļos:

1) noviržu kvadrātu summas sadalīšana komponentēs;
2) brīvības pakāpju skaita sadalīšana komponentēs;
3) dispersiju aprēķināšana;
4) F_{emp} aprēķināšana;
5) F_{krit} atrašana tabulās un iegūto rezultātu interpretācija.[1,2]

Manuālajiem aprēķiniem darbietilpīgākais un sarežģītākais ir pirmais solis – noviržu kvadrātu summu aprēķināšana.

Noviržu kvadrātu summu būtības izprašanai der apskatīt 6.2. attēlu un tā skaidrojumu.

Noviržu kvadrātu summas apzīmē ar lielo latīņu burtu Q.

Kopējā noviržu kvadrātu summa raksturo datu izkliedi ap kopējo vidējo, kas formulas veidā ir:

$$Q = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2 , \qquad (6.1.)$$

kur i – gradācijas klases;
k – gradācijas klašu skaits;
j – novērojuma numurs gradācijas klasē;
n_i – novērojumu skaits i gradācijas klasē;
x_{ij} – i klases j novērojums;
$\sum_{i=1}^{k}$ – summēšana pa grupām (izdalītajām gradācijas klasēm);
$\sum_{j=1}^{n_i}$ – summēšana gradācijas klases ietvaros no pirmā līdz pēdējam (n_i) novērojumam i klasē.

Izskaidrotā noviržu kvadrātu summa raksturo grupu vidējo novirzi no kopējā vidējā:

$$Q_A = \sum_{i=1}^{k} \left(\overline{x_i} - \overline{x}\right)^2 * n_i , \qquad (6.2.)$$

kur $\overline{x_i}$ – i grupas vidējais;
Q_A – izskaidrotā noviržu kvadrātu summa.

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 180. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 732.

Neizskaidrotā noviržu kvadrātu summa ir datu izkliede ap grupu vidējiem, ko apzīmē ar Q_Z:

$$Q_Z = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x}_i\right)^2 \qquad (6.3.)$$

Manuālajos aprēķinos ir ļoti svarīgi vienkāršot aprēķinus. No šī apsvēruma arī izriet bieži lietotais Q_Z nosaukums – atlikuma noviržu kvadrātu summa. Zinot to, ka kopējā noviržu kvadrātu (Q) summa ir izskaidroto (Q_A) un neizskaidroto (Q_Z) noviržu kvadrātu summa un ir aprēķināts Q un Q_A, var viegli aprēķināt Q_Z kā atlikumu:

$$Q_Z = Q - Q_A, \qquad (6.4.)$$

Tas, ko neizskaidro grupējums, ir fona noviržu kvadrātu summa.[1,2,3]

Sakarā ar to, ka metode radās, kad nebija pieejama efektīva skaitļošanas tehnika, ir izstrādātas momentu formulas. Šīs formulas vizuāli šķiet sarežģītākas par noviržu formulām (6.1., 6.2., 6.3.), tās arī neatklāj pētīšanas būtību, bet ievērojami samazina nepieciešamo kalkulāciju apjomu.

Tālāk ir dotas momentu formulas noviržu kvadrātu summu aprēķināšanai:

$$Q = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - C \qquad (6.5.)$$

Vārdiem izsakot, šī formula ir visu datu kvadrātu summa mīnus korekcijas loceklis, ko savukārt aprēķina pēc 6.6.formulas:

$$C = \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{N}, \qquad (6.6.)$$

kur skaitītājā ir visu datu summas kvadrāts un saucējā ir kopējais novērojumu skaits visās gradācijas klasēs ($N = \sum_{1}^{k} n_i$).

Jāatceras, ka kvadrātu summa nav tas pats, kas summas kvadrāts.

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 724.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 139. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 181. lpp.

Veicot aprēķinus bez datora datu analīzes rīka iespējām, ir jāsaprot nepieciešamo aritmētisko darbību secība.

Ilustrācijai mazs piemērs. $x_1 = 2$ un $x_2 = 3$.

$$\sum_{i=1}^{n} x_i^2 = 2^2 + 3^2 = 4 + 9 = 13 \neq \left(\sum_{i=1}^{n} x_i \right)^2 = (2+3)^2 = 5^2 = 25$$

Izskaidroto noviržu kvadrātu summu aprēķina pēc 6.7. formulas:

$$Q_A = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} x_{ij} \right)^2}{n_i} - C \qquad (6.7.)$$

jeb vārdiem tas būtu: grupu datu summu kvadrātus dala ar novērojumu skaitu grupā. Šos dalījumus summē un atņem korekcijas locekli.

Ja ir homogēns statistiskais komplekss – visās grupās ir vienāds novērojumu skaits ($n=n_1 = n_2 = ... = n_i$), tad $1/n$ var iznest ārpus summēšanas (iekavām), un formula iegūst šādu veidu:

$$Q_A = \frac{1}{n} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} x_{ij} \right)^2 - C \qquad (6.8.)$$

Q_Z aprēķiniem arī ir momentu formula (6.9.formula), bet manuālajos aprēķinos to praktiski nekad neizmanto un Q_Z aprēķina pēc atlikuma principa (6.4. formula).

Q_Z momentu formula ir:

$$Q_Z = \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{j=1}^{n_i} x_{ij} \right)^2}{n_i} \right) \qquad (6.9.)$$

Vārdiem tas skan tā – no grupu datu kvadrātu summām atņem grupu datu summas kvadrātu, kas dalīts ar grupas novērojumu skaitu (grupas korekcijas loceklis). Tad summē pa visām grupām.[1,2]

Turklāt aprēķinu vienkāršošanai, rēķinot ar vienkāršu skaitļošanas tehniku (skaitīkļi, logaritmiskie lineāli, mehāniskās reizināšanas un dalīšanas mašīnas –

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 181.lpp.
[2] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 109. lpp.

Felikss, kalkulatori parādījās tikai 20. gadsimta 70. gados), plaši izmantoja dispersijas īpašību, kas ļauj no visiem novērojumiem atņemt vienu lielumu (konstanti), un dispersija nemainās. Šādi aprēķinātas variantes dēvēja par kodētām variantēm. Neskatoties uz to, ka mūsdienās visiem ir kalkulatori, varianšu kodēšana paātrina aprēķinu veikšanu un samazina kļūdīšanās iespējas, veicot aprēķinus manuāli. Lai izprastu dispersijas aprēķinu būtību, formulas, manuālo aprēķinu metodiku dots 6.1. piemērs.

6.1. piemērs. Lauksaimniecības tehnoloģiskajos pētījumos tiek novērtētas 4 rapša mēslošanas kombinācijas, kuras atšķiras ar mēslošanas laiku un devu sadalījumu. Katrai metodei bija 5 izmēģinājuma lauki 0,5 ha platībā katrs. Iegūti šādi rezultāti, pārrēķinot ražību cnt/ha:

1. metode	2. metode	3. metode	4. metode
18 cnt/ha	20	24	15
19	23	25	20
21	19	22	18
20	21	27	17
22	22	24	16

Kā redzams, rezultāti svārstās ap 20 cnt/ha, tāpēc kodēšanai par konstanti izvēlas 20 (no katras variantes vērtības atņem 20. Kodētās variantes var būt arī negatīvi skaitļi, svarīgi, lai tos var viegli kāpināt kvadrātā). Kodētās variantes ieraksta darba tabulā, aprēķina šo varianšu kvadrātus un grupu vidējos, kā arī kopējo vidējo.

6.1. tabula

Darba tabula viena faktora dispersijas analīzei, lietojot momentu formulas

Gradācijas klases	$x_{ij} - 20$	$(x_{ij} - 20)^2$	*Vidējie*
A_1 (1. metode)	-2 -1 1 0 2	4 1 1 0 4	
Σ (kopā klasē)	0	10	$\overline{x_1} = 0/5 + 20 = 20$
A_2	0 3 -1 1 2	0 9 1 1 4	
Σ (kopā klasē)	5	15	$\overline{x_2} = 5/5 + 20 = 21$
A_3	4 5 2 7 4	16 25 4 49 16	
Σ (kopā klasē)	22	110	$\overline{x_3} = 22/5 + 20 = 24{,}4$
A_4	-5 0 -2 -3 -4	25 0 4 9 16	
Σ (kopā klasē)	-14	54	$\overline{x_4} = -14/5 + 20 = 17{,}2$
Σ Σ (visi novērojumi kopā)	13	189	$\overline{x} = 13/20 + 20 = 20{,}65$

Kā redzams darba tabulā, arī aritmētisko vidējo aprēķināšanai izmanto aritmētiskā vidējā īpašību, kas ļauj atņemt no visiem novērojumiem vienu konstantu skaitli. Starpībām aprēķina vidējo un, lai iegūtu labi saprotamus rezultātus sākotnējā datu mērogā, aprēķinātajam vidējam pieskaita atņemto konstanti.

Ja aprēķiniem lietotu noviržu formulas, tad katrai variantei būtu jārēķina novirze no kopējā vidējā, no grupas vidējā, iegūtās novirzes jākāpina kvadrātā. Kalkulāciju būtu ievērojami vairāk. Jāņem vērā – lai iegūtu statistiski nozīmīgus rezultātus, izlases ir krietni lielākas, mērījumu skaitļi ir precīzāki (tajos ir vairāk ciparu). Apskatītajā piemērā vidējie dalās precīzi, bet bieži vien vidējie nav precīzi skaitļi, un tad rodas vēl papildus noapaļošanas kļūda. Šie apsvērumi ir pamatā tam, kāpēc manuālajiem aprēķiniem izmanto momentu formulas.

Kad ir sastādīta darba tabula un aprēķinātas vajadzīgās summas, tad tās liek noviržu kvadrātu summu aprēķināšanas formulās (6.6., 6.5., 6.7., 6.4).

$C = 13^2/20 = 8{,}45$

$Q = 189 - 8{,}45 = 180{,}55$

$Q_A = 1/5(0^2 + 5^2 + 22^2 + (-14)^2) - 8{,}45 = 132{,}55$

$Q_z = 180{,}55 - 132{,}55 = 48$

Neviena no noviržu kvadrātu summām nevar būt negatīva. Ja kāda aprēķināto kvadrātu summa ir iznākusi ar mīnus zīmi, tad tas liecina par kļūdu aprēķinos.

Turpmākos dispersijas analīzes soļus izpilda dispersijas analīzes tabulā. [1,2]

6.2. tabulā ir parādīti apzīmējumi un formulas, kā aprēķināt attiecīgajās vietās ierakstāmos lielumus, bet 6.3. tabulā ir 6.1. piemēra dati.

6.2. tabula

Dispersijas analīzes tabula: apzīmējumi un formulas

Izkliede	*Noviržu kvadrātu summas*	*Brīvības pakāpes*	*Dispersijas*	F_{emp}	F_{krit}
Faktora	Q_A	$\nu_A = k - 1$	$s_A^2 = Q_A/\nu_A$	s_A^2/s_Z^2	$F\alpha; \nu_A; \nu_Z$
Atlikuma	Q_Z	$\nu_Z = N - k$	$s_Z^2 = Q_Z/\nu_Z$		
Kopējā	Q	$\nu = N - 1$			

k – gradācijas klašu skaits; N – kopējais novērojumu skaits.

Brīvības pakāpēm, līdzīgi kā noviržu kvadrātus summai, ir jāpiepildās nosacījumam: $\nu = \nu_A + \nu_Z$. No tā izriet, $\nu_Z = \nu - \nu_A = N - 1 - (k - 1) = N - k$.

Dispersija ir noviržu kvadrātu summa, dalīta ar brīvības pakāpju skaitu (5.13. formula). Kopējo dispersiju arī var aprēķināt, bet tā nav nepieciešama tālākajiem aprēķiniem, tāpēc parasti to nedara.

Empīrisko Fišera kritēriju aprēķina, dalot faktoriālo dispersiju ar atlikuma dispersiju.

Kritisko Fišera kritēriju atrod tabulās, un to nosaka izvēlētais nozīmības līmenis α un brīvības pakāpju skaits dispersijām, kuras tika izmantotas empīriskā Fišera kritērija aprēķināšanai, respektīvi, ν_A ir tabulas ailēs norādītās brīvības pakāpes ν_1, bet ν_Z – tabulas rindās norādītās brīvības pakāpes ν_2 (kā to atrast, ir parādīts 5.7. attēlā).

[1] Liepa, I. (1974). Biometrija: mācību līdzeklis augstskolu studentiem. Rīga: Zvaigzne. 110. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 744.

6.3. tabula

Dispersijas analīzes tabula 6.1. piemēram

Izkliede	*Noviržu kvadrātu summas*	*Brīvības pakāpes*	*Dispersijas*	F_{emp}	F_{krit} $\alpha = 0{,}05$
Faktora	*132,55*	*3 (4 –1)*	*44,18*	*14,73*	*3,239*
Atlikuma	*48*	*16 (20 – 4)*	*3*		
Kopējā	*180,55*	*19 (20 – 1)*			

Galvenais secinājums tiek izdarīts, salīdzinot F_{emp} ar F_{krit}.

Tā kā F_{emp} ir vairākas reizes lielāks par F_{krit} (14,73>3,239), tad ar ļoti lielu pārliecību var noraidīt nulles hipotēzi un droši apgalvot, ka pētāmais faktors (rapša mēslošanas metodes) ir būtisks.

Tas nesniedz atbildi par to, kas ir labāks, vai pētāmais faktors izskaidro lielu vai mazu datu izkliedes daļu. Salīdzinājums F_{emp} un F_{krit} rāda tikai to, vai ar noteikto varbūtību atšķirība starp grupām ir pierādīta.

ν_Z nevajadzētu būt mazākam par 10. Ja atlikuma brīvības pakāpju ir mazāk par 10, tad vai nu jāpalielina novērojumu skaits, vai arī jāsamazina izdalīto gradācijas klašu skaits nepamatoti sīka grupējuma gadījumā, tās apvienojot.

Vēl var novērtēt, cik liels ir izskaidroto noviržu kvadrātu īpatsvars kopējā noviržu kvadrātu summā (6.4. tabula).

6.4. tabula

Izkliedes īpatsvara aprēķināšana

Izkliede	*Noviržu kvadrātu summas*	*Izkliedes īpatsvars*
Faktora	*132,55*	*73%*
Atlikuma	*48*	*27%*
Kopējā	*180,55*	*100%*

Kā redzams, ar pētāmo faktoru (mēslošanas variantiem) tiek izskaidroti 73% no visas datu izkliedes.

Vēl viens jautājums, kuru var atrisināt dispersijas analīzē, ir noteikt, starp kurām gradācijas klasēm pastāv būtiska atšķirība. Mūsdienās šo pārbaudi viegli var veikt, pārbaudot nulles hipotēzes, salīdzinot kopas pa pāriem. Manuālajiem aprēķiniem ir izstrādāta virkne kritēriju:

- Stjūdenta;
- Tjūkija;
- Dunkana;
- Ņūmana - Kjūla.

Šeit tiks apskatīts tikai viens no šiem kritērijiem (Stjūdenta) un izdarīti secinājumi par 6.1. piemēru. Dispersijas analīzē pieņem, ka gradācijas klašu vidējo aritmētisko reprezentācijas kļūdas ir vienādas:

$$s_{\overline{x_1}} = s_{\overline{x_2}} = \ldots = s_{\overline{x}} \text{, tad}$$

$$s_{\overline{x}} = \sqrt{\frac{s_Z^2}{n}}, \qquad (6.10.)$$

kur s_Z^2 – atlikuma dispersija;
n – novērojumu skaits gradācijas klasē (formula der homogēniem statistiskajiem kompleksiem – visās klasēs ir vienāds novērojumu skaits).

s_D jeb $s_{\overline{x}_1 - \overline{x}_2}$ ir starpības reprezentācijas kļūda, to aprēķina pēc 6.11. formulas:

$$s_D = \sqrt{s_{\overline{x_1}}^2 + s_{\overline{x_2}}^2} = \sqrt{2s_{\overline{x}}^2} = 1{,}41 s_{\overline{x}} \qquad (6.11.)$$

Kritisko starpības lielumu starp gradāciju klašu lielumiem aprēķina pēc 6.12. formulas:

$$\gamma_\alpha = t_{\alpha:\nu} s_D, \qquad (6.12.)$$

kur $t_{\alpha;\nu}$– kritiskā Stjūdenta kritērija vērtība, to nosaka izvēlētais nozīmības līmenis α un atlikuma brīvības pakāpju skaits ν_Z. Šo vērtību nolasa tabulās.[1]

6.1. piemēram aprēķini ir šādi:

$$s_{\overline{x}} = \sqrt{\frac{3}{5}} = 0{,}77$$

s_D = 1,41 $*$ 0,77 = 1,09.

Ja α izvēlas 0,05, ν_Z = 16, tad t_{krit} = 2,12

$\gamma_{0,05}$ = 2,12 $*$ 1,09 = 2,32 cnt/ha

Ja α izvēlas 0,01, ν_Z = 16, tad t_{krit} = 2,921

$\gamma_{0,01}$ = 2,921 $*$ 1,09 = 3,18 cnt/ha

Tālāk aprēķina un ieraksta tabulā starpības starp gradācijas klašu vidējiem aritmētiskajiem. Ja starpības absolūtā vērtība (nav svarīgi, kuras klases vidējais ir lielāks, novērtē, cik statistiski nozīmīga ir atšķirība) ir lielāka par kritisko vērtību 0,05 nozīmības līmenim, tad pasvītro ar vienu svītru, ja tā ir lielāka par starpību nozīmības līmenim 0,01, tad – ar divām svītrām.

[1] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 118. lpp.

6.5. tabula

Gradācijas klašu aritmētisko vidējo starpību tabula 6.1. piemēram

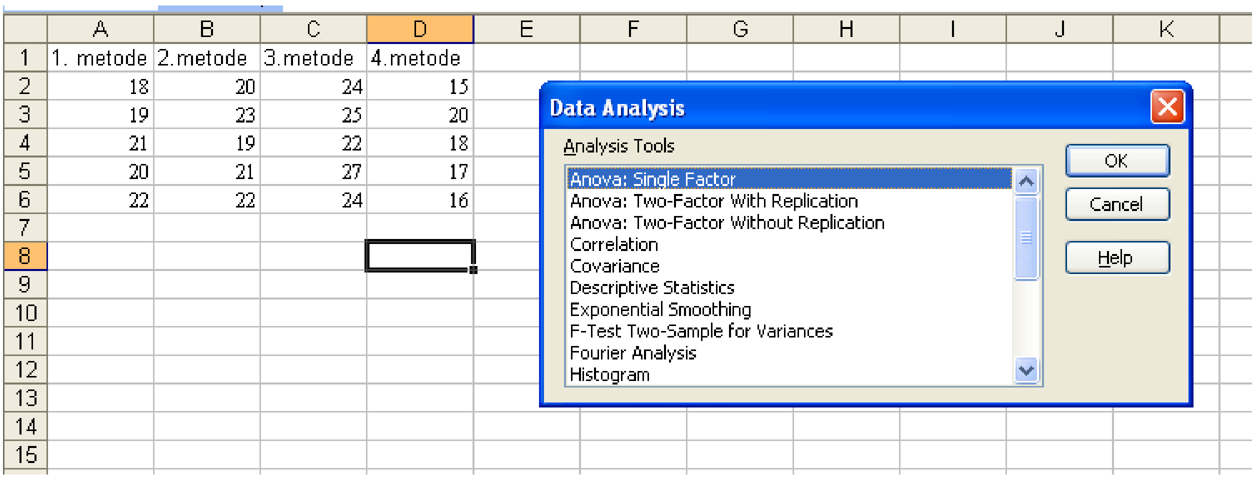
	1. metode $\overline{x}_1 = 20$ *cnt/ha*	*2. metode* $\overline{x}_2 = 21$ *cnt/ha*	*3. metode* $\overline{x}_3 = 24{,}4$ *cnt/ha*
2. metode $\overline{x}_2 = 21$ *cnt/ha*	*1 cnt/ha (21 – 20)*		
3. metode $\overline{x}_3 = 24, 4$ *cnt/ha*	*4,4 cnt/ha*	*3,4 cnt/ha*	
4. metode $\overline{x}_4 = 17{,}2$ *cnt/ha*	*2,8 cnt/ha*	*3,8 cnt/ha*	*7,2 cnt/ha*

Kā redzams no starpību tabulas, tad ar nozīmības līmeni 0,05 nevar noraidīt nulles hipotēzi starpībai starp 1. un 2.mēslošanas metodiku. Ja izvirza augstāku statistiskās nozīmības līmeni – $\alpha = 0{,}01$*, tad nulles hipotēzi nevar noraidīt arī starpībai starp 1. un 4. grādācijas klases vidējiem. Pārējās starpības starp gradācijas klasēm ir būtiskas arī ar šo augsto nozīmības līmeni.*

Svarīgi ir "pārtulkot" iegūtos statistiskos secinājumus sākotnējās hipotēzes tēzēs. Tātad mēslošanas laiki un devu kombinācija būtiski ietekmē rapša ražu, turklāt droši var apgalvot, ka 3. mēslošanas metodika ir pati labākā un to ieteikt ieviest ražošanā.

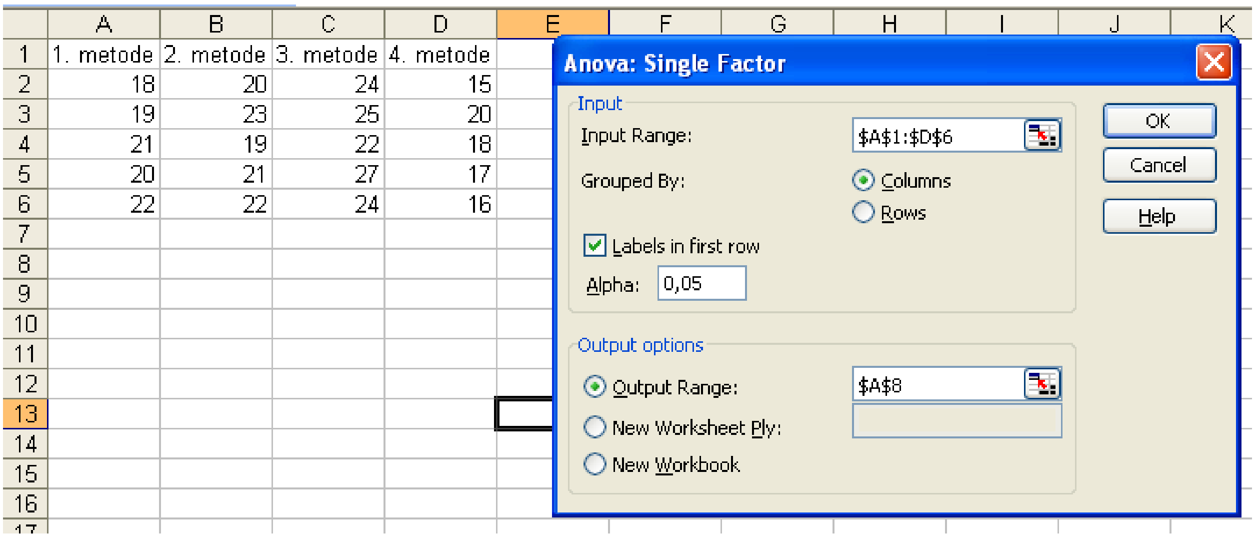
6.3. Viena faktora dispersijas analīze ar programmu *Excel*

Tagad 6.1. piemērs tiks izanalizēts ar mūsdienu metodēm *Microsoft Office Excel* programmā.



6.3. attēls. ***Excel*** **darba lapas fragments ar ievadītajiem datiem un "Data Analysis" dialoga logu**

Dispersijas analīze – *Analysis of variance*, saīsinājumā – *ANOVA*. Dialoga logā ir piedāvāti trīs dispersijas analīzes varianti. Tā kā šajā gadījumā vajadzīga viena faktora dispersijas analīze – "*Anova:Single Factor*", ir jāuzklikšķina ar peles kreiso taustiņu uz dotā analīzes veida un jānospiež poga "OK".



6.4. attēls. ***Excel*** **darba lapas fragments ar vienfaktora dispersijas analīzes dialoga logu**

Logā "*Input range*" norāda datu apgabalu. To var izdarīt, vai nu iedrukājot šūnu adreses no tastatūras, vai iezīmējot datu apgabalu ar datorpeli – kursoru novieto vienā datu apgabala stūrī, piespiež kreiso peles taustiņu un pārvieto kursoru pa diagonāli uz pretējo stūri, taustiņu atlaiž.

Dati var būt grupēti ailēs (*columns*) vai rindās (*rows*). Ja automātiskajā piedāvājumā grupējuma virziens (ailes vai rindas) ir nepareizs, tad kursoru pārvieto uz aplīti pretī pareizajam grupējumam un piespiež peles kreiso taustiņu, aplītī parādās zaļš punktiņš.

Datus labāk ievadīt ar to apzīmējumiem, lai vēlāk būtu vieglāk saprast aprēķinātos rezultātus. Piemērā nosaukumi ir pirmajā rindā, tādēļ jāieliek ķeksītis lodziņā pretī "*Labels in first row*". Ja grupējums ir rindās, tad lodziņa nosaukums ir "*Labels in first column*".

Ja grupu nosaukumi ir norādīti datu apgabalā, bet lodziņā par apzīmējumu esamību nav ķeksīša, tad aprēķini tiks veikti skaitliskajiem lielumiem, ignorējot tekstu, bet rezultātu izdrukā apzīmējumu vietā būs nosaukumi *"Variable 1"* utt.

Ja norādītajā datu apgabalā ir tikai dati bez apzīmējumiem, bet ir atzīmēts, ka apzīmējumi ir, tad pirmās rindas (vai ailes, ja grupējums ir rindās) lielumi tiks atzīti par virsrakstiem un aprēķinos netiks izmantoti.

"*Alpha*" ir secinājumu izdarīšanas nozīmības līmenis (α). Automātiski piedāvā 0,05 – kļūdas varbūtība ir līdz 5 %. Šo lielumu ir iespējams labot un ierakstīt konkrētā pētījuma izvirzīto nozīmības līmeni. Ja datu ir nedaudz, tiem ir liela iekšgrupu izkliede un vairāk datu iegūt ir sarežģīti vai neiespējami, tad α lielumu skaitliski palielina, piemēram, līdz 0,1, bet secinājumu nozīmības līmenis samazinās. Ja vēlas lielāku drošību par to, ka secinājumos nav kļūdu, tad nozīmības līmeni paaugstina jeb samazina α skaitlisko vērtību.

Rezultātus var izvadīt jaunā darba burtnīcas lapā (*New Worksheet Ply*), blakus lodziņā norādot izvadāmās lapas nosaukumu, vai jaunā *Excel* darba burtnīcā (*New Workbook*). Šīs izvēlnes parasti netiek lietotas, to darīt ir vērts, ja ir lieli datu masīvi.

Lai izvadītu datus tajā pašā darba burtnīcas lapā, atzīmē (ieliek punktu lodziņā) izvēlni "*Output range*", novieto kursoru blakus esošajā lodziņā un vai nu iedrukā tieši šūnas adresi, vai arī ar peli uzklikšķina uz tās šūnas, kas būs izvadāmo rezultātu kreisais augšējais stūris. No šīs šūnas pa labi un uz leju ir jābūt pietiekami daudz brīvas vietas, kur izvadīt rezultātus (parasti izvēlas tādu vietu darba lapā, kur uz leju un pa labi vairs nav nekādu datu). Ja šādi dati tomēr ir, tad programma dialoga logā prasīs, vai pārrakstīt (*Output range will overwrite existing data, Press OK to overwrite data in range*) esošos datus. Ja tur atrodas iepriekš veiktie dispersijas analīzes rezultāti, kuru aprēķinos (izmantoto lielumu norādēs) ir bijušas kādas kļūdas un tagad tie nav vajadzīgi, tad spiež *OK*, pretējā gadījumā – "*Cancel*" un atrod citu datu izvades sākumpunktu.

Kad tas ir izdarīts, spiež *OK* un iegūst rezultātus, kas parādīti 6.5. attēlā.

Secinājumus par to, kura metode ir labāka, izdara pēc vidējiem lielumiem. Šajā gadījumā pozitīvi tiek vērtēta lielāka raža, tādēļ par labāko ir jāatzīst 3.metode. Sliktākā metode ir 4., kurai ir zemākā vidējā ražība, kā arī plašākā datu izkliede (*variance*)

8	Anova: Single Factor						
9							
10	SUMMARY						
11	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
12	1. metode	5	100	20	2,5		
13	2.metode	5	105	21	2,5		
14	3.metode	5	122	24,4	3,3		
15	4.metode	5	86	17,2	3,7		
16							
17							
18	ANOVA						
19	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
20	Between Groups	132,55	3	44,18333	14,72778	7,28E-05	3,238872
21	Within Groups	48	16	3			
22							
23	Total	180,55	19				
24							

Secinājumu par to, vai grupējums izskaidro datu izkliedi, izdara, salīdzinot *F* un *Fcrit*. Šajā gadījumā modelis ir statistiski nozīmīgs – mēslošanas metodes būtiski ietekmē ražas lielumu

6.5. attēls. ***Excel*** **darba lapas fragments ar dispersijas analīzes rezultātiem un komentāriem**

Iegūtās tabulas var pārkopēt *Word* dokumentā vai *Power Point* prezentācijā un analizēt tur. Analīzei var izmantot tikai iegūtos galvenos rezultātus, atsaucoties uz *Excel* aprēķiniem.

Rezultātos ir divas tabulas. Dispersijas analīzes (*ANOVA*) tabula ir gandrīz tāda pati, kā tika iegūta manuālajos aprēķinos (6.3. tabula):

- *Source of Variation* – izkliedes avots;
- *Between Groups* – starpgrupu jeb izskaidrotā datu izkliede;
- *Within Groups* – iekšgrupu jeb neizskaidrotā izkliede;
- *Total* – kopējā datu izkliede.
- *SS* – noviržu kvadrātu summas;

- *df* – brīvības pakāpes;
- *MS* – dispersijas (noviržu kvadrātu summa uz vienu brīvības pakāpi);
- *F* – empīriskā Fišera kritērija vērtība;
- *P-value* – varbūtība, ka nulles hipotēze ir pareiza, manuālajos aprēķinos šādu lielumu neiegūst;
- F_{crit} – kritiskā Fišera kritērija vērtība.

No dispersijas analīzes (*ANOVA*) tabulas pamatsecinājumus izdara, salīdzinot *F* ar F_{crit}. Secinājumu būtība tika apskatīta jau iepriekš. Šajā gadījumā $F > F_{crit}$, tādēļ nulles hipotēzi noraida un pieņem alternatīvo hipotēzi, kas, "pārtulkojot" to pētāmās hipotēzes valodā, nozīmē – grupējums atklāj būtisku iemeslu datu variācijai jeb mēslošanas laiki un devas būtiski ietekmē ražas lielumu.

P-value dod tos pašus secinājumus – varbūtība, ka pētāmis faktors neietekmē rezultātus un to izkliede ir nejauša, ir ļoti niecīga – 7,28E-05 = 0,0000728. Tādēļ ar ļoti lielu pārliecību varam apgalvot, ka mēslošanas devas un laiki ir būtiski ražas lielumam.

Vēl no šīs tabulas var iegūt ietekmes īpatsvaru – izskaidrotā noviržu kvadrātu summa ir 132,55 no 180,55 kopējās noviržu kvadrātu summas, t. i., ar doto grupējumu tiek izskaidroti 73 % (132,55/180,55*100 %) no kopējās datu izkliedes un neizskaidroto faktoru ietekme ir 27 %. (manuālajos aprēķinos – 6.4. tabula).

Kopsavilkuma (*SUMMARY*) tabulā ir parādīts novērojumu skaits (*Count*) grupā, datu summa (*Sum*), grupu vidējās vērtības (*average*) un datu izkliedes rādītājs – dispersija (*Variance*) pa grupām.

Svarīgākais no šiem rādītājiem ir vidējais. Tā traktējums ir atkarīgs no pētījuma būtības. Šajā piemērā tika pētīta raža, tādēļ labāks ir variants, kas dod augstāko ražu. Ja pētītu resursu, izmaksu vai laika patēriņu, tad labāks būtu variants ar mazāko vidējo vērtību. Šos secinājumus nedod statistiskais pētījums, bet gan tā nozare, kura izmanto statistiskās pētniecības instrumentus.

Mazāk svarīgs ir dispersijas lielums. 3. mēslošanas metode dod lielāko ražu, bet tai ir arī otrs lielākais izkliedes rādītājs, tātad fona (neizskaidroto faktoru) ietekme varētu būt lielāka nekā citām metodēm. Dažreiz (piemēram, kvalitātes pētījumos) dispersijas lielumam var būt svarīgāka nozīme – svarīgs ir ne tikai vidējais izmērs, bet arī, cik lieli ir atsevišķu izstrādājumu izmēri, kas veido šo vidējo vērtību.

Jau iepriekš tika minēts, ka var izanalizēt arī vairāku faktoru ietekmi uz rezultējošo pazīmi, tas tiks apskatīts nākamajā apakšnodaļā.

6.4. Divu faktoru dispersijas analīzes būtība un manuālie aprēķini

Veicamās aprēķinu procedūras pēc būtības ir analogas viena faktora dispersijas analīzei. Analīzē ieviešot jaunu faktoru, pieaug aprēķināmo noviržu kvadrātu summu skaits, statistiskais komplekss un formulas kļūst sarežģītākas. Papildus tam, ka ir jauna papildus faktora iedarbība, ir jāaprēķina arī pētāmo faktoru mijiedarbības efekts. Piemēram, var konstatēt, ka būtiska ietekme ir agroklimatiskajiem apstākļiem (A faktors) un kultūraugu šķirnei (B faktors) un būtiska ir arī šo faktoru mijiedarbība – kāda šķirne ir piemērotāka sausiem, bet cita mitriem laika apstākļiem. Var būt, ka pētāmie faktori atsevišķi nav statistiski nozīmīgi rezultāta veidošanā, bet šo faktoru mijiedarbība būtiski ietekmē rezultātu. Ar divu faktoru dispersijas analīzi var pētīt dažādu tehnoloģisko procesu un izejvielu piegādātāju ietekmes efektu u.tml.

Divu faktoru dispersijas analīzē ir tādi paši soļi kā viena faktora dispersijas analīzē. Lai varētu uzreiz veikt nepieciešamos aprēķinus, tālāk ir dots 6.2. piemērs.

Lai vieglāk varētu veikt manuālos aprēķinus, arī šim piemēram varianšu vērtības tiks kodētas. Pazīmes variācija ir no 26 līdz 37 ballu kopsummas, piemērota konstante samazināšanai ir 30.

Pirms darba tabulas veidošanas der apskatīties, kādas formulas var lietot noviržu kvadrātu summu aprēķināšanai. Līdzīgi kā viena faktora dispersijas gadījumā, var lietot noviržu formulas, kas ir saprotamākas, tās atspoguļo aprēķinu būtību, bet prasa vairāk kalkulācijas. Un var lietot momentu formulas, kas vizuāli šķiet sarežģītākas, bet ievērojami samazina nepieciešamās kalkulācijas apjomu.[1,2,3]

Kopējo noviržu kvadrātu summu aprēķina pēc 6.13. formulas:

$$Q = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{l=1}^{n_{ij}} \left(x_{ijl} - \bar{x} \right)^2 , \qquad (6.13.)$$

kur $\sum_{i=1}^{k}$ – summēšana pa A faktora gradācijas klasēm no 1. līdz pēdējai k klasei;

$\sum_{j=1}^{m}$ – summēšana pa B faktora klasēm no 1. līdz pēdējai m klasei;

$\sum_{l=1}^{n_{ij}}$ – summēšana statistiskā kompleksa struktūras elementa (*A* faktora *i* klases un *B* faktora *j* klases novērojumi – atkārtojumi) iekšienē no 1. līdz pēdējam n_{ij} atkārtojumam;

x_{ijl} – *A* faktora *i* gradācijas klases un *B* faktora *j* gradācijas klases l variante (novērojums).

Faktiski šī formula nozīmē, ka no katras novērotās vērtības (variantes) atņem visu novērojumu vidējo vērtību, iegūtās starpības kāpina kvadrātā un tad visus noviržu kvadrātus summē.

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 764.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 192. lpp.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga : Izglītības soļi.145. lpp.

6.2. piemērs. Vīnu ražotnes tehnologi vēlas pārbaudīt 4 gadu vīnus, vai to atšķirības ir tik lielas, lai atsevišķu gadu ražojumiem būtu pamats noteikt atšķirīgu vairumcenu. Šo vīnu vērtēšanai ir pieaicināti 5 eksperti – degustatori, kuri vērtē garšu, izskatu, aromātu u.tml.. Vērtējums ir ballēs, kuras pēc noteikta algoritma summē. Ekspertiem paraugus piedāvā jauktā secībā un testēšanu atkārto 3 dienas pēc kārtas.
Ir jānoskaidro:
1) vai ražošanas gads padara vīnu atšķirīgu, vai šīs variācijas var uzskatīt par tik lielām, lai noteiktu atšķirīgu vairumcenu (faktors A);
2) vai ir atšķirīga ekspertu vērtējumu stingrība - jutīgums (faktors B);
3) vai eksperti atšķirīgi vērtē dažādās vīnu variācijas (faktoru mijiedarbība AB).
Informācija par vīnu variāciju vērtējumiem ir dota tālāk:

	1. vīna variācija	2. vīna variācija	3. vīna variācija	4. vīna variācija
1. degustators	32 35 37	36 37 33	34 32 30	31 33 29
2. degustators	37 34 29	36 37 32	35 36 30	32 34 26
3. degustators	34 32 36	35 37 37	35 34 32	31 34 33
4. degustators	32 36 37	37 36 37	32 34 34	34 33 31
5. degustators	29 34 32	32 37 35	30 35 32	26 33 31

$$Q_A = mn\sum_{i=1}^{k}\left(\bar{x}_i - \bar{x}\right)^2 \qquad (6.14.)$$

$$Q_B = kn\sum_{j=1}^{m}\left(\bar{x}_j - \bar{x}\right)^2, \qquad (6.15.)$$

kur m – B faktora gradācijas klašu skaits;
n – novērojumu skaits vienā struktūras elementā (atkārtojumu skaits);
k – A faktora gradācijas klašu skaits;
$\bar{x}_i$ – A faktora i gradācijas klases aritmētiskais vidējais;
$\bar{x}_j$ – B faktora j gradācijas klases aritmētiskais vidējais.

Šie aprēķini rāda grupu vidējo (grupējums pēc katra faktora atsevišķi) izkliedi ap kopējo vidējo.

Faktoru mijiedarbību AB aprēķina pēc 6.16. formulas:

$$Q_{AB} = n\sum_{i=1}^{k}\sum_{j=1}^{m}\left(\overline{x_{ij}} - \overline{x_i} - \overline{x_j} - \overline{x}\right)^2 , \tag{6.16.}$$

kur $\overline{x_{ij}}$ – statistiskā kompleksa elementa *i A* faktora un *j B* faktora gradācijas klases atkārtojumu vidējā vērtība.

Fona jeb neizskaidroto noviržu kvadrātu summu aprēķina pēc 6.17. formulas:

$$Q_Z = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n}\left(x_{ijl} - \overline{x_{ij}}\right)^2 \tag{6.17.}$$

Fona noviržu kvadrātu summu rāda datu izkliedi ap statistiskā kompleksa elementu vidējām vērtībām.

Tālāk tiks apskatītas manuālajiem aprēķiniem piemērotākās momentu formulas. [1,2]

$$C = \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n} x_{ijl}\right)^2}{N} \tag{6.18.}$$

Ar *C* apzīmē korekcijas locekli, formula nozīmē, ka visu novērojumu summa ir kāpināta kvadrātā un dalīta ar kopējo novērojumu skaitu.

Kopējo noviržu kvadrātu summu aprēķina pēc 6.19.formulas:

$$Q = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n} x_{ijl}^2 - C , \tag{6.19.}$$

kas nozīmē no visu datu kvadrātu summas (iepriekš bija skaidrojums, ar ko atšķiras summas kvadrāts no kvadrātu summas) atņemt korekcijas locekli.

Faktoru noviržu kvadrātu summas aprēķina pēc 6.20. un 6.21. formulas:

$$Q_A = \sum_{i=1}^{k}\frac{\left(\sum_{j=1}^{m}\sum_{l=1}^{n} x_{ijl}\right)^2}{mn} - C , \tag{6.20.}$$

---

[1] Raščevska, M., Kristapsone, S. (2000). Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem. Rīga: Izglītības soļi. 144. lpp.
[2] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 134. lpp.

$$Q_B = \sum_{j=1}^{m} \frac{\left(\sum_{i=1}^{k} \sum_{l=1}^{n} x_{ijl} \right)^2}{kn} - C \qquad (6.21.)$$

Formulas nozīmē, ka datus summē pa faktoru gradācijas klasēm, iegūtās summas kāpina kvadrātā, dala ar atbilstošā faktora gradācijas klasē esošo novērojumu skaitu, summē pa visām pētāmā faktora klasēm un tad atņem korekcijas locekli.

Ir jāatrod faktoru mijiedarbības (Q_{AB}) noviržu kvadrātu summa, bet, tā kā vieglāk ir aprēķināt kopējo izskaidroto jeb faktoriālo noviržu kvadrātu summu, tad vispirms aprēķina to (Q_f) pēc 6.22. formulas:

$$Q_f = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(\sum_{l=1}^{n} x_{ijl} \right)^2}{n} - C, \qquad (6.22.)$$

tas nozīmē, ka statistiskā kompleksa elementos summē novērotās vērtības, iegūtās summas kāpina kvadrātā, dala ar novērojumu (atkārtojumu) skaitu, iegūtos rezultātus summē gan pa *A*, gan pa *B* faktora gradācijas klasēm un tad atņem korekcijas locekli.

Tā kā izskaidrotā noviržu kvadrātu summa veidojas no *A* un *B* faktora un faktoru mijiedarbības *AB* noviržu kvadrātu summām, tad faktoru mijiedarbības noviržu kvadrātu summu (Q_{AB}) aprēķina pēc 6.23.formulas:

$$Q_{AB} = Q_f - Q_A - Q_B, \qquad (6.23.)$$

Fona noviržu kvadrātu summu (Q_Z) aprēķina pēc atlikuma principa ar 6.24. formulu:

$$Q_Z = Q - Q_f \qquad (6.24.)$$

6.6. tabula

Statistiskā kompleksa darba tabula 6.2. piemēram, veidota aprēķiniem pēc momentu formulām, variantes kodētas x_{ijl} – 30, blakus ailēs $(x_{ijl} - 30)^2$

Faktors B	Faktors A								nj	1.	2.	3.
	A1		A2		A3		A4					
B1	2 5 7	4 25 49	6 7 3	36 49 9	4 2 0	16 4 0	1 3 -1	1 9 1				
4., 5.	14	78	16	94	6	20	3	11	12	39	203	33,25
B2	7 4 -1	49 16 1	6 7 2	36 49 4	5 6 0	25 36 0	2 4 -4	4 16 16				
4., 5.	10	66	15	89	11	61	2	36	12	38	252	33,17
B3	4 2 6	16 4 36	5 7 7	25 49 49	5 4 2	25 16 4	1 4 3	1 16 9				
4., 5.	12	56	19	123	11	45	8	26	12	50	250	34,17
B4	2 6 7	4 36 49	7 6 7	49 36 49	2 4 4	4 16 16	4 3 1	16 9 1				
4., 5.	15	89	20	134	10	36	8	26	12	53	285	34,42
B5	-1 4 2	1 16 4	2 7 5	4 49 25	0 5 2	0 25 4	-4 3 1	16 9 1				
4., 5.	5	21	14	78	7	29	0	26	12	26	154	32,17
ni	15		15		15		15		60			
6., 7.	56	310	84	518	45	191	21	125		206	1144	
8..	33,73		35,60		33,00		31,40					

Tabulas ailēs un rindās norādīto numuru atšifrējumi:

1. $\sum_{i=1}^{k}\sum_{l=1}^{n}\left(x_{ijl}-30\right)$
2. $\sum_{i=1}^{k}\sum_{l=1}^{n}\left(x_{ijl}-30\right)^2$
3. $\overline{x}_j$

4. $\sum_{l=1}^{n}\left(x_{ijl}-30\right)$

5. $\sum_{l=1}^{n}\left(x_{ijl}-30\right)^2$

6. $\sum_{j=1}^{m}\sum_{l=1}^{n}\left(x_{ijl}-30\right)$

7. $\sum_{j=1}^{m}\sum_{l=1}^{n}\left(x_{ijl}-30\right)^2$

8. $\overline{x_i}$

Nepieciešamās summas no darba tabulas ievieto noviržu kvadrātu summu aprēķināšanas momentu formulās (6.18. –6.24. formula).

$C = 206^2/60 = 707{,}27$

$Q = 1144 - 707{,}27 = 436{,}73$

$Q_A = 1/5*3(56^2 + 84^2 + 45^2 + 21^2) - 707{,}27 = 136{,}60$

$Q_B = 1/4*3(39^2 + 38^2 + 50^2 + 53^2 + 26^2) - 707{,}27 = 38{,}56$

$Q_f = 1/3(14^2 + 16^2 + 6^2 + 3^2 + 10^2 + 15^2 + 11^2 + 2^2 + 12^2 + 19^2 + 11^2 + 8^2 + 15^2 + 20^2 + 10^2 + 8^2 + 5^2 + 14^2 + 7^2 + 0^2) - 707{,}27 = 191{,}40$

$Q_{AB} = 191{,}40 - 136{,}60 - 38{,}56 = 16{,}24$

$Q_Z = 436{,}73 - 191{,}40 = 245{,}33$

Tālākos aprēķinus veic dispersijas analīzes tabulā.[1,2] Vispirms 6.7. tabulā ir parādīti apzīmējumi un aprēķinu formulas, bet 6.8. tabulā – aprēķini 6.2. piemēram.

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 773.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 196. lpp.

6.7. tabula

Divfaktoru dispersijas analīzes tabula ar apzīmējumiem un formulām

Izkliede	*Noviržu kvadrātu summas*	*Brīvības pakāpes*	*Dispersija*	F_{emp}	F_{krit}
Faktors *A*	Q_A	$\nu_A=k-1$	$S_A^2= Q_A/ \nu_A$	S_A^2/ S_Z^2	α, ν_A, ν_Z
Faktors *B*	Q_B	$\nu_B=m-1$	$S_B^2= Q_B/ \nu_B$	S_B^2/ S_Z^2	α, ν_B, ν_Z
Mijiedarbība *AB*	Q_{AB}	$\nu_{AB}= \nu_A * \nu_B$	$s_{AB}^2= Q_{AB}/ \nu_{AB}$	s_{AB}^2 /S_Z^2	α, ν_{AB}, ν_Z
Atlikums	Q_Z	$\nu_Z= \nu- \nu_A- \nu_B- \nu_{AB}$	$S_Z^2= Q_Z/ \nu_Z$		
Kopējā	Q	$\nu=N-1$			

6.8. tabula

Dispersijas analīzes tabula 6.2.piemēram

Izkliede	*Noviržu kvadrātu summas*	*Brīvības pakāpes*	*Dispersija*	*Femp*	*Fkrit*
Faktors A	136,60	3	45,53	7,43	2,84
Faktors B	38,56	4	9,64	1,57	2,61
Mijiedarbība AB	16,24	12	1,35	0,22	2,00
Atlikums	245,33	40	6,13		
Kopējā	436,73	59			

Brīvības pakāpes faktoriem aprēķina, no faktora gradācijas klašu skaita atņemot 1.

$\nu_A = k - 1 = 4 - 1 = 3$

$\nu_B = m - 1 = 5 - 1 = 4$

Kopējai izkliedei brīvības pakāpes arī aprēķina līdzīgi kā citos gadījumos – no kopējā novērojumu skaita atņem 1.

$\nu = n - 1 = 60 - 1 = 59$

Faktoru mijiedarbībai AB brīvības pakāpes aprēķina, sareizinot faktoru brīvības pakāpes.

$\nu_{AB} = \nu_A * \nu_B = 3 * 4 = 12$

Fona izkliedei brīvības pakāpes aprēķina pēc atlikuma principa – no kopējās izkliedes brīvības pakāpēm atņem izskaidrotās datu izkliedes brīvības pakāpes.

$\nu_Z = \nu - \nu_A - \nu_B - \nu_{AB} = 59 - 3 - 4 - 12 = 40$

Dispersijas aprēķina parastajā kārtībā, dalot noviržu kvadrātu summu ar brīvības pakāpēm.

Empīrisko Fišera kritēriju aprēķina, dalot atbilstošo izskaidrotās dispersijas elementu ar atlikuma (neizskaidroto) dispersiju.

Kritisko Fišera kritērija vērtību atrod tabulās atbilstoši izvēlētajam nozīmības līmenim α un empīriskā Fišera kritērija aprēķināšanai izmantotajām dispersiju brīvības pakāpēm ν.

ν_1 ir brīvības pakāpes skaitītājā esošajai dispersijai, bet ν_2 ir brīvības pakāpes saucējā esošajai dispersijai.

Tā 6.2. piemērā kritiskās Fišera kritērija vērtības noteikšanai faktoram A ir $\nu_1 = 3$, bet $\nu_2 = 40$, faktoram B - $\nu_1 = 4$, bet $\nu_2 = 40$ un faktoru mijiedarbībai – $\nu_1 = 12$, bet $\nu_2 = 40$.

Secinājumus par rezultātiem izdara parastajā kārtībā – ja $F_{emp} > F_{krit}$, tad pētāmā faktora vai faktoru mijiedarbības ietekme ir būtiska; ja $F_{emp} < F_{krit}$, tad pētāmā faktora vai faktoru mijiedarbības ietekme nav būtiska un nulles hipotēzi noraidīt nevar.

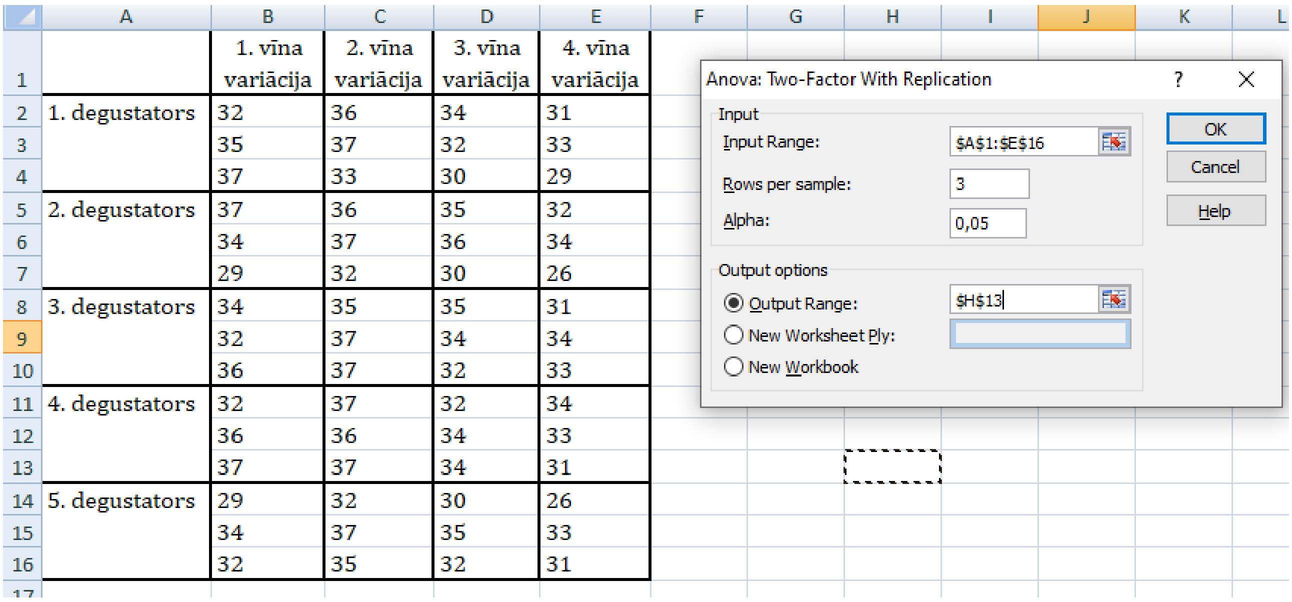
Secinājumi 6.2.piemēram ir šādi:

- *ražošanas gads ir būtiski ietekmējis dzēriena kvalitāti (Faktors A bija vīna ražošanas gads). Tādēļ būtu pamats domāt par atsevišķu gadu vīnu kā atšķirīgu produktu virzīšanu tirgū. Pēc ekspertu vērtējuma labākā ir A2 variācija ar vidējo vērtējumu 35,6 balles;*
- *eksperta (B) faktors nav statistiski nozīmīgs, ekspertu vērtēšanas skala visiem ekspertiem ir samērā līdzīga. Ekspertu komanda ir saderīga;*
- *faktoru mijiedarbība ir ļoti maz ietekmējusi datu izkliedi, tas nozīmē, ka ekspertiem nav atšķirīga vērtēšanas stingrība.*

Šāda tipa secinājumi ir tieši. kamdēļ tiek veikti statistiskie pētījumi, kaut arī tie vairāk nav tieši statistiskie secinājumi. Iegūtie statistiskie secinājumi par nulles un alternatīvajām hipotēzēm, pētījumu beidzot, ir "jāpārtulko" sākotnējās hipotēzes terminos, noraidot to vai apstiprinot.

6.5. Divfaktoru dispersijas analīze ar programmu *Excel*

Tālāk ir dots šī paša uzdevuma risinājums ar programmu *Excel*, dialoga logs ir redzams 6.6. attēlā.



	1. vīna variācija	2. vīna variācija	3. vīna variācija	4. vīna variācija
1. degustators	32	36	34	31
	35	37	32	33
	37	33	30	29
2. degustators	37	36	35	32
	34	37	36	34
	29	32	30	26
3. degustators	34	35	35	31
	32	37	34	34
	36	37	32	33
4. degustators	32	37	32	34
	36	36	34	33
	37	37	34	31
5. degustators	29	32	30	26
	34	37	35	33
	32	35	32	31

6.6. attēls. ***Excel*** **darba lapas fragments ar ievadītajiem datiem un aizpildītu dialoga logu divfaktoru dispersijas analīzei**

Datu izvietojums darba lapā ir līdzīgs tam, kā to veidoja manuāli statistiskajā kompleksā, atšķirība tā, ka nav jāatstāj rindas kopsummām.

Savukārt dialoga logā atšķirībā no vienfaktora dispersijas analīzes nav jautājuma par grupējumu (rindās vai ailēs), jo grupējums ir gan rindās, gan ailēs.

Atkārtojumi ir jāizvieto uz leju un ir jānorāda, cik rindās ir vienas, piemēram, A1B1, grupas novērojumi. Nospiež *OK* un iegūst rezultātus, kas doti 6.7. attēlā.

Rezultātus analizē, sākot ar *ANOVA* tabulu, kur izdara secinājumus par faktoriem, kas ir vai nav būtiski datu izkliedes izskaidrošanā. Secinājumi jau tika apskatīti pēc manuālajiem aprēķiniem (pēc 6.8. tabulas).

Manuālo aprēķinu tabulā nav "*P-value*", tāpēc jāpievēršas šo lielumu analīzei. Var secināt, ka vērtēšanas rezultātu statistiski nozīmīgi ietekmē vienīgi vīna ražošanas gads. Tomēr diezgan liela ietekme ir bijusi arī degustatoriem. Varbūtība, ka nulles hipotēze ir pareiza, ir 0,2, bet tas nozīmē, ka varbūtība, ka degustatori ir ar atšķirīgu stingrību vērtējumos, ir 0,8 (1–0,2). Tas tomēr ir diezgan augsts rādītājs.

Taču tas nav svarīgākais šajā piemērā (svarīgums izriet no uzdevuma loģikas), galvenais, degustatori vīnu variācijas ir sarindojuši praktiski vienādi (*interaction P-value* ir gandrīz viens).

Anova: Two-Factor With Replication

SUMMARY	1. liķiera variācija	2. liķiera variācija	3. liķiera variācija	4. liķiera variācija	Total
1. degustators					
Count	3	3	3	3	12
Sum	104	106	96	93	399
Average	34,66667	35,33333	32	31	33,25
Variance	6,333333	4,333333	4	4	6,931818
2. degustators					
Count	3	3	3	3	12
Sum	100	105	101	92	398
Average	33,33333	35	33,66667	30,66667	33,16667
Variance	16,33333	7	10,33333	17,33333	11,9697
3. degustators					
Count	3	3	3	3	12
Sum	102	109	101	98	410
Average	34	36,33333	33,66667	32,66667	34,16667
Variance	4	1,333333	2,333333	2,333333	3,787879
4. degustators					
Count	3	3	3	3	12
Sum	105	110	100	98	413
Average	35	36,66667	33,33333	32,66667	34,41667
Variance	7	0,333333	1,333333	2,333333	4,628788
5. degustators					
Count	3	3	3	3	12
Sum	95	104	97	90	386
Average	31,66667	34,66667	32,33333	30	32,16667
Variance	6,333333	6,333333	6,333333	13	8,878788
Total					
Count	15	15	15	15	
Sum	506	534	495	471	
Average	33,73333	35,6	33	31,4	
Variance	7,209524	3,4	4	6,828571	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	38,56667	4	9,641667	1,572011	0,20045	2,605975
Columns	136,6	3	45,53333	7,423913	0,000458	2,838745
Interaction	16,23333	12	1,352778	0,220562	0,996375	2,003459
Within	245,3333	40	6,133333			
Total	436,7333	59				

6.7. attēls. **Divfaktoru dispersijas analīzes rezultāti, iegūti ar *Excel***

Kopsavilkuma tabulas analizē pēc būtības – augstākais vērtējums ir 2. vīna variācijai. Var paanalizēt arī degustatoru vērtējumus. Vidēji augstākos vērtējumus ir piešķīris 3. un

4. degustators, plašākā vērtējumu amplitūda ir bijusi 2.degustatoram. Šos secinājumus nevar izdarīt ar nozīmības līmeni 0,05, bet varbūtība, ka tie ir reāli, ir pietiekami augsta.

Dispersijas analīze noskaidro to, vai pētāmais faktors būtiski ietekmē rezultējošās pazīmes izmaiņas. Var aprēķināt faktoru ietekmes īpatsvaru, bet vairāk interesē, kāda ir sakarības forma, kā, ietekmējot faktoriālo pazīmi, var panākt vēlamo rezultātu. Ja faktoriālā pazīme variē kvantitatīvi (rezultējošajai pazīmei kvantitatīvi jāvariē arī dispersijas analīzē), tad cenšas noskaidrot arī sakarības formu, un to dara regresijas analīzē, par ko tiks runāts nākamajā nodaļā.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
Atlikuma (neizskaidrotā) noviržu kvadrātu summa	Residual (within groups) sum of squares	Внутригрупповая сумма квадратов	Iekšgrupu datu izkliede – datu izkliede ap grupu vidējiem
Brīvības pakāpes	Degree of freedom	Степени свободы	Novērojumu skaits mīnus saistošo nosacījumu skaits
Daudzfaktoru dispersijas analīze	Multiple analysis of variance	Многофакторный дисперсионный анализ	Datu variāciju izskaidro ar grupējumu pēc diviem vai vairāk faktoriem
Dispersijas analīze	Analysis of variance	Дисперсионный анализ	Analīzes metode, kad salīdzina ar grupējumu izskaidroto datu izkliedi ar neizskaidroto datu izkliedi
Fatoriālā (izskaidrotā) dispersija	Between group variance	Факториальная дисперсия	Ar grupējumu izskaidrotā datu izkliede
Fišera kritērijs	F - test	F- критерий	Fišera kritērijs, izmanto dispersiju salīdzināšanai. Ja empīriskā vērtība ir lielāka par kritisko, tad nulles hipotēzi noraida, datu izkliede atšķiras būtiski, dispersijas analīzē pētāmais faktors būtiski ietekmē rezultātus
Fona (neizskaidrotā) dispersija	Within groups variance	Остаточная дисперсия	Datu izkliede ap izdalīto grupu vidējiem
Kopējā dispersija	Total variance	Общая дисперсия	Datu izkliede ap kopējo (visu novērojumu) vidējo
Kopējā noviržu kvadrātu summa	Total sum of squares	Общая сумма квадратов	Noviržu kvadrātu summa ap kopējo vidējo
Starpgrupu (izskaidrotā) noviržu kvadrātu summa	Between groups sum of squares	Межгрупповая сумма квадратов	Ar grupējumu izskaidrotā noviržu kvadrātu summa – grupu vidējo izkliede ap kopējo vidējo
Vienfaktora dispersijas analīze	ANOVA (analysis of variance): single factor	Однофакторный дисперсионный анализ	Dati ir grupēti pēc vienas pazīmes – faktora

7. KORELĀCIJAS UN REGRESIJAS ANALĪZE

Pēc nodaļas apgūšanas studentiem:

- *jāizprot korelācijas un regresijas analīzes būtība;*
- *jāprot konstruēt korelācijas diagramma;*
- *jāprot aprēķināt Pīrsona korelācijas koeficientu un lineārās regresijas vienādojuma konstantes, konstruēt aprēķināto regresijas taisni;*
- *jāsaprot korelācijas koeficienta un regresijas vienādojuma koeficientu statistiskā nozīmīguma izvērtēšanas būtība;*
- *jāzina neparametriskās sakarību pētīšanas metodes;*
- *jāprot konstruēt dažādu formu regresijas līknes, iegūt regresijas vienādojumu un determinācijas koeficientu ar Microsoft Excel palīdzību;*
- *jāprot aprēķināt korelācijas koeficientus un daudzfaktoru lineārās regresijas vienādojumu ar Microsoft Excel datu analīzes rīku, interpretēt iegūtos rezultātus.*

7.1. Korelācijas būtība un Pīrsona korelācijas koeficients

Iepriekšējā nodaļā tika apskatīts, kādā veidā var novērtēt faktora ietekmi rezultējošās pazīmes variācijā, kā pārbaudīt, vai aprēķinātā ietekme ir statistiski nozīmīga. Dispersijas analīzi var lietot, ja grupēts ir gan pēc kvantitatīvas, gan atribūtīvas pazīmes (lauksaimniecības kultūraugu šķirnes, dažādas tehnoloģijas, tautības u. tml.). Ja abas pazīmes – gan faktoriālā, gan rezultatīvā – mainās kvantitatīvi, tad var izmantot korelācijas un regresijas analīzes metodes.

Daudzas sociāli ekonomiskās parādības ir savstarpēji saistītas. Piemēram, var atrast sakarību starp subsīdijām lauksaimniekiem un apstrādātajām zemes platībām, algām un darbaspēka migrāciju, izmaksām un saražotās produkcijas apjomu utt.

Faktoriālo (neatkarīgo) pazīmi apzīmē ar X un atbilstošās pazīmes vērtības (variantes) – ar x_1, x_2, x_3, … x_n, rezultējošo (atkarīgo) pazīmi apzīmē ar Y un atbilstošās pazīmes vērtības – ar y_1, y_2, y_3, … y_n.[1,2]

Sakarības var būt vienkāršas, kad viena faktoriālā pazīme ietekmē vienu rezultējošo pazīmi, un kombinētas, kad ir vairākas faktoriālās pazīmes, kas ietekmē vienu rezultējošo pazīmi. Šajā gadījumā faktoriālās pazīmes apzīmē ar

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 157. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 221. lpp.

X_1, X_2, … X_m, kur m ir kopējais faktoriālo pazīmju skaits. Atbilstošās faktoriālās pazīmes vērtības apzīmē ar diviem indeksa skaitļiem – pirmais ir pazīmes numurs, otrais – variantes numurs pēc kārtas, piemēram, x_{11} (1. faktoriālās pazīmes 1. novērojums), x_{12}, … x_{1n}, x_{21} (2. faktoriālās pazīmes 1. novērojums, iegūts kopā ar y_1 un x_{11} un citu pētījumā iekļauto faktoriālo pazīmju atbilstošajiem novērojumiem), x_{22}, …x_{2n} līdz pat x_{m1}, x_{m2}, x_{m3}, … x_{mn}.

Sakarība starp pazīmēm var būt funkcionāla, kad katrai x vērtībai atbilst tikai viena y vērtība[1], piemēram, uzņēmumā ir gabaldarba samaksa – 0,5 EUR par izstrādājuma izgatavošanu, darba samaksa ir funkcionāli atkarīga no tā, cik izstrādājumus strādnieks izgatavos.

Sakarība starp pazīmēm var būt stohastiska (korelatīva), kad katrai faktoriālās pazīmes vērtībai atbilst nevis konkrēta rezultējošās pazīmes vērtība, bet gan šīs pazīmes vidējā vērtība.[2] Dispersijas analīzē jau tika skatīts jautājums, ka rezultējošās pazīmes vērtību ietekmē faktoriālās pazīmes vērtības un arī daudzi citi neidentificēti faktori jeb savādāk to sauc par fona ietekmi. Korelācijas un regresijas analīzē bez identificētā faktora uz rezultējošo pazīmi iedarbojas vēl daudzi citi, bieži vien neregulāras iedarbības nevēlami (piemēram, kļūdas ražošanas tehnoloģijās), neitrāli vai pozitīvi apkārtējās vides faktori. Šo fona faktoru ietekmē rezultējošās pazīmes vērtība var būt gan lielāka, gan mazāka par vidējo sagaidāmo pazīmes vērtību. Piemēram, daudzos tehnoloģiskajos procesos sakarība starp izlietoto materiālu apjomu un saražotās produkcijas apjomu būs korelatīva – cukura saturs cukurbietēs noteiks cukura iznākumu, graudu rupjums ietekmēs miltu iznākumu, tāpat iznākumu ietekmēs tehnoloģiskā procesa svārstības, kļūdas, kas rada brāķi u.tml. Materiālu patēriņa normu nosaka kā vidējo patēriņu normālam (standarta) ražošanas procesam. Ļoti daudzās situācijās sakarības var novērtēt tikai pētījuma rezultātā. Piemēram, var izvirzīt hipotēzi par to, ka vidējais algas lielums valstī ietekmē migrācijas procesus (cik cilvēku izbrauc vai iebrauc valstī). Bet nekur nav noteikta norma, cik cilvēkiem ir jāatstāj valsti, ja vidējā alga ir kāda noteikta lieluma. Iepriekš izvirzīto hipotēzi var vai nu apstiprināt, vai noraidīt, iegūstot skaitļus novērojumu rezultātā un pēc tam matemātiski apstrādājot.

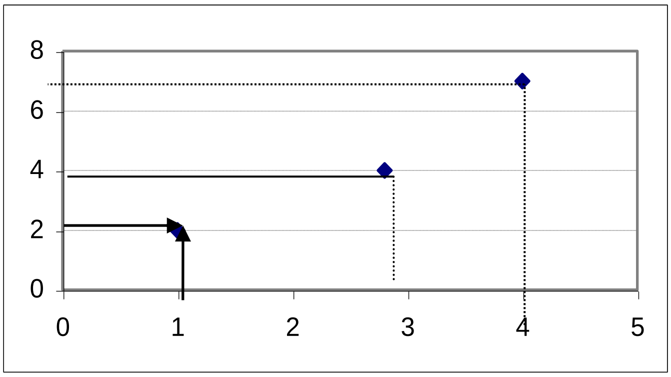
Statistikas pētniecības objekts ir variējošas pazīmes, tāpēc funkcionālās sakarības tā nepēta. Kaut arī korelatīvā sakarības forma kā dabas, tā arī sociāli ekonomiskajos pētījumos ir sastopama biežāk nekā funkcionālā sakarības forma.

Parasti sakarību pētīšana sākas ar korelācijas diagrammas izveidošanu – Dekarta koordinātu sistēmā uz abscisu (X) ass atliek faktoriālās pazīmes vērtības, bet uz ordinātu (Y) ass atliek atbilstošās rezultējošās pazīmes vērtības.

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 157. lpp.
[2] Turpat.

Katrs punkts grafikā tiek raksturots ar diviem lielumiem (*x* un *y* vērtības). Kā punktus atliek, parādīts 7.1. attēlā.



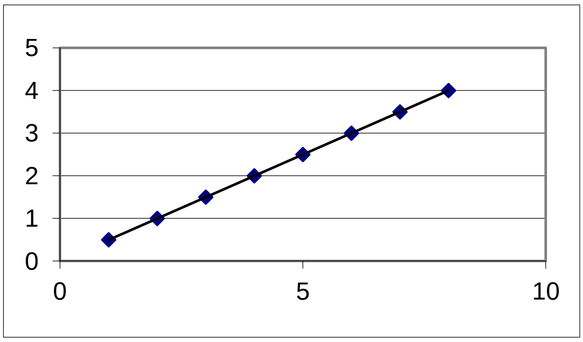
7.1. attēls. **Punktu diagrammmas veidošanas piemērs**

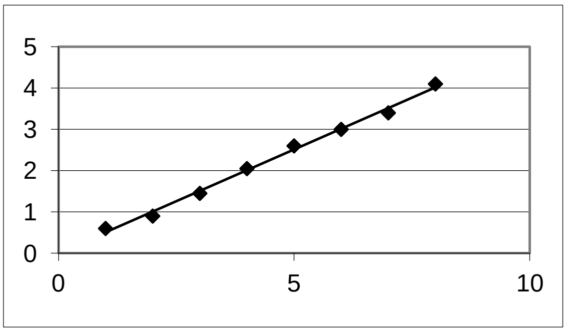
Grafikā ir atlikti ir trīs punkti $x_1 = 1$; $y_1 = 2$; $x_2 = 2{,}8$; $y_2 = 4$; $x_3 = 4$; $y_3 = 7$.

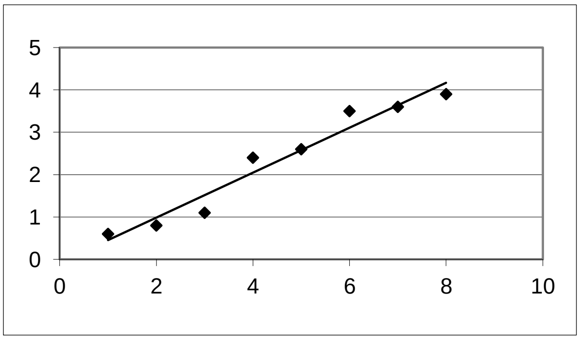
Lai atliktu punktu, uz koordinātu asīm atrod atbilstošos lielumus (obligāti ir jāievēro mērogs – nedrīkst atlikt katru nākamo novērojumu par vienādu attālumu tālāk), tad no šiem lielumiem velk taisnes (parasti domās, uz papīra velk tikai, lai iemācītos atlikt punktus), paralēlas koordinātu asīm. Tur, kur šīs taisnes krustojas, ir punkts, kas raksturo, ka tādai *x* vērtībai ir atbilstošā y vērtība. 7.1. attēlā ar bultiņām ir parādīts, kā tiek atrasts x_1; y_1 punkts, pārējiem punktiem ir norādītas taisnes.

Atkarībā no punktu izvietojuma korelācijas diagrammā, vizuāli novērtē, vai pastāv sakarība, cik tā ir cieša un kāda ir tās forma.[1,2,3] 7.2. attēlā ir parādītas, kādas var būt sakarības formas (iespējamie korelācijas diagrammas paraugi).

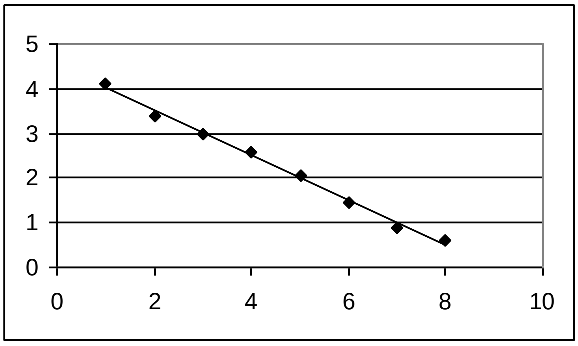
[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 163. lpp.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 222. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 408.

<table>
<tr>
<td>


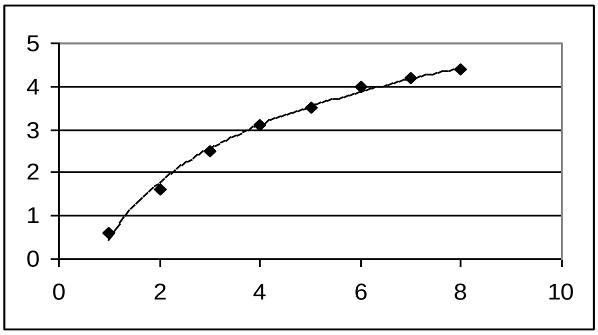
a) Funkcionāla sakarība, visi punkti atrodas uz sakarību aprakstošās līnijas (attēlā tā ir taisne, bet vispār var būt jebkura līkne)
</td>
<td>


b) Lineāra pozitīva korelācija. Pieaugot x vērtībām, pieaug arī y vērtības. Punkti nav izvietoti tieši uz līnijas, bet grupējas ap to. Jo tuvāk empīriskie novērojumu punkti atrodas teorētiskajai sakarību aprakstošajai līknei, jo ciešāka ir korelācija
</td>
</tr>
<tr>
<td>


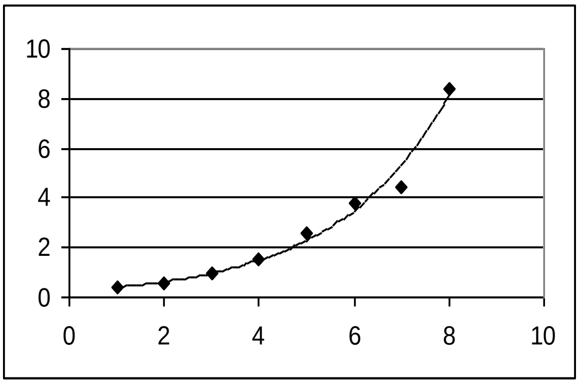
c) Arī lineāra pozitīva korelācija, taču sakarība ir vājāka nekā b) attēlā (punkti novietoti tālāk no teorētiskās līnijas)
</td>
<td></td>
</tr>
</table>



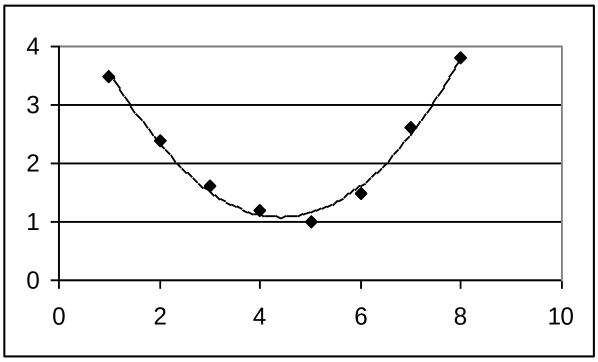
d) Lineāra negatīva korelācija, pieaugot *x* vērtībām, *y* vērtības samazinās



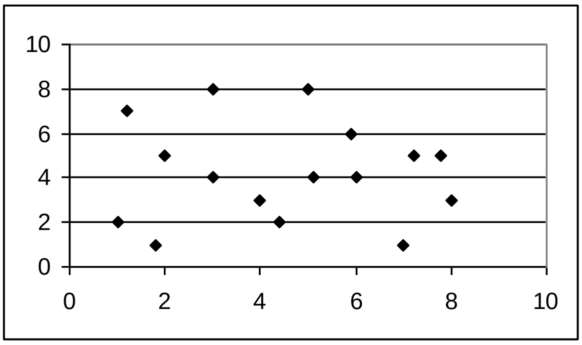
e) Nelineāra korelācija, punkti grupējas ap līkni, piemērā eksponenciālā līkne



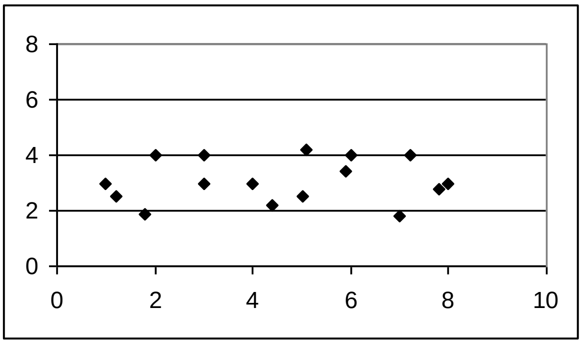
f) Nelineāra korelācija (hiperboliskā sakarība)



g) Nelineārā korelācija (paraboliskā sakarība)



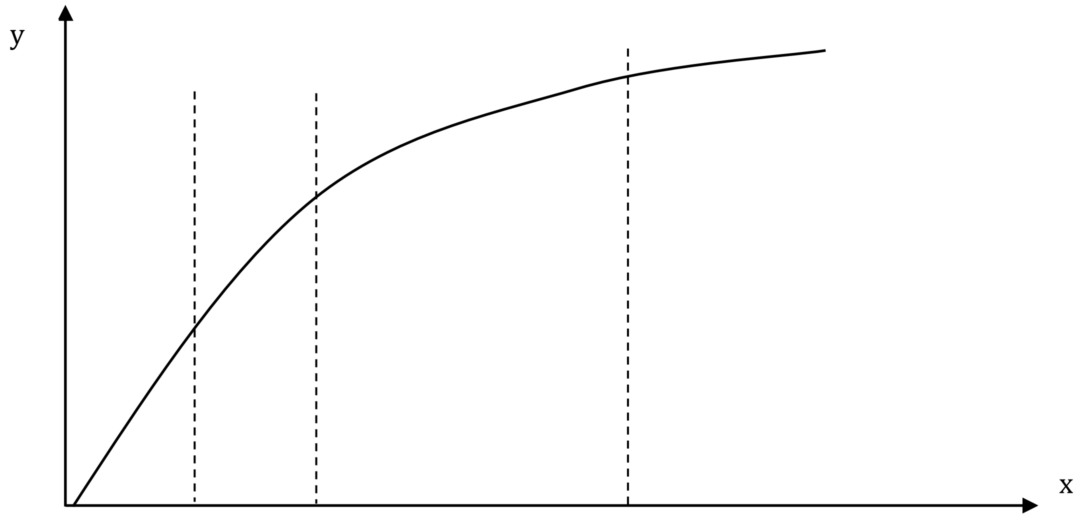
h) Nekorelēta punktu kopa, jebkurai *x* vērtībai var atbilst jebkura *y* vērtība



i) Nekorelēta punktu kopa, no h) attēla atšķiras ar to, ka y vērtības variē šaurākā diapazonā, bet, neskatoties uz to, vizuāli nevar redzēt, vai kādas *x* izmaiņas ietekmētu *y* izmaiņas

7.2. attēls. **Korelācijas diagrammu veidi**

Vairums parādību sociāli ekonomiskajos un dabas pētījumos mainās nelineāri – pieauguma vai samazinājuma tempi laika gaitā, vai arī būtiski pieaugot citai faktoriālās pazīmes vērtībai, mainās, parādās jauni faktori, kas sākotnēji nebija statistiski nozīmīgi u.tml. Taču, neskatoties uz iepriekš minēto, praksē lineārā sakarības forma tiek lietota daudz biežāk nekā nelineārā. Pamatojumu tam var apskatīt 7.3. attēlā.



7.3. attēls. **Lineāro sakarību lietošana nelineāras sakarības vietā**

7.3. attēlā ir parādīta nelineāra sakarība. Ja līkni sadala īsākos posmos (attēlā sadalīts ar raustītām līnijām), tad katru atsevišķo posmu var aprakstīt ar lineāro (taisnes) funkciju. Nelinearitāte bieži vien izpaužas ilgākā laika periodā. Piemēram, attīstās jaunas sadzīves tehnoloģijas (mobilie telefoni, DVD rakstītāji u.tml.), sākuma posmā pārdošanas apjomu pieaugums ir ļoti straujš, bet tad kādā brīdī tirgus ir piesātināts un pieaugums strauji sarūk. Gaidāmo pieauguma tempu samazinājumu var kvalitatīvi (nepamatojot ar konkrētiem skaitļiem) paredzēt, bet nav statistisko datu, kas ļautu šo funkciju aprēķināt matemātiski.

Populārākais sakarības ciešuma mērs, ko lieto, ja abi mainīgie ir mērīti intervālu vai proporcionālajā skalā (kvantitatīvi) un sakarības forma ir lineāra, ir **Pīrsona korelācijas koeficients**. Parasti praksē to sauc vienkārši par korelācijas koeficientu.[1,2,3]

Daudz retāk sakarību pētīšanai lieto Spīrmena vai Kendela rangu korelācijas koeficientus. **Rangu korelācijas koeficientus** lieto, ja viens vai abi

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 417.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 218. lpp.
[3] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 163. lpp.

mainīgie ir mērīti rangu skalā (variantes var saranžēt – piešķirt vietas vērtības pieauguma secībā, bet attālumi – vērtību starpības starp tām nav vienādas).[1,2,3]

Ja mainīgie ir mērīti nominālajā skalā (atributīvi), tad sakarības ciešuma raksturošanai lieto kontingences koeficientus.[4]

Pīrsona korelācijas koeficientu apzīmē ar "*r*". Līdzīgi kā citām statistikā lietotajām formulām, arī korelācijas koeficienta aprēķināšanai ir gan noviržu, gan momentu formulas. Turklāt pēdējās var būt dažādās izveduma formās.

Pīrsona korelācijas koeficienta aprēķināšanas noviržu formula ir:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \qquad (7.1.)$$

Tālāk dota viena no momentu formulām:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}} \qquad (7.2.)$$ [5]

Formulās ir jāsaprot, ka reizinājumu summa (Σx*y) nav tas pats, kas summu reizinājums (Σx*Σy) un kvadrātu summa (Σx^2) ir atšķirīga no summas kvadrāta – $(\Sigma x)^2$.

Formulu lietošana tiks apskatīta 7.1. piemēram.

---

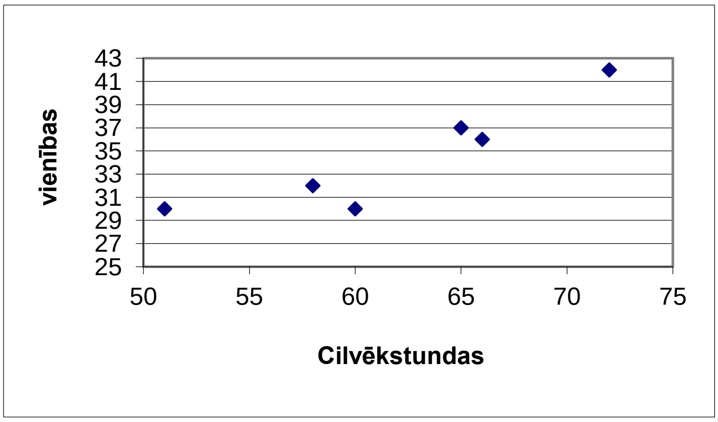
[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 201. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 846.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 408. lpp.
[4] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 205. lpp.
[5] Turpat, 163. lpp.

7.1. piemērs. Ražošanas uzņēmumā tiek pārbaudīts sakarības ciešums starp saražotās produkcijas apjomu un patērēto laiku. Tālāk ir doti novērojumi par 6 dienām.

Nostrādātās cilvēkstundas (x)	Saražotā produkcija vienībās (y)
65	37
60	30
66	36
72	42
51	30
58	32

Jāaprēķina korelācijas koeficients.

Vispirms korelācijas diagrammā novērtē, vai sakarība ir lineāra.



7.4. attēls. **Korelācijas diagramma 7.1. piemēram**

7.1. piemērā ir maz skaitļu, bet, vizuāli vērtējot, var pieņemt, ka sakarība ir tuva lineārajai. Pēc šāda secinājuma var aprēķināt Pīrsona korelācijas koeficientu.

Tālāk parādīti aprēķini pēc abām (7.1. un 7.2.) formulām. Manuālos aprēķinus formulām, kurās ir summēšanas zīmē, veic darba tabulās. 7.1. tabulā ir parādīta darba tabula korelācijas koeficienta aprēķināšanai pēc noviržu formulas.

7.1. tabula

Darba tabula korelācijas koeficienta aprēķināšanai pēc noviržu formulas 7.1. piemēram

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1.	65	37	3	2,5	9	6,25	7,5
2.	60	30	-2	-4,5	4	20,25	9
3.	66	36	4	1,5	16	2,25	6
4.	72	42	10	7,5	100	56,25	75
5.	51	30	-11	-4,5	121	20,25	49,5
6.	58	32	-4	-2,5	16	6,25	10
Σ	372	207	0	0	266	111,5	157

Vispirms aprēķina vidējo nostrādāto cilvēkstundu un saražoto vienību skaitu: $\Sigma x/n = 372/6 = 62$ un $\Sigma y/n = 207/6 = 34{,}5$.

$$r = \frac{157}{\sqrt{266 * 111{,}5}} = 0{,}91$$

Piemērs ir neliels, skaitļi ir ļoti parocīgi noviržu formulas aprēķiniem, vidējie ir precīzi skaitļi (aprēķinos neveidosies noapaļošanas kļūdas), starpības ir nelielas, aprēķinus var veikt gandrīz bez kalkulatora, tāpēc aprēķini ir pat vienkāršāki, nekā lietojot momentu formulu.

7.2. tabulā ir dota darba tabula aprēķinam pēc 7.2. formulas.

7.2. tabula

Darba tabula korelācijas koeficienta aprēķināšanai pēc momentu formulas 7.1. piemēram

i	x_i	y_i	x_i^2	y_i^2	$x_i{*}y_i$	$x_i + y_i$	$(x_i + y_i)^2$
1.	*65*	*37*	*4225*	*1369*	*2405*	*102*	*10404*
2.	*60*	*30*	*3600*	*900*	*1800*	*90*	*8100*
3.	*66*	*36*	*4356*	*1296*	*2376*	*102*	*10404*
4.	*72*	*42*	*5184*	*1764*	*3024*	*114*	*12996*
5.	*51*	*30*	*2601*	*900*	*1530*	*81*	*6561*
6.	*58*	*32*	*3364*	*1024*	*1856*	*90*	*8100*
Σ	*372*	*207*	*23330*	*7253*	*12991*	*579*	*56565*

Pēdējās divas ailes darba tabulā ir aprēķinu pareizības kontrolei, korelācijas koeficienta aprēķināšanai tie nav nepieciešami. Pārbaudi veic pēc šādas sakarības: $\Sigma(x_i + y_i)^2 = \Sigma x_i^2 + 2\Sigma x_i y_i + \Sigma y_i^2$.[1] Piemērā šī sakarība piepildās: $56565 = 23330 + 2*12991 + 7253$.

$$r = \frac{12991 - \frac{372 * 207}{6}}{\sqrt{\left(23330 - \frac{372^2}{6}\right)\left(7253 - \frac{207^2}{6}\right)}} = 0{,}91$$

Svarīga ir iegūto rezultātu interpretācija. Korelācijas koeficients ir relatīvs rādītājs, tāpēc to var lietot vienu pašu (nav nepieciešams salīdzināt ar citiem korelācijas koeficientiem u.tml.) sakarību vērtēšanai.

[1] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 182. lpp.

Informācija, kas jāzina par korelācijas koeficientu:

1) korelācijas koeficienta iespējamās vērtības ir no –1 līdz +1. Ja aprēķinu rezultātā ir iegūts skaitlis, kas atrodas ārpus šīm robežām, tad tas liecina par kļūdu aprēķinos;
2) korelācijas koeficienta zīme norāda sakarības izmaiņu virzienu (koeficienta vērtība nerāda izmaiņu straujumu). Ja korelācijas koeficients ir ar "–" zīmi, tad tas nozīmē, ka, pieaugot neatkarīgās (X) pazīmes vērtībām, samazinās rezultatīvās (Y) pazīmes vērtības. Savukārt pozitīva korelācijas koeficienta vērtība norāda, ka, pieaugot X vērtībai, pieaug arī Y vērtība;
3) korelācijas koeficienta vērtība rāda, cik tuvu empīriskie novērojumu punkti atrodas teorētiskajai regresijas taisnei. Ja korelācijas koeficients ir –1 vai +1, tad tas nozīmē, ka sakarība ir funkcionāla – katrai X vērtībai ir tikai viena iespējamā Y vērtība (7.2.a) attēls, visi empīriskie novērojumu punkti ir uz regresijas taisnes);
4) ļoti striktu nosacījumu korelācijas koeficienta novērtēšanai nav. Jo tuvāk $|r|$ (r modulis – korelācijas koeficienta vērtība, ignorējot zīmi) ir vieniniekam, jo ciešāka korelācija. Jo tuvāk $|r|$ ir nullei, jo vājāka korelācija. Praksē bieži pieturas šādiem korelācijas koeficientu vērtību traktējumiem:
   - Ja $|r| > 0,8$, tad korelācija ir cieša;
   - Ja $0,5 < |r| < 0,8$, tad korelācija ir vidēja;
   - Ja $0,3 < |r| < 0,5$, tad korelācija ir vāja;
   - Ja $|r| < 0,3$, tad korelācija nepastāv.[1,2,3]

Iepriekš minētie traktējumi tomēr nedod atbildi uz to, vai aprēķinātais korelācijas koeficients ir statistiski nozīmīgs. Citiem vārdiem sakot, ir jāveic vēl papildus aprēķini, lai varētu apgalvot, ka sakarība pastāv ne tikai izlasē, pēc kuras datiem ir aprēķināts korelācijas koeficients, bet arī visā ģenerālkopā, vai, veicot atkārtotu izlasi (turpinot pētījumu) no tās pašas ģenerālkopas, saglabāsies iepriekš noteiktā sakarība.

7.2. Korelācijas koeficienta vērtēšana

Līdzīgi kā iepriekšējās tēmās, arī korelācijas koeficienta būtiskuma vērtēšana balstās uz ticamības intervāliem, salīdzinājumam izmantojot normālo (parasti

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 218. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 167. lpp.
[3] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 417.

Stjūdenta sadalījumu, jo aprēķini balstās uz izlases datiem). Ja no vienas ģenerālkopas veido daudzas izlases un izveido izlašu korelācijas koeficientu sadalījumu, tad iegūst asimetrisku sadalījumu. Īpaši asimetrisks šis sadalījums ir, ja korelācijas koeficienta vērtība tuvojas vieniniekam (ciešāka korelācija). Izskaidrojums tam – korelācijas koeficienta vērtības ir ierobežotas (no –1 līdz +1), bet normālais vai Stjudenta sadalījums ir neierobežots. Ja korelācijas koeficienta vērtība ir augsta, tad novirze uz vienu pusi var būt liela, bet uz otru pusi maza.

Līdzīgi kā salīdzinot divus relatīvos biežumus, arī pārbaudot korelācijas koeficienta būtiskumu, ir jāveic Fišera transformācija, kas ir sadalījuma, neatbilstoša normālajam, pārveidošana par normālo sadalījumu ar aritmētisko manipulāciju palīdzību.

Pārbaudot korelācijas koeficienta būtiskumu, ir jāveic vairāki soļi.

1. Formulē nulles un alternatīvo hipotēzi – H_0: $\rho = 0$ un H_1: $\rho \neq 0$ (abpusējai hipotēzei) vai H_1: $\rho > 0$ (pozitīva korelācija), vai H_1: $\rho < 0$ (negatīva korelācija). ρ (grieķu burts, izrunā „ro”) ir korelācijas koeficients ģenerālkopā.
2. Veic Fišera transformāciju pēc izlases datiem aprēķinātajam korelācijas koeficientam, izmantojot 7.3.formulu
$$z = \frac{1}{2}\ln\frac{1+r}{1-r} \qquad (7.3.)$$
Formulu var aprēķināt ar zinātnisko kalkulatoru, *Excel* programmā, ievadot formulu, vai sameklēt statistikas mācību grāmatās pielikumu tabulās.
3. Aprēķina transformētās korelācijas koeficienta vērtības – z reprezentācijas kļūdu, izmantojot 7.4.formulu:
$$s_Z = \frac{1}{\sqrt{n-3}} \qquad (7.4.)$$
kur n – novērojumu skaits.
4. Aprēķina empīrisko Stjūdenta kritēriju pēc 7.5. formulas:
$$t_{emp} = \frac{z}{s_Z} \qquad (7.5.)$$
5. Tabulās (vai ar *Excel* funkciju *„TINV”*) atrod t_{krit}. Tā vērtību nosaka izvēlētais būtiskuma līmenis α un brīvības pakāpju skaits $\upsilon=n-2$ (korelācijas koeficientam to aprēķina no novērojumu skaita, atņemot 2).
6. Izdara secinājumus par iegūtajiem rezultātiem. Ja $t_{emp}>t_{krit}$, tad nulles hipotēzi noraida, un tas nozīmē, ka sakarība pastāv arī ģenerālkopā. Ja $t_{emp}<t_{krit}$, tad nulles hipotēzi noraidīt nevar vai nu nepietiekami liela datu apjoma dēļ (ja tas tā ir, tad ievāc papildus informāciju un analīzi atkārto), vai arī ģenerālkopā šī pētāmā sakarība nepastāv.[1]

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 167. lpp.

7.1. piemēram tas ir šādi:

- $z = \frac{1}{2}\ln\frac{1+0{,}91}{1-0{,}91} = 1{,}53$;
- $s_Z = \frac{1}{\sqrt{6-3}} = 0{,}577$
- $t = \frac{1{,}53}{0{,}577} = 2{,}651$
- $t_{\alpha;\nu=n-2} = t_{0,05;4} = 2{,}776$

Nulles hipotēzi nevar noraidīt. datu savstarpējā sakarība nav pierādīta, pat neskatoties uz ļoti augsto korelācijas koeficientu vērtību. Šajā gadījumā tas nozīmē, ka novērojumu ir pārāk maz, bet sakarību varētu pierādīt ar iepriekš definēto nozīmības līmeni (α=0,05), ja palielinātu novērojumu skaitu. To var pamatot arī ar to, ka kritiskā vērtība ir tikai nedaudz lielāka par aprēķināto.

Vēl der apskatīt divas metodes, kuras var izmantot korelācijas koeficienta vērtēšanai.

Ja korelācija nav cieša ($|r|\leq 0{,}5$) un ir liels novērojumu skaits (n>50), tad t_{emp} var aprēķināt ar vienkāršākām formulām.

Korelācijas koeficienta reprezentācijas kļūdu aprēķina ar 7.6. formulu un empīrisko Stjūdenta kritērija vērtību ar 7.7. formulu:

$$s_r = \sqrt{\frac{1-r^2}{n-2}} \qquad (7.6.)$$

$$t_{emp} = \frac{r}{s_r} \qquad (7.7.)$$ [1]

Pārējā vērtēšanas procedūra ir analoga iepriekš aprakstītajai metodei – nulles un alternatīvās hipotēzes noformulēšana, t_{krit} atrašana un secinājumu izdarīšana.

Otra metode balstās uz ticamības intervāliem – ar iepriekšnoteiktu varbūtību atrod korelācijas koeficienta intervālu.

Visa matemātiskā procedūra būtu šāda:

- transformē pēc izlases datiem aprēķināto korelācijas koeficientu ar 7.3. formulu;
- aprēķinam z vērtības standartkļūdu ar 7.4. formulu;
- izvēlas būtiskuma līmeni α atbilstoši tam un brīvības pakāpēm;
- Stjūdenta kritērija kritisko vērtību tabulā nolasa t_{krit};
- aprēķina intervāla augšējo un zemāko robežu pēc 7.8. un 7.9. formulas:

$$z_z = z - ts_z \qquad (7.8.)$$

$$z_a = z + ts_z \qquad (7.9.)$$

- iegūtās intervāla robežas transformē atpakaļ (veic inverso transformāciju) pēc 7.10. formulas:

---

[1] Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 189. lpp.

$$r = \frac{e^{2z} - 1}{e^{2z} + 1} \quad (7.10.)$$

kur e – naturālo logaritmu bāze, konstante – 2,71...

- novērtē iegūtās vērtības:

$$r_z \leq \rho \leq r_a$$

Ja intervāls neietver nulli, tad nulles hipotēzi noraida, ja ietver, tad ar izvirzīto rezultātu nozīmības līmeni nulles hipotēzi noraidīt nevar.[1]

Paanalizējot kritisko vērtību tabulu, redzams – jo mazāks ir novērojumu skaits, jo ciešākai ir jābūt korelācijai, lai to varētu uzskatīt par statistiski nozīmīgu. Sakarības ciešums, ko izsaka korelācijas koeficients, un tā statistiskais nozīmīgums ir gan savstarpēji cieši saistīti, bet tomēr atšķirīgi traktējumi. Augsta korelācijas koeficienta vērtība drošāk nozīmē, ka korelācija pastāv, tomēr tas nenozīmē, ka atkārtojot pētījumu iegūsim līdzīgu rezultātu. Ja ir pierādīts, ka korelācijas koeficients ir statistiski nozīmīgs, tas nozīmē, ka, atkārtojot pētījumu, rezultāts būs līdzīgs. Problēmas var rasties, ja ir maza izlase, aprēķināta augsta korelācijas koeficienta vērtība ($|r|$ vērtība tuvojas vienam), bet tā nav pierādīta, ja $t_{emp} < t_{crit}$.(7.1. piemērā). Parasti, palielinot pētījuma apjomu, sakarību izdodas pierādīt.

Kā pētīt sakarību ciešumu, ja mainīgie tiek mērīti kārtas (rangu) skalā vai arī nominālajā skalā (atributīvas pazīmes), tiks apskatīts nākamajā apakšnodaļā.

7.3. Sakarību ciešuma mērīšanas neparametriskās metodes

Viens vai abi mainīgie var būt mērīti rangu skalā. Piemēram, eksperti sarindo izstrādājumus pēc kvalitātes, subjektīvi novērtējot to kvalitāti ballēs u.tml. Ir zināma varianšu secība, bet starpības starp diviem blakus esošiem rangiem nav identiskas. Pīrsona korelācijas koeficients šajā situācijā nederēs, ir jālieto kāds no rangu korelācijas koeficientiem. Šie koeficienti nav tik jūtīgi uz krasām varianšu vērtību izmaiņām. Arī situācijās, ja mainīgie tiek mērīti proporcionālajā vai intervālu skalā, bet mērījumu nav daudz un kāds būtiski atšķiras, tad var lietot rangu korelācijas koeficientu. Atšķirīgā variantes vērtība var rasties kā novērojuma, reģistrēšanas kļūda vai kādu netipisku apstākļu ietekmes rezultātā. Lietojot parametrisko metodi – Pīrsona korelācijas koeficientu, šāda variante no aprēķiniem ir jāizslēdz, bet analīzes tekstā tā ir jāizvērtē papildus. Lietojot rangu korelācijas koeficientus, to var neizslēgt.

Tālāk tiks apskatīti divi rangu korelācijas koeficientu (Spīrmena un Kendela) veidi.

---

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 234. lpp.

Spīrmena korelācijas koeficients balstās uz rangu pāru starpībām un to aprēķina pēc 7.11. formulas:

$$r_S = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)}, \tag{7.11.}$$

kur x_i un y_i – varianšu rangi (numuri pēc kārtas varianšu pieaugšanas secībā);
n – novērojumu skaits.

Ja kāda no pazīmēm – faktoriālā vai rezultatīvā – sākotnēji nav izteikta rangos, tad to pārveido rangu skalā.[1,2]

7.2. piemērs. Pārbaudīt studentu attieksmes pret bohēmisku dzīvesveidu un sekmības sakarības ciešumu.

Attieksme pret bohēmisko dzīvesveidu ir vērtēta pēc šādas sistēmas:

0 – nekādu izklaižu, tikai mācības;

1 – izklaide ļoti reti – īpašos gadījumos, primārās ir mācības;

2 – izklaide reti, ja netraucē mācībām;

3 – izklaide reizēm, mācības dažreiz var pagaidīt;

4 – izklaide samērā bieži, mācības gan jau kaut kad;

5 – izklaide bieži, mācības „nav zaķis, mežā neaizmuks";

6 – izklaide pāri visam – vai tad studenti mācās?!

Atbilstošais sekmju vērtējums iegūts kā pēdējās sesijas vidējā atzīme.
Tālāk doti 10 studentu aptaujas rezultāti.

Students	*Attieksme pret bohēmisko dzīvesveidu*	*Vidējā atzīme pēdējā sesijā*
1.	2	6,7
2.	4	5,3
3.	1	8,5
4.	2	7,4
5.	5	4,6
6.	3	5,5
7.	3	7,2
8.	0	8,3
9.	2	8,5
10.	6	4,2

Lai aprēķinātu Spīrmena rangu korelācijas koeficientu, iepriekšējie vērtējumi ir jāpārveido rangos (jāpiešķir vietas).

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 408. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 201. lpp.

$$r_S = 1 - \frac{6 * 307{,}3}{10 * \left(10^2 - 1\right)} = 1 - \frac{1843{,}8}{990} = 1 - 1{,}862 = -0{,}862$$

7.3. tabula

Spīrmena korelācijas koeficienta aprēķināšanas palīgtabula 7.2. piemēram

Students	*Attieksme pret bohēmisko dzīvesveidu, rangs*	*Vidējā atzīme pēdējā sesijā, rangs*	x_i-y_i	(x_i-y_i)^2
1.	*4.(3.-5.)*	*5.*	*-1*	*1*
2.	*8.*	*3.*	*5*	*25*
3.	*2.*	*9,5.(9.-10.)*	*-7,5*	*56,25*
4.	*4.(3.-5.)*	*7.*	*-3*	*9*
5.	*9.*	*2.*	*7*	*49*
6.	*6,5.(6.-7.)*	*4.*	*2,5*	*6,55*
7.	*6,5.(6.-7.)*	*6.*	*0,5*	*0,25*
8.	*1.*	*8.*	*-7*	*49*
9.	*4.(3.-5.)*	*9,5.(9.-10.)*	*-5,5*	*30,25*
10.	*10.*	*1.*	*9*	*81*
			0	*307,3*

Starpību summai ir jābūt nullei

Vai aprēķinātais korelācijas koeficients ir statistiski nozīmīgs, pārliecinās, salīdzinot aprēķināto korelācijas koeficientu ar kritisko vērtību.

Šo vērtību nosaka izvēlētais nozīmības līmenis (α) un novērojumu skaits (n).

10 novērojumiem ar α = 0,05 r_{krit}*= 0,648.*

Kā redzams, tas ir diezgan augsts un liecina par to, ka nopietnam statistiskajam pētījumam būtu jābūt lielākam (vairāk novērojumu).

Taču, neskatoties uz iepriekš minēto, nulles hipotēzi var noraidīt un atzīt, ka pastāv negatīva korelācija starp studentu bohēmisko dzīvesveidu un sekmību.

Otru rangu korelācijas koeficientu ir izstrādājis Kendels, un tas balstās uz to, cik ir nesakritību rangos. **Kendela korelācijas koeficientu** aprēķina pēc 7.12. formulas:

$$r_K = \frac{P - S}{0{,}5n\left(n - 1\right)}, \quad (7.12.)$$

kur P – par y_i rangiem lielāko rangu skaita summa visiem novērojumiem (i);

S – par y_i rangiem mazāko rangu skaita summa visiem novērojumiem (*i*).[1,2]

Lai aprēķinātu *P* un *S* lielumus, jāsaraksta faktoriālās pazīmes rangi pieaugošā secībā, tiem piekārto atbilstošos rezultējošās pazīmes rangus, tad skaita, cik no atbilstošā punkta pa labi ir lielāku un mazāku rangu rezultējošajai pazīmei, to ieraksta papildus rindās.

7.4. tabula

Kendela rangu korelācijas koeficienta aprēķināšanas palīgtabula 7.2. piemēram

x_i *rangs*	*1.*	*2.*	*4.*	*4.*	*4.*	*6,5.*	*6,5.*	*8.*	*9.*	*10.*	*Σ*
Atbilstošais y_i *rangs*	*8.*	*9,5.*	*5.*	*7.*	*9.5.*	*4.*	*6.*	*3.*	*2.*	*1.*	*X*
P	*2*	*0,5*	*3*	*1*	*0*	*1*	*0*	*0*	*0*	*X*	*7,5*
S	*7*	*7,5*	*4*	*5*	*5*	*3*	*3*	*2*	*1*	*X*	*37,5*
P+S	*9*	*8*	*7*	*6*	*5*	*4*	*3*	*2*	*1*	*X*	*45*

Šajā piemērā ir redzama zināma nenoteiktība, jo 4. x rangs ir trim novērojumiem un, atbilstoši mainot y_i *rangu secību, tiks iegūts atšķirīgs Kendela korelācijas koeficients.*

Pēdējā – P+S rinda kalpo kontrolei, katrā nākamajā ailē summai ir jābūt par vienu mazākai.

Skaitļus ievieto 7.12. formulā un aprēķina:

$$r_K = \frac{7{,}5 - 37{,}5}{0{,}5 * 10 * (10 - 1)} = \frac{-30}{45} = -0{,}667$$

Ja dalītās vietas sarindotu savādāk (dilstošā secībā), tad rezultāts būtu savādāks (7.5. tabula).

7.5. tabula

Kendela rangu korelācijas koeficienta aprēķināšanas palīgtabula 7.2. piemēram (2. variants, savādāk sarindojot dalītās vietas)

x_i *rangs*	*1.*	*2.*	*4.*	*4.*	*4.*	*6,5.*	*6,5.*	*8.*	*9.*	*10.*	*Σ*
Atbilstošais y_i *rangs*	*8.*	*9,5.*	*9,5.*	*7.*	*5.*	*6.*	*4.*	*3.*	*2.*	*1.*	*X*
P	*2*	*0,5*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*X*	*3,5*
S	*7*	*7,5*	*7*	*6*	*4*	*4*	*3*	*2*	*1*	*X*	*41,5*
P+S	*9*	*8*	*7*	*6*	*5*	*4*	*3*	*2*	*1*	*X*	*45*

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 205. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 411. lpp.

$$r_K = \frac{3,5 - 41,5}{0,5 * 10 * (10 - 1)} = \frac{-38}{45} = -0,844$$

Šis otrais vērtējums faktiski ir analogs Spīrmena korelācijas koeficientam.

Kendela korelācijas koeficientu ērtāk lietot, ja var abas pazīmes viennozīmīgi saranžēt, kad nav dalīto vietu.

Ja vismaz vienu pazīmi (faktoriālo vai rezultatīvo) nevar saranžēt, tā tiek mērīta nominālajā skalā, tad sakarības ciešuma mērīšanai lieto **kontingences koeficientu**. [1,2]

Tas ir, vai var novērot, ka kāda faktoriālās pazīmes nozīme sekmē vai, gluži pretēji, ierobežo rezultējošās pazīmes nozīmi, tad tiek runāts par pazīmju kontingenci.

Izšķir polihoros (*poli* – daudz nozīmju) un tetrahoros (*tetra* – četras nozīmes) kontingences koeficientus. Četras nozīmes veidojas, ja abas pazīmes ir alternatīvas.

Tagad tiks apskatīts Čuprova kontingences (polihoro) koeficients.

$$k = \sqrt{\frac{\chi^2}{n(k_x - 1)(k_y - 1)}}, \qquad (7.13.)$$

kur χ^2 – *hī* kvadrāts, aprēķina pēc 5.2.formulas:

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}},$$

kur n_{ij} – x_i un y_j klasei atbilstošā empīriskā frekvence;
$n`_{ij}$ – x_i un y_j klasei atbilstošā vienmērīga sadalījuma teorētiskā frekvence;
k_x un k_y – x un y pazīmju klašu skaits.[3]

[1] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 411. lpp.
[2] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 205. lpp.
[3] Liepa, I. (1974). *Biometrija: mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 206. lpp.

7.3. piemērs. Uzņēmuma personāla daļa veic pētījumu par strādnieku stresa līmeni, pildot savus darba pienākumus, atkarībā no to vecuma. Iegūtā informācija ir šāda:

Vecums	Augsts	Mērens	Zems	Kopā
Līdz 25 gadiem	4	3	18	25
25-45 gadi	5	3	46	54
Vairāk nekā 45 gadi	28	16	7	51
Kopā	37	22	71	130

Aprēķināt kontingences koeficientu.

Paskatoties piemēra datus, šķiet, ka vairāk streso vecāki darbinieki, bet – cik cieša ir šī sakarība un vai tā ir statistiski nozīmīga, tas ir jāaprēķina.

Vispirms aprēķina teorētiskās frekvences vienmērīgam sadalījumam pēc 7.14. formulas:

$$n_{tij}' = \frac{\sum_{i=1}^{k} n_{ij} * \sum_{j=1}^{m} n_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij}} \qquad (7.14.)$$

7.6. tabula

Teorētisko frekvenču aprēķināšana 7.3. piemēram

Vecums	*Augsts*		*Mērens*		*Zems*		*Kopā*
	Empīriskā frekvence	*Teorētiskā frekvence*	*Empīriskā frekvence*	*Teorētiskā frekvence*	*Empīriskā frekvence*	*Teorētiskā frekvence*	
Līdz 25 gadiem	*4*	*7,12*	*3*	*4,23*	*18*	*13,65*	*25*
25-45 gadi	*5*	*15,37*	*3*	*9,14*	*46*	*29,49*	*54*
Vairāk nekā 45 gadi	*28*	*14,52*	*16*	*8,63*	*7*	*27,85*	*51*
Kopā	*37*		*22*		*71*		*130*

37*51/130=14,52 – tā elementa, kuram rēķina atbilstošo frekvenci, ailes kopsummu reizina ar rindas kopsummu un dala ar visu novērojumu summu.

Tālāk izveido tabulu χ^2 aprēķināšanai.

7.7. tabula

Empīriskā χ^2 kritērija vērtības aprēķināšanas tabula

n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$(n_i - n_i')^2/n_i'$
4	7,12	-3,12	9,71	1,36
5	15,37	-10,37	107,52	7,00
28	14,52	13,48	181,83	12,53
3	4,23	-1,23	1,51	0,36
3	9,14	-6,14	37,68	4,12
16	8,63	7,37	54,31	6,29
18	13,65	4,35	18,89	1,38
46	29,49	16,51	272,50	9,24
7	27,85	-20,85	434,88	15,61
		0,00	0,00	57,90

Pēdējās ailes kopsumma ir empīriskā χ^2 vērtība. Ievieto skaitļus 7.13. formulā un atrod kontingences koeficienta vērtību:

$$k = \sqrt{\frac{57{,}90}{130(3-1)(3-1)}} = 0{,}33$$

Sakarība statistiski nozīmīga ir tad, ja $\chi^2_{emp} > \chi^2_{krit}$.

Kritisko kritērija vērtību nolasa tabulās vai atrod ar Excel funkciju CHINV izvēlētajam nozīmības līmenim (α) un brīvības pakāpēm ($\nu=(k-1)(m-1)$).

$\chi^2_{0,05:4}$=9,49, un tas ir būtiski mazāks par empīrisko kritērija vērtību (57,90), tātad sakarība ir statistiski nozīmīga – vecāki darbinieki vairāk streso, pildot savus darba pienākumus. Pēdējais secinājums izriet no pētījuma loģikas, un tas ir statistiskās hipotēzes „tulkojums” pētāmajā problēmā.

Ja sadalījumi ir kvantitatīvi, tad parasti paralēli korelācijas analīzei tiek veikta arī regresijas analīze, par tiks runāts nākamajā apakšnodaļā.

7.4. Regresijas analīze

Lēmumu pieņemšanai par sakarības ciešumu svarīgāk ir zināt sakarības formu – vienādojumu, kurš apraksta likumsakarību, kā, mainoties faktoriālās pazīmes vērtībām, mainās rezultatīvās pazīmes vērtības.

Par sakarības formām tika runāts, pētot korelācijas diagrammas (7.2. attēls).

Regresijas analīzē pēc empīriskajiem datiem izvēlas vienādojuma veidu un aprēķina vienādojuma konstantes.

Līdzīgi kā korelācijas analīzē, arī regresijas analīzē var izdalīt:

- pāru regresiju, kad pēta rezultatīvās pazīmes atkarību no vienas faktoriālās pazīmes. Vispārējā formā to pieraksta: $y = f(x)$ – vārdiski saka, ka y ir funkcija no x. Veicot konkrētu pētījumu, izvēlas konkrētu regresijas formu. Piemēram, ja empīrisko novērojumu punkti korelācijas diagrammā grupējas ap taisni, tad izvēlas lineāro funkciju. Lineārās funkcijas vienādojuma vispārējā forma ir:
$$y = a + bx, \tag{7.15.}$$
kur a un b ir vienādojuma konstantes, kuras regresijas analīzē aprēķina no empīriskajiem datiem;
- multiplā (daudzfaktoru) regresija, kad pēta, kā rezultatīvo pazīmi ietekmē divas vai vairāk pazīmes. Funkcijas vispārējais pieraksts būs $y = f(x_1,x_2,...x_m)$, bet konkrēti, piemēram, divfaktoru lineārās regresijas vienādojums būs:
$$y = a + b_1x_1 + b_2x_2, \tag{7.16.}$$
kur a, b_1 un b_2 – vienādojuma konstantes, kuras aprēķina pēc empīriskajiem (novērojumos) iegūtajiem datiem.

Pēc cita dalījuma (gan pāru, gan multiplā) regresija var būt:

- lineāra, pāru regresijai to ģeometriski var aprakstīt ar taisni, divfaktoru regresijai – ar plakni telpā, bet vairāk (3 un vairāk) faktoru regresijai nav ģeometriskas interpretācijas. Taču kopējā ideja ir tāda, ka vienādām faktoriālās pazīmes (regresora) izmaiņām atbilst vienādas rezultatīvās pazīmes (regresenta) izmaiņas;
- nelineāra (pāru regresijai to ģeometriski apraksta līkne, vairākfaktoru regresijai ģeometriski to aprakstīt ir sarežģīti vai pat neiespējami). Kopējā ideja ir tāda, ka vienādām faktoriālās pazīmes (regresora) izmaiņām pie dažādām to sākuma vērtībām neatbilst vienādas rezultatīvās pazīmes (regresenta) izmaiņas. Piemēram, ar hiperbolas funkciju:
$$y = a + \frac{b}{x} \tag{7.17.}$$
apraksta piesātinājuma situāciju, kad pie mazām x vērtībām y pieaugums ir straujāks nekā pie lielām. Šādu funkciju izmanto, prognozējot pieprasījuma izmaiņas jaunam produktam, no sākuma tirgus attīstās ļoti strauji – arvien vairāk pircēju vēlas šo preci iegādāties, bet pēc kāda laika tirgus tiek piesātināts un pieprasījums pēc šīs preces stabilizējas, pieauguma temps sarūk.

Vēl daudzfaktoru regresijas gadījumā izšķir **parciālo** regresiju. Parciāls nozīmē daļējs – no daudzfaktoru regresijas vienādojuma izdala un aprēķina atsevišķus (parciālos) vienfaktora regresijas vienādojumus, noskaidro. kāda ir viena (izdalīta) faktora tīrā ietekme uz rezultējošo pazīmi.

Regresijas analīzē izšķir trīs posmus:

1) noskaidro, kādai sakarības formai atbilst empīriskie dati, uzraksta regresijas vienādojumu vispārējā formā;
2) aprēķina regresijas koeficienta vērtības;
3) novērtē regresijas vienādojumu.

Ja aprēķinātais regresijas vienādojums ir statistiski nozīmīgs (ir noraidīta nulles hipotēze – regresijas koeficienti, vismaz viens, nav nulles), tad aprēķināto regresijas vienādojumu izmanto, lai pamatotu un pieņemtu lēmumus. Šīs darbības vairs nav tieši statistikas uzdevums, bet tas ir tieši tas, kāpēc tiek veikti statistiskie aprēķini. Statistika ir instruments citām nozarēm, dod zinātnisku pamatojumu lēmumiem.[1,2]

Sakarības formu izvēloties, visbiežāk pārbauda, vai var izmantot lineāro sakarību. 7.3. attēlā un tā analīzē bija pamatots, kāpēc bieži izvēlas lineāro regresijas formu nelineārās vietā, kaut arī ilgākā laika periodā vai plašākā datu apgabalā vairums sociāli ekonomisko parādību mainās nelineāri.

Korelācijas diagramma ir labs sākums sakarības formas noteikšanai. Izvēloties nelineāras sakarības formu manuālajos aprēķinos, ir ļoti svarīgas analītiskās ģeometrijas zināšanas (jāzina, kas ir un kā izskatās hiperbola, parabola, n pakāpes polinoms u.tml.). Mūsdienās izvēli ļoti atvieglo dators, kur piedāvātajā dialoga logā jau tiek piedāvāti vizuālie līkņu veidi un labāko regresijas vienādojumu varam piemeklēt ar kļūdu – mēģinājumu metodi.

Nelineārā regresija un datora lietošana tiks apskatīta vēlāk, bet nākamā apakšnodaļa ir veltīta lineāro pāru regresijas analīzei.

7.5. Vienfaktora lineārā regresija

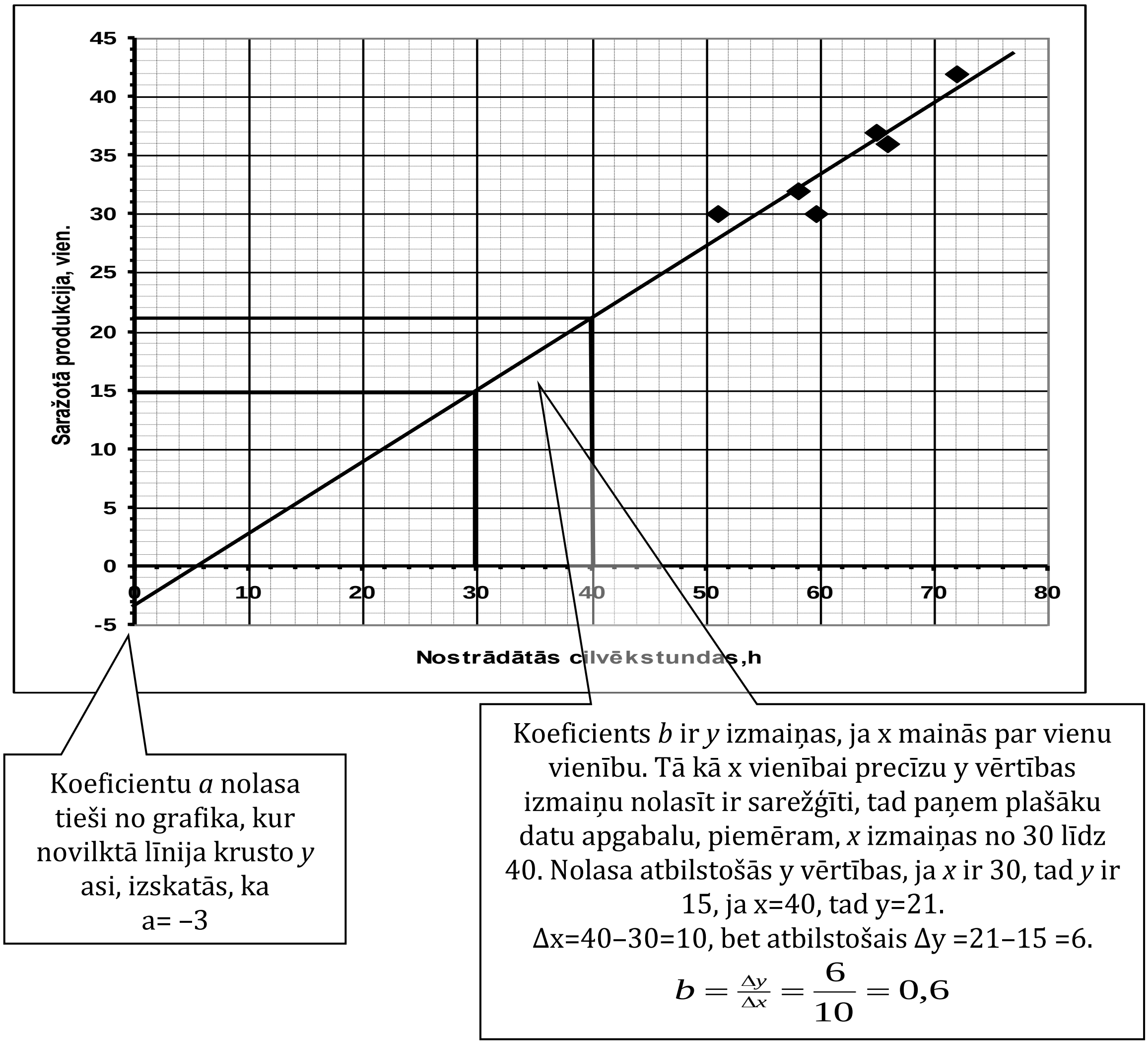
Manuālajos aprēķinos vienkāršākais veids, kā noteikt regresijas koeficientus, ir uzzīmēt korelācijas (punktu) diagrammu, pacensties novilkt taisni tā, lai tā būtu pēc iespējas tuvāk visiem empīriskajiem novērojumu punktiem un tad nolasīt (uzmērīt un aprēķināt) regresijas vienādojuma koeficientus grafikā. Metode ir neprecīza, puslīdz derīgus rezultātus var iegūt, ja korelācija ir cieša (empīrisko novērojumu punkti ir tuvu teorētiskajai regresijas taisnei). Metode ir subjektīva, vieniem un tiem pašiem datiem divi pētnieki nolasīs kaut nedaudz, bet tomēr atšķirīgus regresijas koeficientus.

---

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 171. lpp.
[2] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 423.

Nopietnos pētījumos šo metodi nelietoja arī tad, kad skaitļošanas tehnika nebija tik ļoti attīstīta kā mūsdienās. Taču šī metode tiks apskatīta, lai labāk saprastu regresijas vienādojumu būtību un koeficientu interpretāciju.

Ilustrācijai var izmantot 7.1. piemēra datus, kurus vēlreiz atzīmē diagrammā. Lai grafiskā metode izdotos, grafikam ir jāatvēl pietiekami daudz vietas. Labāk to darīt uz milimetru papīra, ja tāds nav pieejams, tad uz rūtiņu papīra. Obligāti ir jāievēro datu mērogs.



7.5. attēls. **Lineārās regresijas vienādojuma nolasīšana no punktu grafika**

“Nolasītais” vienādojums ir y = -3 + 0,6x.

Svarīga ir iegūtā vienādojuma ekonomiskā interpretācija. Uzdevums bija izpētīt saražotās produkcijas un patērētā laika sakarības. Neviens no šiem rādītājiem nevar būt negatīvs. No vienādojuma izriet, ka līdz 6 nostrādātajām cilvēkstundām ne tikai neko nesaražo, bet pat patērē. Empīrisko datu apgabals ir tālu no koordinātu sākumpunkta. Informācija par patērēto laiku un atbilstošo saražotās produkcijas apjomu ir no 51 līdz 72 cilvēkstundām un no 30 līdz 42 produkcijas vienībām.

Lai iegūtu regresijas vienādojumu, veic datu ekstrapolāciju. **Ekstrapolācija** ir nezināmu datu meklēšana ārpus zināmā datu apgabala, izmantojot zināmajā datu apgabalā konstatētās likumsakarības. Ekstrapolāciju lēmumu pieņemšanai parasti izmanto datu apgabalam, kas ir tuvu jau zināmajam datu apgabalam, piemēram, prognozes veido nākamajam gadam, nevis gadu desmitiem vai simtiem. Datu ekstrapolāciju lieto tikai tad, ja loģiskā analīze liecina, ka zināmajā datu apgabalā konstatētās likumsakarības saglabāsies arī tam tuvajos (lielākos vai mazākos) datu apgabalos.

Taču iepriekš minētais neattiecas uz a koeficienta atrašanu (aprēķināšanu). Šajā gadījumā *a* vērtība ir tuva nullei un iespējams, ka, palielinot novērojumu skaitu, iegūtais vienādojums būtu, kur koeficients *a* vienāds ar nulli.

Ja tomēr *a* ir izteikti negatīvs situācijā, kad *y* loģiski ir iespējams tikai pozitīvs, tad šo negatīvo vērtību var pamatot tā, ka patiesībā sakarība ir nelineāra, bet tiek izmantots viens tās posms, kuru tuvināti atzīst par lineāru.

Mazāko kvadrātu metode ir algebriska metode, kad ar matemātisku manipulāciju palīdzību atrod tādu regresijas taisnes pozīciju, kas nodrošina tās atrašanos maksimāli tuvu visiem empīrisko novērojumu punktiem. Tiek ievērots mazāko noviržu kvadrātu summas princips, tikai aritmētiskā vidējā vietā šeit ir regresijas taisne.

Ar formulu to var izteikt šādi:

$$Q_Z = \sum_i (y_i - \hat{y}_i)^2 = Q_{Z\,\min}\,, \qquad (7.17.)$$

kur y_i – empīrisko novērojumu y vērtības;

$\hat{y}_i$ – i novērojuma x vērtībai aprēķinātā teorētiskā y vērtība.

Uz šo ideju balstās izstrādātās formulas, bet arī manuālajos aprēķinos parasti neizmantoja matemātiskās pamatidejas aprēķinus, bet jau gatavas formulas. Tās var būt dažādās izveduma pakāpēs, viena no tādām ir 7.18. formula:

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} \qquad (7.18.)$$

a koeficientu aprēķina, pieņemot, ka regresijas taisne noteikti iet caur vidējām vērtībām:

$$a = \bar{y} - b\bar{x}$$ [1,2,3] (7.19.)

7.1. piemēra atrisināšanai nepieciešamie lielumi jau ir aprēķināti 7.2. tabulā.

$\sum xy$ – *xy reizinājumu summa ir 12991;*

$\sum x$ – *x summa ir 372;*

$\sum y$ – *y summa ir 207;*

$\sum x^2$ – *x kvadrātu summa ir 23 330;*

n – *6 novērojumi.*

$$b = \frac{12991 - \frac{372 * 207}{6}}{23330 - \frac{372^2}{6}} = \frac{157}{266} = 0{,}59,$$

*Kā redzams, analītiski aprēķinātais b koeficients nedod pretrunīgus secinājumus, bet tomēr atšķiras no grafikā nolasītā koeficienta 0,59–0,6=0,01 jeb 0.01/0,59*100 % =1,7 %, kas nav liela kļūda.*

$\bar{x} = \frac{372}{6} = 62$ *un* $\bar{y} = \frac{207}{6} = 34{,}5$

$a = 34{,}5 - 0{,}59 * 62 = -2{,}08$

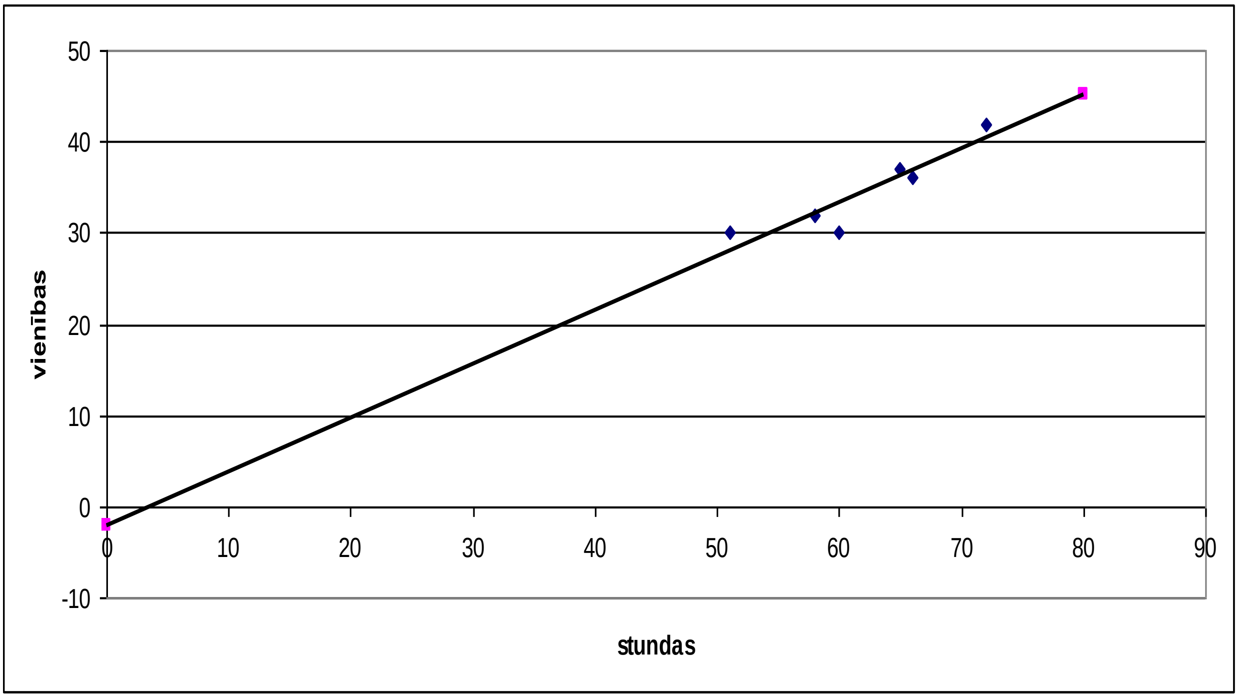
Aprēķinātais regresijas vienādojums ir: $y = -2{,}08 + 0{,}59x$

*Lai atliktu taisni, aprēķina divu punktu koordinātes. Viena x vērtība ir nulle (y=a=-2,08), bet otra ir lielākā x vērtība, kas tiek attēlota punktu diagrammā (x=80; y =-2,08+0,59*80=45,12).*

7.6. attēlā ir parādīta korelācijas (punktu) diagramma ar atlikto regresijas taisni.

---

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 423.
[2] Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 236. lpp.
[3] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 207. lpp.



7.6. attēls. **Korelācijas diagramma un aprēķinātā regresijas taisne 7.1. piemēram**

Par savu manuālo aprēķinu pareizību (vai nav pielaista kļūda aprēķinos) pārliecinās, atliekot aprēķināto regresijas taisni punktu diagrammā. Ja līnija iet ārpus punktu zonas vai ar savādāku slīpuma izvietojumu, tas nozīmē, ka ir pieļauta kļūda. Kā redzams, taisne iet pa vidu empīriskajiem novērojumu punktiem.

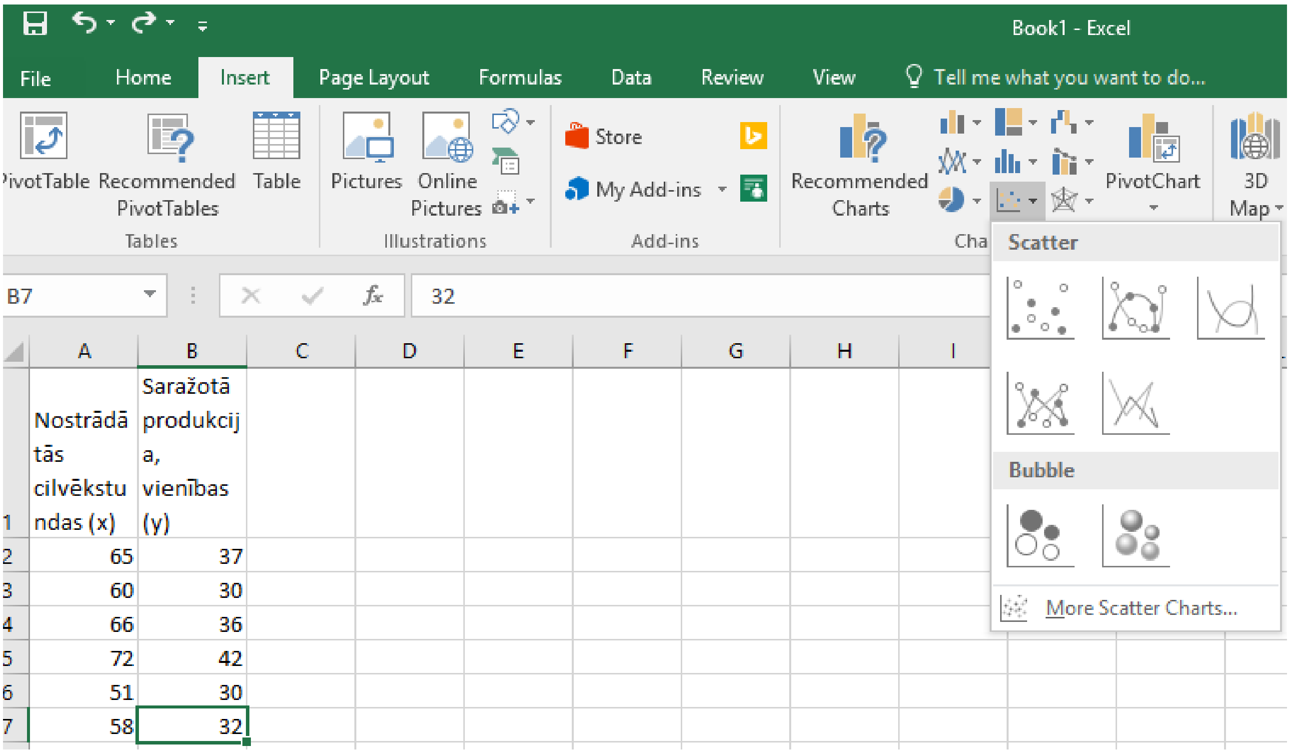
Tālāk tiks apskatīts, kā var iegūt regresijas vienādojumu ar programmas *Microsoft Excel* palīdzību.

7.6. Regresijas un korelācijas analīze ar programmu *Excel*

Regresijas vienādojumu ar *Excel* var atrast divos veidos. Ja ir pāru (viena neatkarīgā – *X* un viena atkarīgā pazīme – *Y*) regresija, tad var lietot grafisko metodi. Metode ir viegli saprotama, vizualizē datus grafikos, turklāt var viegli iegūt nelineārās regresijas vienādojumus.

Otra metode ir lietot datu analīzes rīku „*Regression*". Šī metode ļauj novērtēt regresijas vienādojuma koeficientu statistisko nozīmību, var iegūt daudzfaktoru lineārās regresijas vienādojumu. Ar šo metodi nevar iegūt nelineārās regresijas vienādojumus.

Vispirms tiks apskatīta 7.1. piemēra apstrāde ar grafisko metodi.



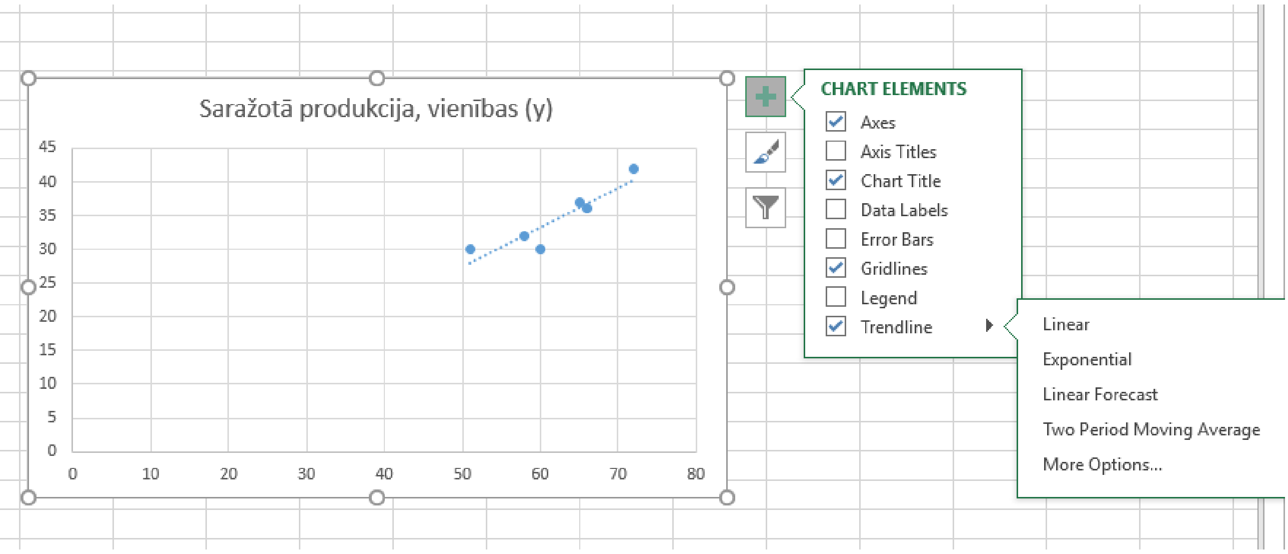
7.7. attēls. **Korelācijas diagrammas izveide ar *Excel***

Vispirms pievieno izkaisīto (*XY Scatter*) diagrammu, izvēlas variantu bez savienojuma līnijām (bez empīriskās regresijas līnijas).

Parasti programma *Excel* pati atrod datus, ja nē, tad, pildot dialoga logos noteiktās prasības, norāda, vai datu apgabals ir jākoriģē.

Tālāk izpilda nepieciešamās formatēšanas darbības – pievieno asu apzīmējumus, noņem fona krāsu, novāc nevajadzīgos virsrakstus (darbības, analogas aprakstītajām 2. nodaļā, veidojot histogrammu) u. tml.

Kad korelācijas diagramma izveidota, tad aktīvam grafikam izvēlnē „*Chart*” pievieno trenda līniju (7.8. attēls).



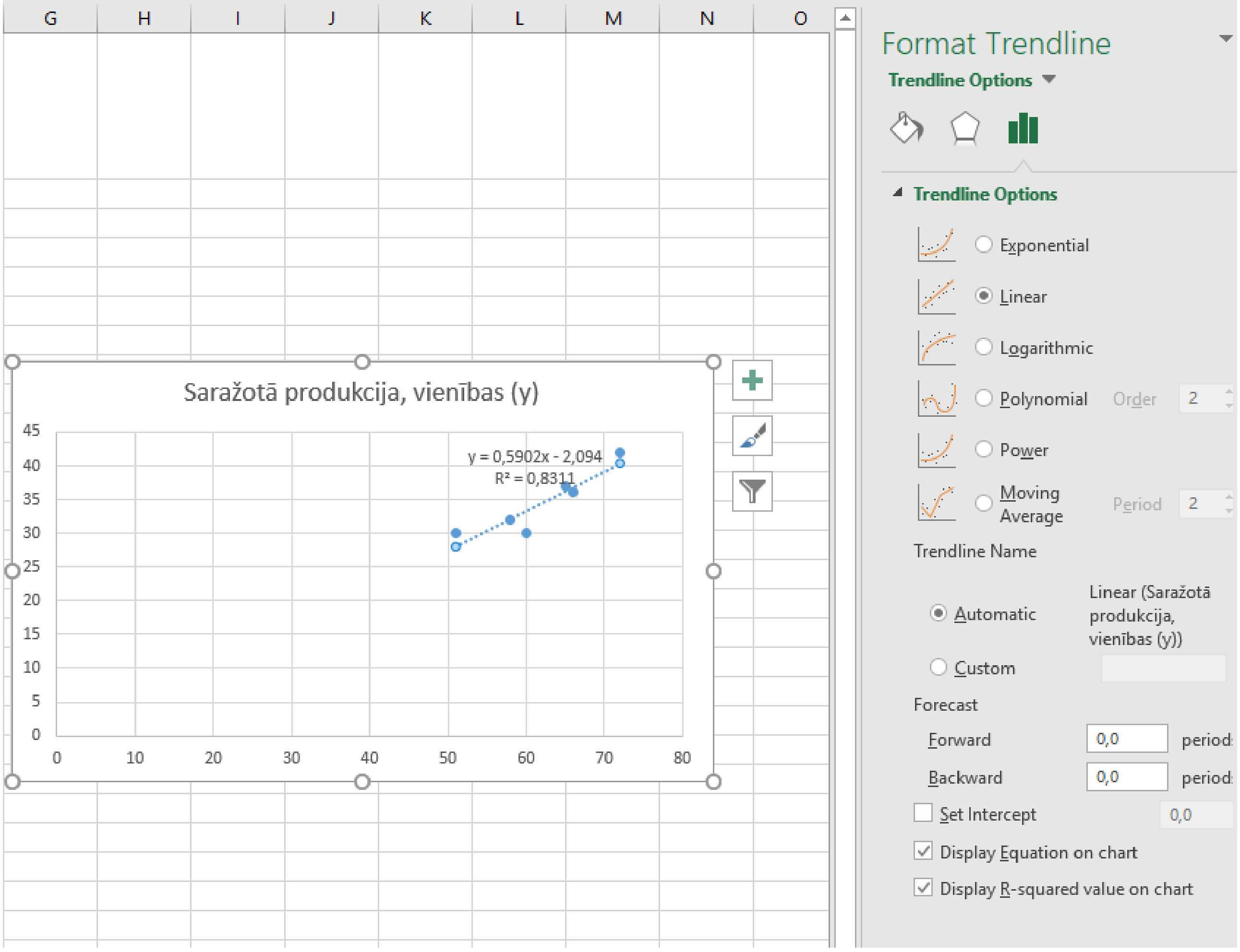
7.8. attēls. **Trenda līnijas pievienošanas dialoga logs**

Kā redzams, mūsdienās vairs nav nepieciešamas analītiskās ģeometrijas zināšanas – paskatās, kurai no piedāvātajām līknēm līdzīgāks ir empīrisko novērojumu punktu izvietojums, un izvēlas atbilstošo regresijas tipu.

Izvēloties regresijas formu (tipu), primārais ir ekonomiskā loģika – vai var izskaidrot, kāpēc *y* mainās pēc šādas sakarības. Visbiežāk lieto lineāro regresiju, retāk logaritmisko, ar kuru var aprakstīt piesātinājuma funkciju. Piemēram, prognozējot pārdošanas apjomus, var paredzēt, ka tie, ieviešot inovatīvu produktu, sākumā pieaugs strauji, bet tad tirgus piesātināsies un pieauguma tempi saruks.

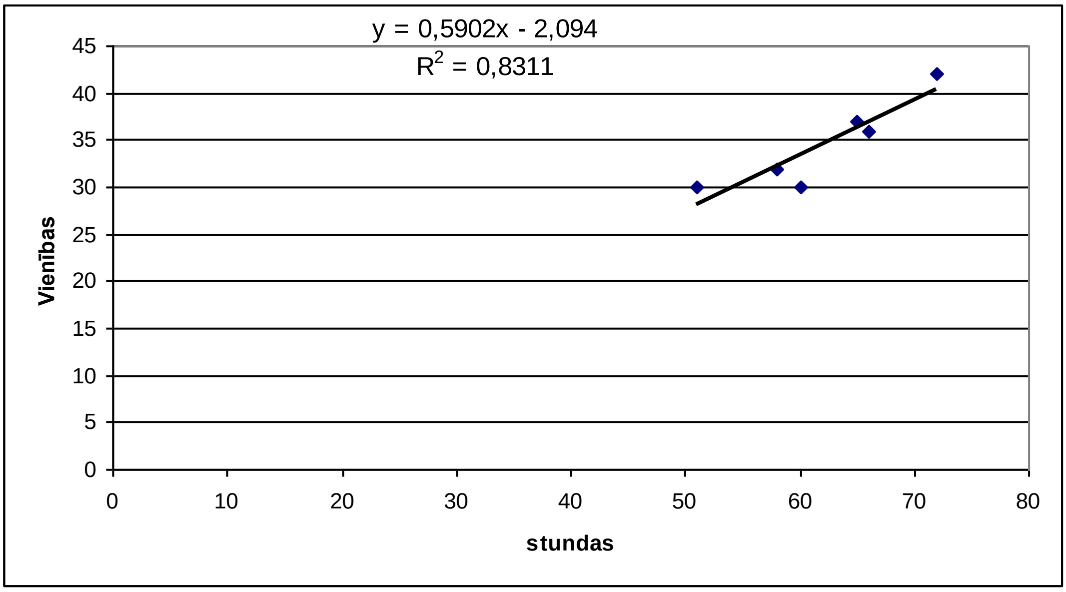
Otrs faktors, kas noteiks regresijas formas izvēli, ir determinācijas koeficients (*R-squred*). Tas ir korelācijas koeficienta kvadrāts. Kāpinot skaitli kvadrātā, tam pazūd negatīvā zīme, ja tāda ir. Determinācijas koeficients svārstās robežās no 0 līdz vienam un rāda ar *x* izmaiņām izskaidroto *y* izmaiņu īpatsvaru kopējās y izmaiņās. Labāks modelis ir tāds, kuram determinācijas koeficients ir lielāks.

Lai parādītu vienādojumu un determinācijas koeficientu, dialoga logā nospiež cilni „*Options*" un tajā atzīmē – parādīt vienādojumu un determinācijas koeficientu grafikā (7.9. attēls).



7.9. attēls. **Dialoga logs regresijas vienādojuma un determinācijas koeficienta aprēķināšanai**

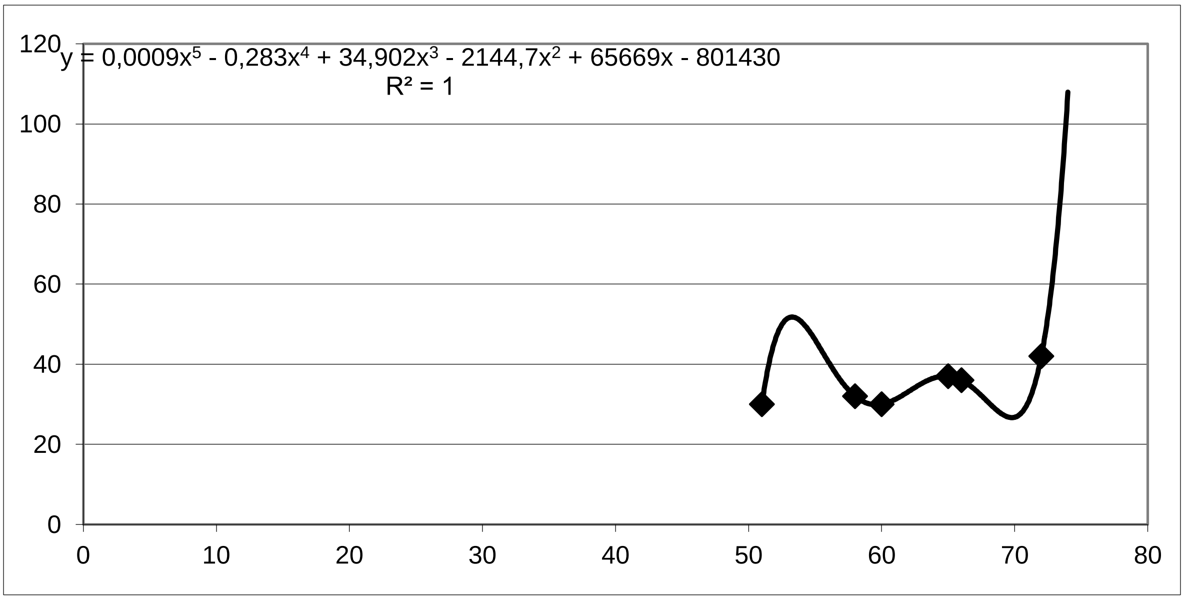
7.10. attēlā ir parādīts aprēķinātais vienādojums un determinācijas koeficients.



7.10. attēls. **Ar grafisko metodi iegūtais regresijas un korelācijas analīzes rezultāts**

Vienādojums sakrīt (nelielas nesakritības noapaļošanas dēļ) ar manuāli aprēķināto vienādojumu (pirms 7.6. attēla). Korelācijas koeficients, aprēķināts pēc 7.1. un 7.2. tabulas, ir 0,91. Kāpinot kvadrātā, tika iegūts 0,83.

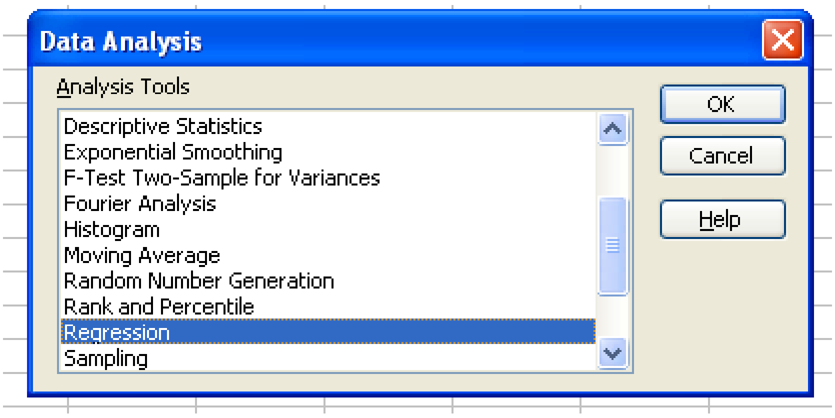
7.11. attēlā ir parādīta loģiski nekorekti veikta regresijas analīze. Izvēlēts ir 6 pakāpes polinoms. Iegūtais determinācijas koeficients ir 1, tas nozīmē, ka visi empīriskā novērojuma punkti atrodas uz teorētiskās regresijas līnijas. No šī rādītāja viedokļa ideāli, bet tas ir apmēram tā, kā novilkt taisni caur diviem punktiem (divi novērojumi), bet nav zināms, vai tā ir likumsakarība, vai nejaušība. Tāpat arī šeit novērojumu ir pārāk maz, lai iegūtu statistiski ticamus rezultātus. Bet galvenais ir tas, ka nevar atrast kādu loģisku izskaidrojumu, kāpēc no 50 līdz apmēram 55 nostrādātām cilvēkstundām ražošanas apjoms strauji aug, bet, ja nostrādā nedaudz vairāk, tad saražo mazāk. Grafikā ir pievienota vēl viena iespēja – prognoze vēl divām papildus cilvēkstundām.



Polinomam ir raksturīgas straujas izmaiņas, ja *x* iziet ārpus zināmā datu apgabala.

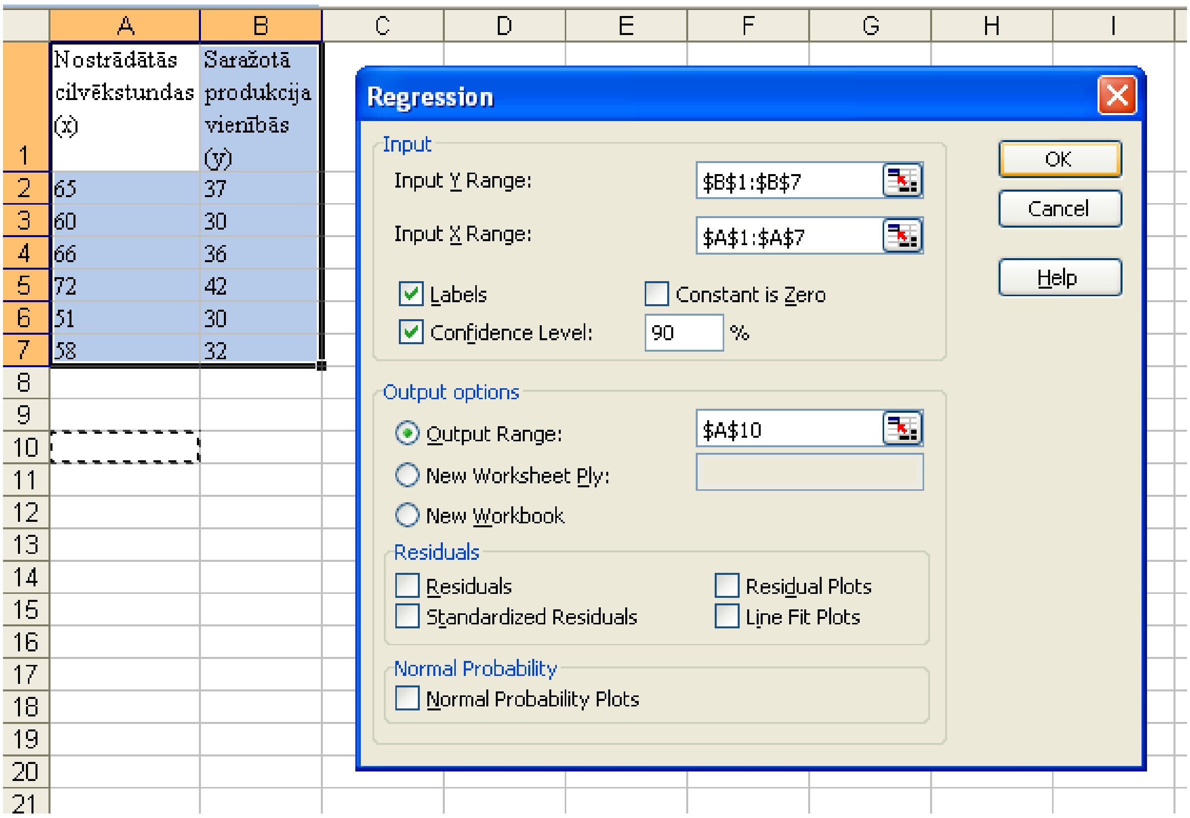
7.11. attēls. **7.1. piemēra dati, aprēķināti ar 6. pakāpes polinomu, un prognoze divām vienībām uz priekšu**

Ja nepieciešama dziļāka analīze, tad izvēlas datu analīzes rīku "*Regression*" (7.12. attēls).



7.12. attēls. **Dialoga logs lineārās regresijas veikšanai**

Nākamajā 7.13. attēlā ir parādīts *Excel* darba lapas fragments ar aizpildītu dialoga logu regresijas analīzes veikšanai.



7.13. attēls. **Dialoga logs ar norādītajām datu adresēm un citiem analīzes parametriem**

Norādot *X* un *Y* apgabalus, jāuzmanās tos nesajaukt. Var iegūt apgriezto vienādojumu – *x=f(y)*.

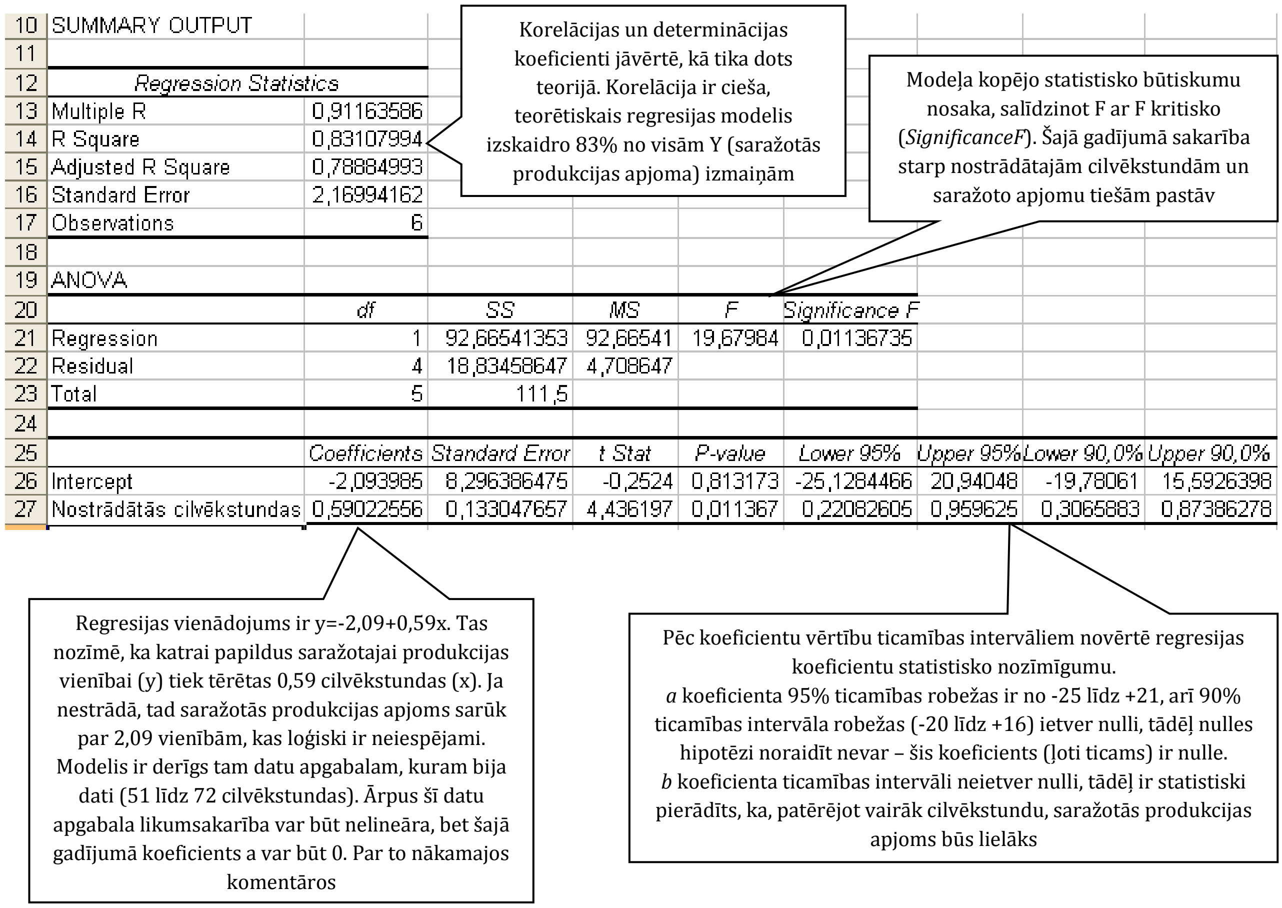
Rekomendē norādīt datus ar apzīmējumiem, ieliekot ķeksīti pret „*Labels*". Ja ir ievadīti tikai dati, bet ir atzīmēts „*Labels*", tad programma pirmos skaitļus uzskata par apzīmējumiem un aprēķinos neiekļauj.

Ticamības līmeni dators automātiski piedāvā 95 %, bet, nemainot šo lodziņu, tad koeficientu 95 % ticamības robežas tiks parādītas divreiz.

Ja datu ir nedaudz (tā kā piemērā), tad nav vērts datus izvadīt citā darba lapā (datora piedāvātā izvēle), tādēļ norāda datu izvades sākumšūnu (uz leju un pa labi ir jābūt brīvam).

Ja atzīmē atlikumus (*Residuals*), tad var novērtēt, vai piepildās regresijas analīzes pieņēmumi. Tie šajā grāmatā nav apskatīti.

Nākamajā 7.14. attēlā ir dots ar *Excel* iegūtais aprēķins un komentāri.



SUMMARY OUTPUT

	Regression Statistics	
13	Multiple R	0,91163586
14	R Square	0,83107994
15	Adjusted R Square	0,78884993
16	Standard Error	2,16994162
17	Observations	6

ANOVA

20		df	SS	MS	F	Significance F
21	Regression	1	92,66541353	92,66541	19,67984	0,01136735
22	Residual	4	18,83458647	4,708647		
23	Total	5	111,5			

25		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 90,0%	Upper 90,0%
26	Intercept	-2,093985	8,296386475	-0,2524	0,813173	-25,1284466	20,94048	-19,78061	15,5926398
27	Nostrādātās cilvēkstundas	0,59022556	0,133047657	4,436197	0,011367	0,22082605	0,959625	0,3065883	0,87386278

7.14. attēls. **Ar „*Regression*" iegūtie rezultāti ar komentāriem**

Manuāli netika aprēķināta dispersijas analīzes tabula (izdarīt to var). Kopējā noviržu kvadrātu summa ir 111,5, ar regresijas modeli izskaidrotā noviržu kvadrātu summa ir 92,7. Izskaidroto noviržu kvadrātus summas īpatsvars ir 92,7/111,5=0,83 (determinācijas koeficients).

Vēl tabulās ir vairāki rādītāji, kuri nozīmē apmēram to pašu, ko komentāros izskaidrotie rādītāji. Arī „*P-value*" rāda to pašu, ko ticamības intervāli – nulles hipotēzi noraida, ja tās varbūtība ir mazāka par 0,05. *a* koeficientam tā ir 0,81, tātad šis koeficients visticamāk ir nulle.

Ar regresijas analīzes rīku var veikt arī daudzfaktoru regresijas analīzi.

7.7. Daudzfaktoru korelācijas un regresijas analīze

Daudzfaktoru korelācijas un regresijas analīzē ir divas vai vairāk faktoriālās (neatkarīgās) pazīmes. Rezultatīvā (atkarīgā) pazīme vienmēr ir tikai viena. Manuāli aprēķināt divfaktoru korelācijas koeficientu var ar 7.20. formulu no pāru korelācijas koeficientiem:

$$r = \sqrt{\frac{{r_{yx_1}}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2} + {r_{yx_2}}^2}{1 - {r_{x_1x_2}}^2}}, \qquad (7.20.)$$

kur indeksi pie *r* norāda, starp kādām pazīmēm ir aprēķināta pāru korelācija.[1]

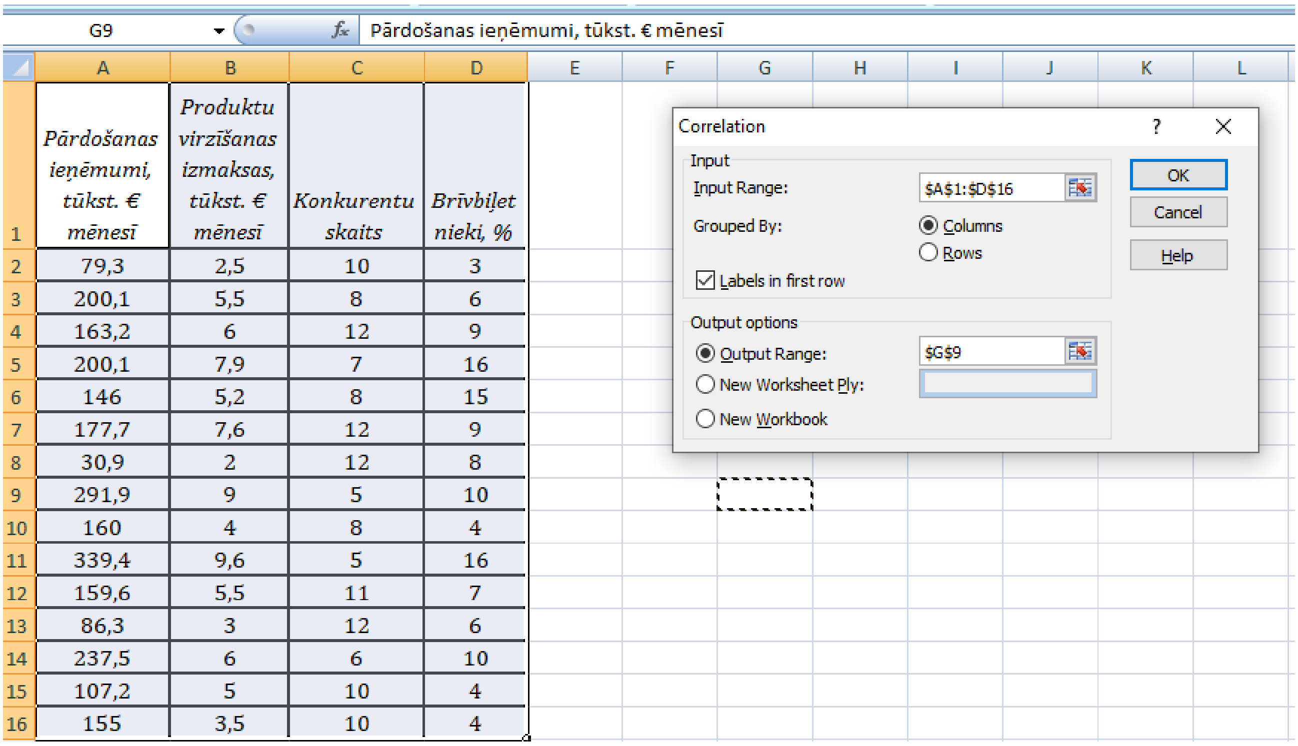
Daudzfaktoru regresijas vienādojuma konstantes manuāli atrod no normālvienādojumu sistēmas ar Gausa metodi.[2] Tas ir darbietilpīgi un prasa labas iemaņas matemātikā. Mūsdienās šo darbu ērti var veikt ar *Microsoft Excel*. Tas tiks apskatīts uz 7.4. piemēra bāzes.

7.4. piemērs. Uzņēmums sniedz sociāli nozīmīgus pakalpojumus tirgus apstākļos un vēlas pārbaudīt, kā mainās pārdošanas ieņēmumi atkarībā no produktu virzības izmaksām, konkurentu skaita attiecīgajā reģionā un brīvbiļetnieku (noteikta klientu kategorija, kas atbilstoši likumdošanai ir jāapkalpo bez maksas) īpatsvara. Informācija par 15 darbības reģioniem ir šāda:

Pārdošanas ieņēmumi, tūkst. € mēnesī	*Produktu virzīšanas izmaksas, tūkst. € mēnesī*	*Konkurentu skaits*	*Brīvbiļetnieki, %*
79,3	2,5	10	3
200,1	5,5	8	6
163,2	6	12	9
200,1	7,9	7	16
146	5,2	8	15
177,7	7,6	12	9
30,9	2	12	8
291,9	9	5	10
160	4	8	4
339,4	9,6	5	16
159,6	5,5	11	7
86,3	3	12	6
237,5	6	6	10
107,2	5	10	4
155	3,5	10	4

Vispirms ir jānovērtē pazīmju pāru korelācija. To izpilda ar datu analīzes rīku „*Correlation*” (7.15. attēls).

[1] Raščevska, M., Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 176. lpp.
[2] Krastiņš, O. (1998). *Statistika un ekonometrija: mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 258. lpp.



	A	B	C	D
1	*Pārdošanas ieņēmumi, tūkst. € mēnesī*	*Produktu virzīšanas izmaksas, tūkst. € mēnesī*	*Konkurentu skaits*	*Brīvbiļet nieki, %*
2	79,3	2,5	10	3
3	200,1	5,5	8	6
4	163,2	6	12	9
5	200,1	7,9	7	16
6	146	5,2	8	15
7	177,7	7,6	12	9
8	30,9	2	12	8
9	291,9	9	5	10
10	160	4	8	4
11	339,4	9,6	5	16
12	159,6	5,5	11	7
13	86,3	3	12	6
14	237,5	6	6	10
15	107,2	5	10	4
16	155	3,5	10	4

7.15. attēls. ***Excel*** **darba lapas fragments ar dialoga logu korelāciju aprēķināšanai**

7.8. tabulā ir parādīti iegūtie pāru korelācijas koeficienti. Tā kā korelācija starp pazīmēm *y* un *x* ir tāda pati kā starp *x* un *y*, korelācijas matricā ir aizpildīta tikai viena puse.

7.8. tabula

Ar datu analīzes rīku iegūta korelācijas koeficientu matrica 7.4. piemēram

	y	*x1*	*x2*	*x3*
y	1			
x1	0,891297	1		
x2	-0,79828	-0,61744	1	
x3	0,561158	0,689676	-0,50478	1

Šo matricu izmanto, lai izvēlētos mainīgos regresijas analīzei. Faktoriālajām pazīmēm vajadzētu cieši korelēt ar rezultatīvo pazīmi un savstarpēji korelēt vāji. Ja faktoriālās pazīmes savstarpēji cieši korelē (r>0,8), tad šo situāciju sauc par multikolinearitāti un vienu no savstarpēji cieši korelējošajām pazīmēm regresijas analīzē neiekļauj.[1] Pirmkārt, regresijas analīzē atstāj to faktoriālo pazīmi, kuras ietekmi var loģiski labāk pamatot, otrkārt, kura ciešāk korelē ar rezultatīvo pazīmi. Šajā gadījumā X_1 cieši korelē ar Y, arī X_2 un Y ir gandrīz 0,8, X_3 ar Y korelē vidēji cieši. Faktoriālās pazīmes savā starpā korelē vidēji cieši. Vadoties pēc korelācijas analīzes, regresijas analīzē atstāj visas pazīmes, kaut arī situācija nav ideāla. Tālāk dara tā, kā bija aprakstīts vienfaktora lineārās regresijas gadījumā, tikai *X* ievada ar datu apgabalu (vairākām blakus esošām ailēm). 7.16. attēlā ir parādīti iegūtie rezultāti.

[1] Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press. p. 484.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,9528295
R Square	0,907884
Adjusted R Square	0,8827615
Standard Error	27,384001
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	81298,399	27099,4663	36,13824	5,45E-06
Residual	11	8248,7185	749,883501		
Total	14	89547,117			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	172,34005	51,380996	3,35415934	0,00643	59,25124	285,4289	59,25124	285,4289
x1	25,949935	4,8774457	5,32039441	0,000245	15,21475	36,68512	15,21475	36,68512
x2	-13,23841	3,6862391	-3,5913041	0,004234	-21,3518	-5,12505	-21,3518	-5,12505
x3	-3,040608	2,3424807	-1,2980291	0,220834	-8,19637	2,115157	-8,19637	2,115157

7.16. attēls. **Ar datu analīze rīku „*Regression*” iegūtie rezultāti**

Iegūtais regresijas vienādojums ir: $y=172,34+25,9x_1-13,2x_2-3,04x_3$.

Apskatot koeficientu ticamības intervālus, ar 95% ticamības līmeni nulles hipotēzi nevar noraidīt b_3 koeficientam. „Brīvbiļetnieku” ietekme uz pārdošanas ieņēmumiem nav pierādīta. Ar modeli tika izskaidrots nepilns 91 % (*R Square*) no kopējās rezultātu izkliedes.

Glosārijs

Latviski	*Angliski*	*Krieviski*	*Skaidrojums*
1	*2*	*3*	*4*
Atkarīgais mainīgais	Dependent variable	Зависимая переменная	Pazīme, kura mainās atkarībā no kādas citas pazīmes izmaiņām
Daudzfaktoru (multiplā) korelācija	Multiple correlation	Множественная корреляция	Sakarība, kad atkarīgā (rezultatīvā) pazīme mainās atkarībā no vairākām neatkarīgām (faktoriālām) pazīmēm
daudzfaktoru (multiplā) regresija	Multiple regression	Множественная регрессия	Matemātiskā sakarības forma, ja ir vairākas faktoriālās pazīmes
Determinācijas koeficients	R square	Коэффициент детерминации	Korelācijas koeficients kvadrātā – rāda modeļa izskaidroto noviržu kvadrātu summas īpatsvaru kopējo noviržu kvadrātu summā
Ekstrapolācija	Ekstarpollation	Экстраполяция	Nezināmu datu meklēšana ārpus zināmā datu apgabala, esošās tendences attiecināšana uz nākotni
Empīriskā regresija	Empirical regression	Эмпирическая регрессия	Empīriski novērotā sakarība – empīriskie novērojumu punkti, savienoti ar lauztu līniju
Kendela rangu korelācijas koeficients	Kendel`s rank coefficient of correlation	Коэффициент ранговой корреляции Кендела	Sakarības ciešuma mērīšanas koeficients, ja pazīmes tiek mērītas rangu (ordinārajā) skalā, izstrādājis Kendels
Kontingences koeficients	Coefficient of contingency	Коэффициент сопряженности	Sakarības ciešuma mērs, ja pazīmes tiek mērītas nominālajā skalā
Korelācijas diagramma	Scatter diagramm	Диаграмма корреляции	Punktu diagramma – vizuāli parāda sakarības formu un ciešumu
Korelācijas koeficients (Pīrsona)	Pearson coefficient of correlation	Коэффициент корреляции (Пирсона)	Lineāras sakarības ciešuma mērs, lieto, ja mērījumi ir intervālu vai proporcionālajā skalā
Lineārā regresija	Linear regression	Линейная регрессия	Regresijas matemātisko sakarību izsaka ar taisnes vienādojumu – vienādām x izmaiņām ir vienādas y izmaiņas neatkarīgi no sākumpunkta

1	2	3	4
Mazāko kvadrātu metode	The least - squares method	Метод наименьших квадратов	Regresijas koeficientu aprēķināšanas metode, kas paredz aprēķināt regresijas līniju, kas iet pēc iespējas tuvāk visiem empīriskā novērojuma punktiem
Neatkarīgais mainīgais	Independent variable	Независимая пременная	Pazīme, kura ietekmē citas pazīmes vērtības
Negatīva sakarība (korelācija)	Negative relationship (correlation)	Негативная (отрицательная) связь (корреляция)	X vērtībām pieaugot, samazinās Y vērtības
Nelineārā regresija	Non linear regression	Нелинеарная регрессия	Regresijas matemātisko formu izsaka līknes vienādojums – pie atšķirīgām sākotnējām X vērtībām vienādas X izmaiņas izsauc atšķirīgas Y izmaiņas
Parciālā regresija	Partial regression	Парциальная регрессия	No daudzfaktoru regresijas vienādojuma izdalīts viena faktora regresijas vienādojums atsevišķa faktora ietekmes izvērtēšanai
Pāru korelācija	Simple correlation	Парная корреляция	Sakarība starp vienu neatkarīgo (X) un vienu atkarīgo (Y) pazīmēm
Pāru regresija	Simple regression	Парная регрессия	Sakarības matemātiskā forma starp vienu neatkarīgo (X) un vienu atkarīgo (Y) pazīmēm
Pozitīva sakarība (korelācija)	Positive relationship (correlation)	Позитивная (положительная) связь (корреляция)	X vērtībām pieaugot, pieaug arī Y vērtības
Rangs	Rank	Ранг	Variantes kārtas numurs pieaugošā secībā. Pazīmes var saranžēt, starp blakus esošiem rangiem nav vienādas atšķirības
Regresija	Regression	Регрессия	Sakarības matemātiskā forma
Regresijas brīvais loceklis (a koeficients)	Intercept	У- пересечение	Regresijas korelācijas koeficients, kurš nav saistīts ar matemātiskām darbībām ar neatkarīgo mainīgo. Y vērtība, ja X ir nulle
Regresijas slīpuma koeficients (b)	Slope	Коэффициент наклона	Lineārās regresijas koeficients, kurš parāda Y izmaiņas, ja X mainās par vienu vienību
Regresijas vienādojums	Regression equition	Уравнение регрессии	Sakarības matemātiskā forma vienādojuma veidā
Spīrmena rangu korelācijas koeficients	Spearman`s rank coefficient of correlation	Коэффициент ранговой корреляции Спирмена	Sakarības ciešuma mērīšanas koeficients, ja pazīmes tiek mērītas rangu (ordinārajā) skalā, izstrādājis Spīrmens

PIELIKUMS

PAŠPĀRBAUDES JAUTĀJUMI

Atbildiet par apgalvojuma patiesumu ar „Jā” vai „Nē”.

1. nodaļa

1. Statistiskās izpētes objekts ir pētāmais uzņēmums.
2. Statistiskās izpētes objekts ir Rēzeknes novada zemnieku saimniecības.
3. Statistiskā kopa ir identificējama pēc kāda pazīmju kopuma, kas to atšķir no citu kopu vienībām.
4. Statistiskajā kopa visām vienībām ir vienādas pētāmās vērtības.
5. Studentu aptaujā bija jautājums par viņu muzikālo gaumi (kāda stila mūzika viņiem patīk), atbildes tika mērītas ordinārajā skalā.
6. Pētījumā bija jautājums par vieglo automobiļu skaitu mājsaimniecībā, iegūtais sadalījums ir diskrēts.
7. Izlase (paraugkopa) statistiskajā pētījumā ir atlasīti labākie studenti, kuri pārstāvēs akadēmiju universiādē.
8. Vēlētāju politiskās simpātijas (atbalstītās partijas) ir nominālās skalas piemērs.
9. Hipotētiskā kopa ir teorētiska, tā nepastāv reālajā vidē.
10. Objekta struktūras izmaiņas laikā vislabāk attēlot ar „Stacked column” vai „Stacked bar” grafikiem.

2. nodaļa

1. Empīriskā sadalījuma biežumus aprēķina ar īpašām matemātiskām formulām.
2. Relatīvos biežumus (frekvences) lieto, ja jāsalīdzina divu nevienāda apjoma kopu sadalījumi.
3. Analizējot pircēju pirkumu lielumu, tiek veidota intervālu variācijas rinda. Ar Breksa formulu aprēķinātais optimālais intervāla garums ir 8,965 €. Veidojot intervālus grupēšanai, labāk ir lietot šo lielumu, un nebūtu pareizi noapaļot uz 10 €.
4. Histogramma ir stabiņu diagramma – uz abscisu ass tiek attēlotas pazīmes vērtības (intervālu robežvērtības), bet uz ordinātas – novērojumu biežumi.
5. Poligonu zīmē, izmantojot grafiku veidu „Radar”.
6. Ja gradācijas klase ir pārstāvēta ar 6 novērojumiem, kopumā ir 60 novērojumu, tad relatīvais biežums procentos būs aprēķināts šādi:
$$\frac{60 \quad kop\bar{e}jais \quad nov\bar{e}rojumu \quad skaits}{6 \quad klases} = 10\,\%$$
7. Kumulatīvie biežumi paliek tajā pašā līmenī, ja gradācijas klase ir tukša (nav reģistrēts neviens novērojums).
8. Aprakstot sadalījuma histogrammu, vienmēr piemin (izceļ) labāk pārstāvētās klases.
9. Grupēšanu lieto tikai, lai atvieglotu statistisko rādītāju (lokācijas rādītāju) aprēķināšanu, tiem nav pastāvīgas nozīmes analīzē.
10. Ja ir iegūta robaina histogramma – novērojumu skaits pa klasēm te pieaug, te samazinās, ir vairāki tukši (bez neviena novērojuma) intervāli, tad tā ir pazīme, ka ir izveidots pārāk daudz klašu konkrētajam novērojumu skaitam.

3. nodaļa

1. Vidējās sesijas atzīmes aprēķināšanai lieto harmonisko vidējo.
2. Moda ir ranžētas (sakārtotas no mazākās uz lielāko vērtību) variācijas rindas vidū esošās variantes vērtība.
3. Aprēķinot vidējo investīciju fondu ienesīgumu, ir jālieto vidējā svērtā aritmētiskā formula.

4. Standartnovirze ir standartos (normatīvos) noteiktā novirze.
5. Ja $s_1^2 > s_2^2$, tad pirmā sadalījuma histogramma būs nobīdīta vairāk pa labi, salīdzinot ar 2. sadalījuma histogrammu.
6. Ja asimetrijas koeficienta vērtība ir nulle, tad visas pazīmes vērtības ir pārstāvētas vienādi.
7. Ja ekscesa koeficienta vērtība ir pozitīva, tad sadalījums ir smailāks nekā normālais sadalījums.
8. Aprakstot iedzīvotāju ienākumus, ir šāda informācija: 20 % respondentu iztiek ar mazāk nekā 3 € dienā – 3 € dienā ir 2. decile.
9. Variācijas amplitūda ir starpība starp lielāko un mazāko reģistrēto vērtību.
10. Aprēķinot aritmētisko vidējo intervālu variācijas rindai pēc svērtās formulas, tiek iegūta aptuvena vērtība.

4. nodaļa

1. Produkcijas kvalitātes pārbaude speciālā testa aparatūrā ir izmēģinājums.
2. Varbūtība izvilkt „kungu" no 52 kāršu kavas ir 4/52. Tāds aprēķins atbilst statistiskajai varbūtībai.
3. Ja monēta, pametot to gaisā 4 reizes, visas 4 reizes ir nokritusi ar aversu uz augšu, tad metējs noteikti šmaucas, jo pareizi būtu, ja tā nokristu divas reizes ar aversu un divas reizes ar reversu uz augšu.
4. Pilna notikumu kopa ir visi iespējamie notikumi, kas var notikt izmēģinājuma vai novērojuma rezultātā.
5. Matemātiskā cerība ir teorētiskā sadalījuma aritmētiskā vidējā vērtība.
6. Nepārtrauktu sadalījumu (praktiski neierobežotu iznākumu skaitu) aizvieto ar diskrētu sadalījumu, lai vienkāršotu prognozēšanu praktiskā uzņēmuma vadīšanā.
7. Normālais sadalījums ir bieži sastopams dabas un sociāli ekonomiskajos pētījumos.
8. Stjūdenta sadalījums ir sadalījums, ko iegūst students, pildot praktisko uzdevumu, ja ir novērotas visas kopas vienības (ģenerālkopa).
9. Reprezentācijas kļūda rodas nepilnajā statistiskajā novērošanā, un tādas nav, ja ir apsekotas visas ģenerālkopas vienības.
10. Ligzdveida izlasi veido, sadalot kopu iekšēji viendabīgās apakškopās pēc pētījumam būtiskas pazīmes un apsekojot dažas izdalītās apakškopas pilnībā.

5. nodaļa

1. Nulles hipotēze nozīmē, ka starp pārbaudāmajām kopām nav statistiski nozīmīgas atšķirības.
2. Hipotēzes pārbaudes nozīmības līmenis (α) nozīmē izdarītā secinājuma pārliecības (ticamības) varbūtību.
3. Ar Stjūdenta testu salīdzina divu detaļu apstrāžu metožu ātrumu. $\bar{x}_1$=30 det./st., $\bar{x}_2$=35 det./st., t_{stat}=1,89, $t_{critical\ one\ tail}$=1,69, α=0,05. Ar 95 % ticamības līmeni var apgalvot, ka 2. metode ļauj saražot vairāk detaļu vienā stundā.
4. Ar Fišera kritēriju iegūti šādi rezultāti: s_1^2 = 2180; s_2^2 = 3360, F = 1,54; α = 0,05, $F_{critical\ one\ tail}$ = 1,84. Otrajai kopai dati izkliedēti plašākā diapazonā nekā pirmajai, bet varbūtība, ka atšķirība ir nejauša, ir lielāka par 5 %.
5. Divpusējās nulles hipotēzi lieto, ja pārbauda ražotās produkcijas atbilstību standartam.
6. χ^2 kritērijs ir parametriska metode, pārbauda divu kopu aritmētisko vidējo statistiski nozīmīgu atšķirību.
7. Tiek pārbaudīta kvalifikācijas kursu efektivitāte. 10 strādnieki tiek testēti uz darba ražīgumu pirms un pēc kvalifikācijas kursiem, pareizais tests būtu „t-Test Paired Two Sample for Means".

8. Ja aprēķinātā (empīriskā) kritērija vērtība ir lielāka par kritisko (tabulās atrasto), tad nulles hipotēze ir patiesa, atšķirība starp kopām nav statistiski nozīmīga.
9. Fišera kritēriju var lietot, ja salīdzina kopas, kuru vērtības ir mērītas nominālajā skalā.
10. $\bar{x}_1$= 1,39, $\bar{x}_2$= 2,15, t_{emp} = 2,18, $t_{\alpha=0,05;\nu=8}$ = 2,31. Nulles hipotēzi nevar noraidīt, bet novērojumu skaits ir pārāk mazs, un ļoti iespējams, ka, paplašinot pētījumu (izveidojot lielāku izlasi), var pierādīt kopu atšķirību.

6. nodaļa

1. Dispersijas analīze parāda, vai grupēšanai izvēlētā pazīme būtiski izskaidro datu variāciju.
2. Starpgrupu dispersija parāda izskaidroto datu izkliedes daļu.
3. Brīvības pakāpes izskaidrotajai dispersijai aprēķina šādi – novērojumu skaits vienā gradācijas klasē mīnus viens.
4.

Izkliede	Noviržu kvadrātu summa	ν	S^2	F	F_{crit}
Faktora	1800	2	900	4,22	3,68
Atlikuma	3200	15	213		
Kopējā	5000	17			

Pētāmā faktora ietekme nav statistiski nozīmīga, jo ar grupējumu izskaidrota tikai aptuveni1/3 daļa no kopējās datu izkliedes.

5. Dispersijas analīzē var izmantot intervālu variācijas rindas grupējumu.
6. Faktors *A* ir darbu izpildes secība, faktors *B* – strādnieki. Iegūti šādi rezultāti:

	F_{emp}	F_{crit}
A	1,03	3,15
B	1,95	2,04
Mijiedarbība *AB*	3,37	1,80

Nav svarīga darbu izpildes secība, arī starp strādnieku ražīgumu nav būtiskas atšķirības, bet svarīgi ir, kādu metodi strādnieks ir apguvis.

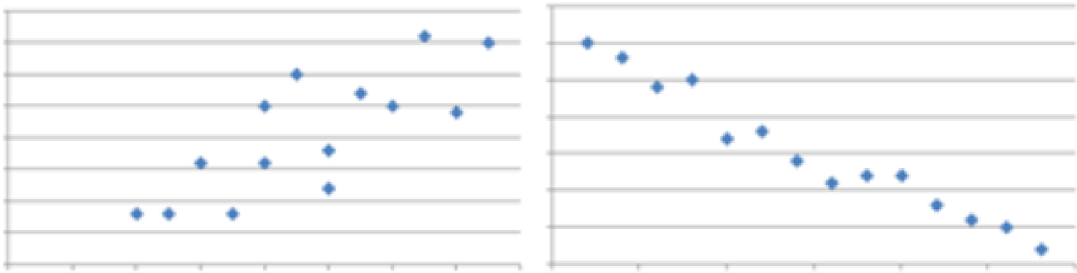
7. Ir izvēlēts hipotēzes pārbaudes nozīmības līmenis α = 0,1, tas nozīmē, ka, noraidot nulles hipotēzi, tomēr pastāv kļūdaina secinājuma varbūtība līdz 0,1.
8. Neizskaidrotā noviržu kvadrātu summa raksturo datu izkliedi ap izdalīto grupu vidējiem.
9. Fišera kritērijs ir grupu aritmētisko vidējo attiecība.
10. Dispersijas analīze pētāmās pazīmes vērtības reģistrē nominālajā skalā.

7. nodaļa

1. Korelācijas analīze izskaidro, cik cieši vienas pazīmes izmaiņas ietekmē otras pazīmes izmaiņas.
2. Regresijas analīze pēta ekonomikas recesiju.
3. Pīrsona korelācijas koeficientu vērtība „–0,86" norāda, ka starp pazīmēm nav sakarības, izmaiņas tuvu haotiskām.
4. Rēķinot korelācijas koeficientu manuāli, iegūts rezultāts „2,3", tas nozīmē, ka aprēķini ir kļūdaini.
5. Ja korelācijas koeficients ir aprēķināts no dažiem novērojumiem, tad pat augsta korelācijas koeficienta vērtība var nenozīmēt pierādītu sakarību.
6. Pētot uzņēmuma aktīvu (*y*) un apgrozījumu (*x*) sakarību, ir iegūts šāds vienādojums: $y = 200 + 0,3\,x$, tas nozīmē, ka vidēji uz katru apgrozījuma eiro pieaugumu aktīvu vērtība paaug par 30 centiem.

7. Galvenais kritērijs piemērotākā regresijas vienādojuma izvēlē ir determinācijas koeficients – jo tas ir augstāks, jo labāks vienādojums, iegūtā regresijas modeļa ekonomiski loģiskais skaidrojums nav svarīgs.
8. Ar regresijas rīku veiktajos aprēķinos ir iegūtas šādas *b* koeficienta 95 % ticamības intervāla robežas: zemākā – „–0,0062”, augstākā – „–0,0148”, tas nozīmē, ka nulles hipotēzi noraidīt nevar, jo, koeficientu noapaļojot, sanāk nulle.

9. un 10. Novērtējiet pēc korelācijas diagrammām.



r_1 r_2

9. $|r_1| < |r_2|$
10. $r_1 > r_2$

Atbildes uz pašpārbaudes jautājumiem

1. nodaļa

1. Nē. 2. Jā. 3. Jā. 4. Nē. 5. Jā. 6. Jā. 7. Nē. 8. Jā. 9. Nē. 10. Jā.

2. nodaļa

1. Nē. 2. Jā. 3. Nē. 4. Jā. 5. Nē. 6. Nē. 7. Jā. 8. Jā. 9. Nē. 10. Jā.

3. nodaļa

1. Nē. 2. Nē. 3. Jā. 4. Nē. 5. Nē. 6. Nē. 7. Jā. 8. Jā. 9. Jā. 10. Jā.

4. nodaļa

1. Jā. 2. Nē. 3. Nē. 4. Jā. 5. Jā. 6. Jā. 7. Jā. 8. Nē. 9. Jā. 10. Nē

5. nodaļa

1. Jā. 2. Nē. 3. Jā. 4. Jā. 5. Jā. 6. Nē. 7. Jā. 8. Nē. 9. Nē. 10. Jā.

6. nodaļa

1. Jā. 2. Jā. 3. Jā. 4. Nē. 5. Nē. 6. Jā. 7. Jā. 8. Jā. 9. Nē. 10. Nē.

7. nodaļa

1. Jā. 2. Nē. 3. Nē. 4. Jā. 5. Jā. 6. Jā. 7. Nē. 8. Nē. 9. Nē. 10. Jā.

Standartizēta normālā sadalījuma integrālā funkcija $F(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{\frac{z^2}{2}}\,dz$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,0	0,50000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,47210	0,46812	0,46414
-0,1	0,46017	0,45620	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
-0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
-0,3	0,38209	0,37828	0,37448	0,37070	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
-0,4	0,34458	0,34090	0,33724	0,33360	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
-0,5	0,30854	0,30503	0,30153	0,29806	0,29460	0,29116	0,28774	0,28434	0,28096	0,27760
-0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,24510
-0,7	0,24196	0,23885	0,23576	0,23270	0,22965	0,22663	0,22363	0,22065	0,21770	0,21476
-0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
-0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
-1,0	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
-1,1	0,13567	0,13350	0,13136	0,12924	0,12714	0,12507	0,12302	0,12100	0,11900	0,11702
-1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
-1,3	0,09680	0,09510	0,09342	0,09176	0,09012	0,08851	0,08691	0,08534	0,08379	0,08226
-1,4	0,08076	0,07927	0,07780	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
-1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
-1,6	0,05480	0,05370	0,05262	0,05155	0,05050	0,04947	0,04846	0,04746	0,04648	0,04551
-1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,03920	0,03836	0,03754	0,03673
-1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
-1,9	0,02872	0,02807	0,02743	0,02680	0,02619	0,02559	0,02500	0,02442	0,02385	0,02330
-2,0	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,01970	0,01923	0,01876	0,01831
-2,1	0,01786	0,01743	0,01700	0,01659	0,01618	0,01578	0,01539	0,01500	0,01463	0,01426
-2,2	0,01390	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,01160	0,01130	0,01101
-2,3	0,01072	0,01044	0,01017	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
-2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
-2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
-2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
-2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
-2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
-2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
-3,0	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,00100
-3,1	0,00097	0,00094	0,00090	0,00087	0,00084	0,00082	0,00079	0,00076	0,00074	0,00071
-3,2	0,00069	0,00066	0,00064	0,00062	0,00060	0,00058	0,00056	0,00054	0,00052	0,00050
-3,3	0,00048	0,00047	0,00045	0,00043	0,00042	0,00040	0,00039	0,00038	0,00036	0,00035
-3,4	0,00034	0,00032	0,00031	0,00030	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024
-3,5	0,00023	0,00022	0,00022	0,00021	0,00020	0,00019	0,00019	0,00018	0,00017	0,00017
-3,6	0,00016	0,00015	0,00015	0,00014	0,00014	0,00013	0,00013	0,00012	0,00012	0,00011
-3,7	0,00011	0,00010	0,00010	0,00010	0,00009	0,00009	0,00008	0,00008	0,00008	0,00008
-3,8	0,00007	0,00007	0,00007	0,00006	0,00006	0,00006	0,00006	0,00005	0,00005	0,00005
-3,9	0,00005	0,00005	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004	0,00003	0,00003

Standartizēta normālā sadalījuma integrālā funkcija

(turpinājums pozitīvām z vērtībām)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

Stjūdenta sadalījuma kritiskās vērtības

ν\α	0,1	0,05	0,025	0,01	0,001
1	6,314	12,706	25,452	63,657	636,619
2	2,920	4,303	6,205	9,925	31,599
3	2,353	3,182	4,177	5,841	12,924
4	2,132	2,776	3,495	4,604	8,610
5	2,015	2,571	3,163	4,032	6,869
6	1,943	2,447	2,969	3,707	5,959
7	1,895	2,365	2,841	3,499	5,408
8	1,860	2,306	2,752	3,355	5,041
9	1,833	2,262	2,685	3,250	4,781
10	1,812	2,228	2,634	3,169	4,587
11	1,796	2,201	2,593	3,106	4,437
12	1,782	2,179	2,560	3,055	4,318
13	1,771	2,160	2,533	3,012	4,221
14	1,761	2,145	2,510	2,977	4,140
15	1,753	2,131	2,490	2,947	4,073
16	1,746	2,120	2,473	2,921	4,015
17	1,740	2,110	2,458	2,898	3,965
18	1,734	2,101	2,445	2,878	3,922
19	1,729	2,093	2,433	2,861	3,883
20	1,725	2,086	2,423	2,845	3,850
21	1,721	2,080	2,414	2,831	3,819
22	1,717	2,074	2,405	2,819	3,792
23	1,714	2,069	2,398	2,807	3,768
24	1,711	2,064	2,391	2,797	3,745
25	1,708	2,060	2,385	2,787	3,725
26	1,706	2,056	2,379	2,779	3,707
27	1,703	2,052	2,373	2,771	3,690
28	1,701	2,048	2,368	2,763	3,674
29	1,699	2,045	2,364	2,756	3,659
30	1,697	2,042	2,360	2,750	3,646
40	1,684	2,021	2,329	2,704	3,551
50	1,676	2,009	2,311	2,678	3,496
60	1,671	2,000	2,299	2,660	3,460
70	1,667	1,994	2,291	2,648	3,435
80	1,664	1,990	2,284	2,639	3,416
90	1,662	1,987	2,280	2,632	3,402
100	1,660	1,984	2,276	2,626	3,390
110	1,659	1,982	2,272	2,621	3,381
120	1,658	1,980	2,270	2,617	3,373
∞	1,645	1,960	2,241	2,576	3,291

Fišera sadalījuma kritiskās vērtības, α=0,05

$\nu_2\backslash\nu_1$	1	2	3	4	5	6	7	8	9	10
1	161,448	199,500	215,707	224,583	230,162	233,986	236,768	238,883	240,543	241,882
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993
70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	2,017	1,969
80	3,960	3,111	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951
90	3,947	3,098	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,975	1,927
110	3,927	3,079	2,687	2,454	2,297	2,182	2,094	2,024	1,966	1,918
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910

$\nu_2 \backslash \nu_1$	11	12	13	14	15	16	17	18	19	20
1	242,983	243,906	244,690	245,364	245,950	246,464	246,918	247,323	247,686	248,013
2	19,405	19,413	19,419	19,424	19,429	19,433	19,437	19,440	19,443	19,446
3	8,763	8,745	8,729	8,715	8,703	8,692	8,683	8,675	8,667	8,660
4	5,936	5,912	5,891	5,873	5,858	5,844	5,832	5,821	5,811	5,803
5	4,704	4,678	4,655	4,636	4,619	4,604	4,590	4,579	4,568	4,558
6	4,027	4,000	3,976	3,956	3,938	3,922	3,908	3,896	3,884	3,874
7	3,603	3,575	3,550	3,529	3,511	3,494	3,480	3,467	3,455	3,445
8	3,313	3,284	3,259	3,237	3,218	3,202	3,187	3,173	3,161	3,150
9	3,102	3,073	3,048	3,025	3,006	2,989	2,974	2,960	2,948	2,936
10	2,943	2,913	2,887	2,865	2,845	2,828	2,812	2,798	2,785	2,774
11	2,818	2,788	2,761	2,739	2,719	2,701	2,685	2,671	2,658	2,646
12	2,717	2,687	2,660	2,637	2,617	2,599	2,583	2,568	2,555	2,544
13	2,635	2,604	2,577	2,554	2,533	2,515	2,499	2,484	2,471	2,459
14	2,565	2,534	2,507	2,484	2,463	2,445	2,428	2,413	2,400	2,388
15	2,507	2,475	2,448	2,424	2,403	2,385	2,368	2,353	2,340	2,328
16	2,456	2,425	2,397	2,373	2,352	2,333	2,317	2,302	2,288	2,276
17	2,413	2,381	2,353	2,329	2,308	2,289	2,272	2,257	2,243	2,230
18	2,374	2,342	2,314	2,290	2,269	2,250	2,233	2,217	2,203	2,191
19	2,340	2,308	2,280	2,256	2,234	2,215	2,198	2,182	2,168	2,155
20	2,310	2,278	2,250	2,225	2,203	2,184	2,167	2,151	2,137	2,124
21	2,283	2,250	2,222	2,197	2,176	2,156	2,139	2,123	2,109	2,096
22	2,259	2,226	2,198	2,173	2,151	2,131	2,114	2,098	2,084	2,071
23	2,236	2,204	2,175	2,150	2,128	2,109	2,091	2,075	2,061	2,048
24	2,216	2,183	2,155	2,130	2,108	2,088	2,070	2,054	2,040	2,027
25	2,198	2,165	2,136	2,111	2,089	2,069	2,051	2,035	2,021	2,007
26	2,181	2,148	2,119	2,094	2,072	2,052	2,034	2,018	2,003	1,990
27	2,166	2,132	2,103	2,078	2,056	2,036	2,018	2,002	1,987	1,974
28	2,151	2,118	2,089	2,064	2,041	2,021	2,003	1,987	1,972	1,959
29	2,138	2,104	2,075	2,050	2,027	2,007	1,989	1,973	1,958	1,945
30	2,126	2,092	2,063	2,037	2,015	1,995	1,976	1,960	1,945	1,932
40	2,038	2,003	1,974	1,948	1,924	1,904	1,885	1,868	1,853	1,839
50	1,986	1,952	1,921	1,895	1,871	1,850	1,831	1,814	1,798	1,784
60	1,952	1,917	1,887	1,860	1,836	1,815	1,796	1,778	1,763	1,748
70	1,928	1,893	1,863	1,836	1,812	1,790	1,771	1,753	1,737	1,722
80	1,910	1,875	1,845	1,817	1,793	1,772	1,752	1,734	1,718	1,703
90	1,897	1,861	1,830	1,803	1,779	1,757	1,737	1,720	1,703	1,688
100	1,886	1,850	1,819	1,792	1,768	1,746	1,726	1,708	1,691	1,676
110	1,877	1,841	1,810	1,783	1,758	1,736	1,716	1,698	1,682	1,667
120	1,869	1,834	1,803	1,775	1,750	1,728	1,709	1,690	1,674	1,659

$\nu_2 \backslash \nu_1$	21	22	23	24	25	26	27	28	29	30
1	248,309	248,579	248,826	249,052	249,260	249,453	249,631	249,797	249,951	250,095
2	19,448	19,450	19,452	19,454	19,456	19,457	19,459	19,460	19,461	19,462
3	8,654	8,648	8,643	8,639	8,634	8,630	8,626	8,623	8,620	8,617
4	5,795	5,787	5,781	5,774	5,769	5,763	5,759	5,754	5,750	5,746
5	4,549	4,541	4,534	4,527	4,521	4,515	4,510	4,505	4,500	4,496
6	3,865	3,856	3,849	3,841	3,835	3,829	3,823	3,818	3,813	3,808
7	3,435	3,426	3,418	3,410	3,404	3,397	3,391	3,386	3,381	3,376
8	3,140	3,131	3,123	3,115	3,108	3,102	3,095	3,090	3,084	3,079
9	2,926	2,917	2,908	2,900	2,893	2,886	2,880	2,874	2,869	2,864
10	2,764	2,754	2,745	2,737	2,730	2,723	2,716	2,710	2,705	2,700
11	2,636	2,626	2,617	2,609	2,601	2,594	2,588	2,582	2,576	2,570
12	2,533	2,523	2,514	2,505	2,498	2,491	2,484	2,478	2,472	2,466
13	2,448	2,438	2,429	2,420	2,412	2,405	2,398	2,392	2,386	2,380
14	2,377	2,367	2,357	2,349	2,341	2,333	2,326	2,320	2,314	2,308
15	2,316	2,306	2,297	2,288	2,280	2,272	2,265	2,259	2,253	2,247
16	2,264	2,254	2,244	2,235	2,227	2,220	2,212	2,206	2,200	2,194
17	2,219	2,208	2,199	2,190	2,181	2,174	2,167	2,160	2,154	2,148
18	2,179	2,168	2,159	2,150	2,141	2,134	2,126	2,119	2,113	2,107
19	2,144	2,133	2,123	2,114	2,106	2,098	2,090	2,084	2,077	2,071
20	2,112	2,102	2,092	2,082	2,074	2,066	2,059	2,052	2,045	2,039
21	2,084	2,073	2,063	2,054	2,045	2,037	2,030	2,023	2,016	2,010
22	2,059	2,048	2,038	2,028	2,020	2,012	2,004	1,997	1,990	1,984
23	2,036	2,025	2,014	2,005	1,996	1,988	1,981	1,973	1,967	1,961
24	2,015	2,003	1,993	1,984	1,975	1,967	1,959	1,952	1,945	1,939
25	1,995	1,984	1,974	1,964	1,955	1,947	1,939	1,932	1,926	1,919
26	1,978	1,966	1,956	1,946	1,938	1,929	1,921	1,914	1,907	1,901
27	1,961	1,950	1,940	1,930	1,921	1,913	1,905	1,898	1,891	1,884
28	1,946	1,935	1,924	1,915	1,906	1,897	1,889	1,882	1,875	1,869
29	1,932	1,921	1,910	1,901	1,891	1,883	1,875	1,868	1,861	1,854
30	1,919	1,908	1,897	1,887	1,878	1,870	1,862	1,854	1,847	1,841
40	1,826	1,814	1,803	1,793	1,783	1,775	1,766	1,759	1,751	1,744
50	1,771	1,759	1,748	1,737	1,727	1,718	1,710	1,702	1,694	1,687
60	1,735	1,722	1,711	1,700	1,690	1,681	1,672	1,664	1,656	1,649
70	1,709	1,696	1,685	1,674	1,664	1,654	1,646	1,637	1,629	1,622
80	1,689	1,677	1,665	1,654	1,644	1,634	1,626	1,617	1,609	1,602
90	1,675	1,662	1,650	1,639	1,629	1,619	1,610	1,601	1,593	1,586
100	1,663	1,650	1,638	1,627	1,616	1,607	1,598	1,589	1,581	1,573
110	1,653	1,640	1,628	1,617	1,606	1,597	1,587	1,579	1,571	1,563
120	1,645	1,632	1,620	1,608	1,598	1,588	1,579	1,570	1,562	1,554

$\nu_2\backslash\nu_1$	40	50	60	70	80	90	100	110	120
1	251,143	251,774	252,196	252,497	252,724	252,900	253,041	253,157	253,253
2	19,471	19,476	19,479	19,481	19,483	19,485	19,486	19,487	19,487
3	8,594	8,581	8,572	8,566	8,561	8,557	8,554	8,551	8,549
4	5,717	5,699	5,688	5,679	5,673	5,668	5,664	5,661	5,658
5	4,464	4,444	4,431	4,422	4,415	4,409	4,405	4,401	4,398
6	3,774	3,754	3,740	3,730	3,722	3,716	3,712	3,708	3,705
7	3,340	3,319	3,304	3,294	3,286	3,280	3,275	3,271	3,267
8	3,043	3,020	3,005	2,994	2,986	2,980	2,975	2,970	2,967
9	2,826	2,803	2,787	2,776	2,768	2,761	2,756	2,751	2,748
10	2,661	2,637	2,621	2,610	2,601	2,594	2,588	2,584	2,580
11	2,531	2,507	2,490	2,478	2,469	2,462	2,457	2,452	2,448
12	2,426	2,401	2,384	2,372	2,363	2,356	2,350	2,345	2,341
13	2,339	2,314	2,297	2,284	2,275	2,267	2,261	2,257	2,252
14	2,266	2,241	2,223	2,210	2,201	2,193	2,187	2,182	2,178
15	2,204	2,178	2,160	2,147	2,137	2,130	2,123	2,118	2,114
16	2,151	2,124	2,106	2,093	2,083	2,075	2,068	2,063	2,059
17	2,104	2,077	2,058	2,045	2,035	2,027	2,020	2,015	2,011
18	2,063	2,035	2,017	2,003	1,993	1,985	1,978	1,973	1,968
19	2,026	1,999	1,980	1,966	1,955	1,947	1,940	1,935	1,930
20	1,994	1,966	1,946	1,932	1,922	1,913	1,907	1,901	1,896
21	1,965	1,936	1,916	1,902	1,891	1,883	1,876	1,870	1,866
22	1,938	1,909	1,889	1,875	1,864	1,856	1,849	1,843	1,838
23	1,914	1,885	1,865	1,850	1,839	1,830	1,823	1,818	1,813
24	1,892	1,863	1,842	1,828	1,816	1,808	1,800	1,795	1,790
25	1,872	1,842	1,822	1,807	1,796	1,787	1,779	1,773	1,768
26	1,853	1,823	1,803	1,788	1,776	1,767	1,760	1,754	1,749
27	1,836	1,806	1,785	1,770	1,758	1,749	1,742	1,736	1,731
28	1,820	1,790	1,769	1,754	1,742	1,733	1,725	1,719	1,714
29	1,806	1,775	1,754	1,738	1,726	1,717	1,710	1,703	1,698
30	1,792	1,761	1,740	1,724	1,712	1,703	1,695	1,689	1,683
40	1,693	1,660	1,637	1,621	1,608	1,597	1,589	1,582	1,577
50	1,634	1,599	1,576	1,558	1,544	1,534	1,525	1,518	1,511
60	1,594	1,559	1,534	1,516	1,502	1,491	1,481	1,474	1,467
70	1,566	1,530	1,505	1,486	1,471	1,459	1,450	1,442	1,435
80	1,545	1,508	1,482	1,463	1,448	1,436	1,426	1,418	1,411
90	1,528	1,491	1,465	1,445	1,429	1,417	1,407	1,399	1,391
100	1,515	1,477	1,450	1,430	1,415	1,402	1,392	1,383	1,376
110	1,504	1,466	1,439	1,418	1,402	1,390	1,379	1,370	1,363
120	1,495	1,457	1,429	1,408	1,392	1,379	1,369	1,360	1,352

Fišera sadalījuma kritiskās vērtības, α = 0,1

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895	1,866
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819
40	2,835	2,440	2,226	2,091	1,997	1,927	1,873	1,829	1,793	1,763
50	2,809	2,412	2,197	2,061	1,966	1,895	1,840	1,796	1,760	1,729
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707
80	2,769	2,370	2,154	2,016	1,921	1,849	1,793	1,748	1,711	1,680
100	2,756	2,356	2,139	2,002	1,906	1,834	1,778	1,732	1,695	1,663
120	2,748	2,347	2,130	1,992	1,896	1,824	1,767	1,722	1,684	1,652

$\nu_2 \backslash \nu_1$	15	20	25	30	40	50	60	80	100	120
1	61,220	61,740	62,055	62,265	62,529	62,688	62,794	62,927	63,007	63,061
2	9,425	9,441	9,451	9,458	9,466	9,471	9,475	9,479	9,481	9,483
3	5,200	5,184	5,175	5,168	5,160	5,155	5,151	5,147	5,144	5,143
4	3,870	3,844	3,828	3,817	3,804	3,795	3,790	3,782	3,778	3,775
5	3,238	3,207	3,187	3,174	3,157	3,147	3,140	3,132	3,126	3,123
6	2,871	2,836	2,815	2,800	2,781	2,770	2,762	2,752	2,746	2,742
7	2,632	2,595	2,571	2,555	2,535	2,523	2,514	2,504	2,497	2,493
8	2,464	2,425	2,400	2,383	2,361	2,348	2,339	2,328	2,321	2,316
9	2,340	2,298	2,272	2,255	2,232	2,218	2,208	2,196	2,189	2,184
10	2,244	2,201	2,174	2,155	2,132	2,117	2,107	2,095	2,087	2,082
15	1,972	1,924	1,894	1,873	1,845	1,828	1,817	1,802	1,793	1,787
20	1,845	1,794	1,761	1,738	1,708	1,690	1,677	1,660	1,650	1,643
25	1,771	1,718	1,683	1,659	1,627	1,607	1,593	1,576	1,565	1,557
30	1,722	1,667	1,632	1,606	1,573	1,552	1,538	1,519	1,507	1,499
40	1,662	1,605	1,568	1,541	1,506	1,483	1,467	1,447	1,434	1,425
50	1,627	1,568	1,529	1,502	1,465	1,441	1,424	1,402	1,388	1,379
60	1,603	1,543	1,504	1,476	1,437	1,413	1,395	1,372	1,358	1,348
70	1,587	1,526	1,486	1,457	1,418	1,392	1,374	1,350	1,335	1,325
80	1,574	1,513	1,472	1,443	1,403	1,377	1,358	1,334	1,318	1,307
90	1,564	1,503	1,461	1,432	1,391	1,365	1,346	1,321	1,304	1,293
100	1,557	1,494	1,453	1,423	1,382	1,355	1,336	1,310	1,293	1,282
110	1,550	1,488	1,446	1,415	1,374	1,347	1,327	1,301	1,284	1,272
120	1,545	1,482	1,440	1,409	1,368	1,340	1,320	1,294	1,277	1,265

Fišera sadalījuma kritiskās vērtības, α = 0,01

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
1	4052,181	4999,500	5403,352	5624,583	5763,650	5858,986	5928,356	5981,070	6022,473	6055,847
2	98,503	99,000	99,166	99,249	99,299	99,333	99,356	99,374	99,388	99,399
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158	10,051
6	13,745	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874
7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620
8	11,259	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814
9	10,561	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257
10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129
30	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,067	2,979
40	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888	2,801
50	7,171	5,057	4,199	3,720	3,408	3,186	3,020	2,890	2,785	2,698
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632
80	6,963	4,881	4,036	3,563	3,255	3,036	2,871	2,742	2,637	2,551
100	6,895	4,824	3,984	3,513	3,206	2,988	2,823	2,694	2,590	2,503
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559	2,472
	15	20	25	30	40	50	60	80	100	120
1	6157,285	6208,730	6239,825	6260,649	6286,782	6302,517	6313,030	6326,197	6334,110	6339,391
2	99,433	99,449	99,459	99,466	99,474	99,479	99,482	99,487	99,489	99,491
3	26,872	26,690	26,579	26,505	26,411	26,354	26,316	26,269	26,240	26,221
4	14,198	14,020	13,911	13,838	13,745	13,690	13,652	13,605	13,577	13,558
5	9,722	9,553	9,449	9,379	9,291	9,238	9,202	9,157	9,130	9,112
6	7,559	7,396	7,296	7,229	7,143	7,091	7,057	7,013	6,987	6,969
7	6,314	6,155	6,058	5,992	5,908	5,858	5,824	5,781	5,755	5,737
8	5,515	5,359	5,263	5,198	5,116	5,065	5,032	4,989	4,963	4,946
9	4,962	4,808	4,713	4,649	4,567	4,517	4,483	4,441	4,415	4,398
10	4,558	4,405	4,311	4,247	4,165	4,115	4,082	4,039	4,014	3,996
15	3,522	3,372	3,278	3,214	3,132	3,081	3,047	3,004	2,977	2,959
20	3,088	2,938	2,843	2,778	2,695	2,643	2,608	2,563	2,535	2,517
25	2,850	2,699	2,604	2,538	2,453	2,400	2,364	2,317	2,289	2,270
30	2,700	2,549	2,453	2,386	2,299	2,245	2,208	2,160	2,131	2,111
40	2,522	2,369	2,271	2,203	2,114	2,058	2,019	1,969	1,938	1,917
50	2,419	2,265	2,167	2,098	2,007	1,949	1,909	1,857	1,825	1,803
60	2,352	2,198	2,098	2,028	1,936	1,877	1,836	1,783	1,749	1,726
70	2,306	2,150	2,050	1,980	1,886	1,826	1,785	1,730	1,695	1,672
80	2,271	2,115	2,015	1,944	1,849	1,788	1,746	1,690	1,655	1,630
90	2,244	2,088	1,987	1,916	1,820	1,759	1,716	1,659	1,623	1,598
100	2,223	2,067	1,965	1,893	1,797	1,735	1,692	1,634	1,598	1,572
110	2,206	2,049	1,947	1,875	1,778	1,716	1,672	1,614	1,577	1,551
120	2,192	2,035	1,932	1,860	1,763	1,700	1,656	1,597	1,559	1,533

χ^2kritērija kritiskās vērtības

ν\α	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01
1	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635
2	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210
3	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345
4	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277
5	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086
6	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812
7	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475
8	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090
9	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666
10	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209
11	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725
12	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217
13	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688
14	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141
15	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578
16	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000
17	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409
18	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805
19	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191
20	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566
21	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932
22	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289
23	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638
24	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980
25	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314
26	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642
27	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963
28	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278
29	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588
30	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892
35	18,509	20,569	22,465	24,797	46,059	49,802	53,203	57,342
40	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691
45	25,901	28,366	30,612	33,350	57,505	61,656	65,410	69,957
50	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154
55	33,570	36,398	38,958	42,060	68,796	73,311	77,380	82,292
60	37,485	40,482	43,188	46,459	74,397	79,082	83,298	88,379
65	41,444	44,603	47,450	50,883	79,973	84,821	89,177	94,422
70	45,442	48,758	51,739	55,329	85,527	90,531	95,023	100,425
75	49,475	52,942	56,054	59,795	91,061	96,217	100,839	106,393
80	53,540	57,153	60,391	64,278	96,578	101,879	106,629	112,329
85	57,634	61,389	64,749	68,777	102,079	107,522	112,393	118,236
90	61,754	65,647	69,126	73,291	107,565	113,145	118,136	124,116
95	65,898	69,925	73,520	77,818	113,038	118,752	123,858	129,973
100	70,065	74,222	77,929	82,358	118,498	124,342	129,561	135,807
105	74,252	78,536	82,354	86,909	123,947	129,918	135,247	141,620
110	78,458	82,867	86,792	91,471	129,385	135,480	140,917	147,414
115	82,682	87,213	91,242	96,043	134,813	141,030	146,571	153,191
120	86,923	91,573	95,705	100,624	140,233	146,567	152,211	158,950

Fišera transformācija

$$\varphi = 2 * \frac{\pi}{180} \arcsin \sqrt{p}$$

p	Argumenta *p* pēdējais cipars									
	0	1	2	3	4	5	6	7	8	9
0,000		0,0200	0,0283	0,0346	0,0400	0,0448	0,0490	0,0529	0,0566	0,0600
0,00		0,0632	0,0895	0,1096	0,1266	0,1415	0,1551	0,1676	0,1791	0,1900
0,0		0,2003	0,2838	0,3482	0,4027	0,4510	0,4949	0,5355	0,5735	0,6094
0,1	0,6435	0,6761	0,7075	0,7377	0,7670	0,7954	0,8230	0,8500	0,8763	0,9021
0,2	0,9273	0,9521	0,9764	1,0004	1,0240	1,0472	1,0701	1,0928	1,1152	1,1374
0,3	1,1593	1,1810	1,2025	1,2239	1,2451	1,2661	1,2870	1,3078	1,3284	1,3490
0,4	1,3694	1,3898	1,4101	1,4303	1,4505	1,4706	1,4907	1,5108	1,5308	1,5508
0,5	1,5708	1,5908	1,6108	1,6315	1,6509	1,6710	1,6911	1,7113	1,7315	1,7518
0,6	1,7722	1,7920	1,8132	1,8338	1,8546	1,8755	1,8965	1,9177	1,9391	1,9606
0,7	1,9823	2,0042	2,0264	2,0488	2,0715	2,0944	2,1177	2,1412	2,1652	2,1895
0,8	2,2143	2,2395	2,2653	2,2916	2,3186	2,3462	2,3746	2,4039	2,4341	2,4655
0,9	2,4981	2,5322	2,5681	2,6061	2,6467	2,6906	2,7389	2,7934	2,8578	2,9413
0,99	2,9413	2,9516	2,9625	2,9741	2,9865	3,0001	3,0150	3,1320	3,0521	3,0783
0,999	3,0783	3,0847	3,0910	3,0973	3,1036	3,1100	3,1163	3,1226	3,1289	3,1353
1	3,1416									

Izmantotās literatūras un avotu saraksts

1. Arhipova, I., Bāliņa, S. (2003). *Statistika ekonomikā. Risinājumi ar SPSS un Microsoft Excel: mācību līdzeklis.* Rīga: Datorzinību Centrs. 349 lpp.
2. Goša, Z. (2003). *Statistika.* Rīga: SIA" Izglītības soļi". 334 lpp.
3. Guļevska, D. (atb.red.). (1987). *Latviešu valodas vārdnīca: A-Ž.* Rīga: Avots. 883 lpp.
4. Krastiņš, O. (1998). *Statistika un ekonometrija : mācību grāmata augstskolām*. Rīga: LR Valsts statistikas komiteja. 436 lpp.
5. Krastiņš, O., Ciemiņa, I. (2003). *Statistika.* Rīga: LR Centrālā statistikas pārvalde. 267 lpp.
6. Lapin, Lawrence L. (1993). *Statistics for Modern Business Decisions* (6th ed.) [n.d.]: The Dryden Press.1265 p.
7. Lasmanis, A. (1999). *Datu ieguves, apstrādes un analīzes metodes pedagoģijas un psiholoģijas pētijumos*. Rīga: Mācību apgāds NT. 190 lpp.
8. Liepa, I. (1974). *Biometrija : mācību līdzeklis augstskolu studentiem.* Rīga: Zvaigzne. 336 lpp.
9. McConnell, C**.**, Brue, S. (1996). *Economics: principles, problems, and policies*. (13th ed.) New York: McGraw-Hill, Inc. 894 p.
10. Moore, D. (2003). *The basic practice of statistics*. (3d ed.) New York: W.H.Freeman and Company. 691 p.
11. Orlovska, A. (2012). *Statistika*. Rīga: RTU Izdevniecība, (RTU). 191 lpp.
12. Raščevska, M**.**, Kristapsone, S. (2000). *Statistika psiholoģijas pētījumos: eksperimentāla mācību grāmata psiholoģijas spec. studentiem*. Rīga: Izglītības soļi. 356 lpp.
13. Revina, I. (2002). *Ekonometrija: mācību līdzeklis* . I. Rīga: Latvijas Universitāte. 270 lpp.
14. Дэвид М. Левин [и др.]; [Пер. с англ. Д. А. Клюшина]. (2005). *Статистика для менеджеров с использованием Microsoft Excel.* (4-е изд.) М. [и др.]: Вильямс, (ГПП Печ. Двор). 1310 с.