J. Dravnieks

Matemātiskās statistikas metodes sporta zinātnē

Mācību grāmata
LSPA studentiem, maģistrantiem, doktorantiem

RĪGA - 2004

SATURS

IEVADS

Vārds "statistika" cēlies no latīņu "status" - stāvoklis. Tam ir vairākas nozīmes. Būtībā statistika ir ļoti sena zinātne, kurai daudz laika ir ziedojuši ne tikai zinātnieki, bet arī politiķi. Tādēļ saprotama statistikas definīciju daudzveidība. Lūk, oficiāli visizplatītākās:

- zinātnes nozare, kas pētī cilvēku sabiedrības, tautas saimniecības attīstības kvantitatīvās pārmaiņas un apstrādā šos datus zinātniskos un praktiskos nolūkos;
- kvantitatīva masu parādību uzskaite;
- matemātiskā statistika - matemātikas nozare, kas aplūko matemātikas metodes, kuras lieto statistikas datu sistematizēšanā, apstrādāšanā un izmantošanā zinātniskiem un praktiskiem secinājumiem.

Ar statistiku sastopamies arī sportā un fiziskajā audzināšanā. Visplašāk pazīstamā sporta statistika ir sacensību protokoli, dažādu rekordu saraksti, gada labāko sportistu saraksti u.c. Sporta mērījumu rezultātu apkopošanai un analīzei izmanto matemātiskās statistikas metodes. Pētot kādu sportista īpašību, uzkrājas liels skaitliskās informācijas daudzums. Cilvēks spēj vienlaicīgi orientēties ne vairāk kā 7 skaitļos. Tāpēc informācija jāapstrādā, padarot to koncentrētu un līdz ar to vieglāk uztveramu. Šo procedūru sauc par statistisko analīzi. Tās rezultāts ir daži skaitļi - statistiskie rādītāji, kuri raksturo vispārējas tendences.

1. PAMATJĒDZIENI

1.1. Matemātiskā statistika

Šai nodaļā aplūkosim galvenos jēdzienus un terminus, ar kuriem bieži sastapsimies un kurus turpmāk izmantosim. Statistika pēta masveida objektus un parādības. Mūsu gadījumā tie var būt sportisti, kas specializējas noteiktā veidā, kādas klases skolnieki, kas piedalās fiziskās audzināšanas stundā, cilvēki, kuri nodarbojas veselības grupā u.c. Tādējādi statistikas pētījuma objekts ir kopa, ko veido tās elementi (cilvēki). Lai iegūtu ziņas par kopu, kura mūs interesē, jānovēro tās elementi un jāreģistrē dati par svarīgākajām īpašībām, kas raksturo katru atsevišķu kopas elementu un līdz ar to visu kopu. Datu savākšanas procedūru sauc par statistisko novērošanu. Pētīšanai izraudzītās īpašības statistikas valodā sauc par pazīmēm. Pazīmju apzīmēšanai izmanto latīņu alfabēta lielos burtus *X, Y, Z*. No iepriekšējā materiāla mums jau zināms, ka pazīmes, kuras var mērīt tieši vai netieši, sauc par kvantitatīvām, bet pazīmes, kuras mērīšanai nepakļaujas - par kvalitatīvām pazīmēm. Pētāmo pazīmi mērot vai vērtējot (ekspertīzes ceļā) iegūtos - skaitļus sauc par variantēm. Variante ir atsevišķa mērījuma vai novērojuma rezultāts, kura vispārīgai apzīmēšanai formulās lieto pazīmes apzīmējumam (alfabēta lielajam burtam) atbilstošo mazo latīņu burtu pievienojot indeksu, kas apzīmē variantes numuru: $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$. Apsekojot vairākus objektus vai vairākkārt vienu un to pašu objektu, iegūto skaitļu - novērojumu rezultātu rindu sauc par nesakārtotu empīrisko rindu, bet šo skaitļu kopu - par paraugkopu. Novērojumos iegūto skaitļu rindu sakārto iegūstot variācijas rindu jeb empīrisko sadalījumu. Šo procedūru sauc par datu pirmapstrādi. Paraugkopas variantes pakļauj statistiskai analīzei, aprēķinot empīriskā sadalījuma parametrus - statistiskos rādītājus. Tagad paanalizēsim situāciju. Paraugkopa ir iegūta apsekojot kādu daļu no visiem iespējamiem vienveidīgiem pētījuma objektiem. Lai formulētu pilnīgi precīzus vispārinošus secinājumus par šo grupu, acīmredzot, būtu jāapseko visi iespējamie šāda tipa objekti. Tādā gadījumā iegūtās variantes veidotu ģenerālkopu. Sarežģījumi sākas ar to, ka lielākajā daļā gadījumu praktiski nav iespējams iegūt ģenerālkopu laika un līdzekļu trūkuma dēļ. Tādēļ izmanto izlases metodi - ievērojot randomizācijas principu apseko daļu no visiem objektiem, analizē iegūto paraugkopu un šīs analīzes rezultātus attiecina uz visiem objektiem ar iepriekš pieņemtu ticamības līmeni t.i. pieļaujot noteiktu kļūdas iespēju. Tādējādi ģenerālkopas vietā analīzei pakļauj tās daļu - paraugkopu, bet iegūtos rezultātus vispārina uz ģenerālkopu izmantojot varbūtību teoriju. Lai paraugkopa būtu reprezentatīva t.i. pietiekoši precīzi atspoguļotu ģenerālkopu, tās sastādīšanai izstrādātas speciālas izlases metodes. Paraugkopas apjomam t.i. varianšu (novērojumu) skaitam jābūt pietiekoši lielam. Ticamu secinājumu iegūšanai nepieciešamo paraugkopas apjomu nosaka ar speciālām metodēm. Jāatceras, ka paraugkopas apjomu t.i. varianšu skaitu tajā apzīmē ar n, un pēc apjoma paraugkopas iedala mazās paraugkopās ($n < 30$) un lielās paraugkopās ($n \leq 30$). Īpaša parādība, ar kuru sastopamies gan ikdienā gan visos statistikas uzdevumos, ir pazīmes variēšana - rezultātu mainīgums atkārtotos mērījumos vai novērojumos. Mērot pētāmo pazīmi vienveidīgu objektu grupā, piemēram, veseliem 5. klases zēniem, iegūstam atšķirīgus rezultātus. Tā ir variācija dotajā skolnieku grupā. Sastopama arī individuālā variācija, piemēram, sportists sacensībās izpildot vairākus mēģinājumus ik reizi sasniedz atšķirīgu rezultātu. Variē ne tikai dzīvu, bet arī nedzīvu (fizikālu) objektu mērīšanas rezultāti. Pazīmes variēšanas cēloņi ir neparedzētas vides izmaiņas un mērīšanas kļūdas, bet dzīviem objektiem arī tā saucamie iekšējie faktori - paša objekta mainīgums. Variē kā kvantitatīvas tā arī kvalitatīvas pazīmes. Kvantitatīvām pazīmēm raksturīga diskrēta vai nepārtraukta variēšana. Šie variēšanas veidi jāzin, lai prastu izvēlēties piemērotas analīzes metodes. Diskrētās variēšanas gadījumā starp divām pazīmes skaitliskām vērtībām ir galīgs, noteikts citu vērtību skaits. Praksē par diskrēti variējošām pazīmēm pieņemts uzskatīt tās, kuru variantes ir veseli skaitļi, piemēram, vingrinājuma izpildes reižu skaits, punktu skaits šaušanā u.c. Pazīmei variējot nepārtraukti, starp tās divām vērtībām teorētiski ir bezgalīgi daudz citu vērtību. Par nepārtraukti variējošām pieņemts uzskatīt tās pazīmes, kuru variantes ir skaitļi ar decimālldaļu. Sīkāk paanalizējot pazīmes variēšanas nepārtrauktības būtību kļūst saprotams, ka variēšanas nepārtrauktība ir relatīvs jēdziens un tieši atkarīga no mērinstrumenta precizitātes. Piemēram, reģistrējot distances veikšanas laiku ar rokas hronometru, var nolasīt rezultātu ar precizitāti līdz 0.1 s, bet ar elektronisko hronometru - līdz 0.01 s. Tas nozīmē, ka abos gadījumos intervalā no 10 s līdz 11 s var fiksēt galīgu vērtību skaitu: pirmajā gadījumā - 10 dažādus rezultātus, bet otrajā - 100, un pazīmes variēšanas "nepārtrauktība" palielinās pieaugot mērinstrumenta precizitātei. Kvalitatīvām pazīmēm raksturīga alternatīva vai nealternatīva variēšana.

Pazīmei variējot alternatīvi, sastopamies ar tās klātbūtni vai iztrūkumu un vērtējumiem "jā" vai "nē", "izpildīts" vai "neizpildīts", "ir" vai "nav". Skaitliski pazīmes klātbūtni vērtē ar "1", iztrūkumu - ar "0". Piemēram, basketbolā spēlētāja meistarības vērtēšanai skaitam iemestos un neiemestos metienus. Nealternatīvās variēšanas gadījumā pētamos objektus var iedalīt dažādās gradācijas grupās, pēc tā kādā veidā vai cik lielā mērā pētamā pazīme objektam piemīt.

1.2. Varbūtību teorija

Matemātiskās statistikas teorētiskā bāze ir varbūtību teorija - zinātne, kas pēta masveida gadījumnotikumu likumsakarības. Tā saistīta ar centieniem novērtēt dažādu parādību un notikumu iespējamību, paredzēt to turpmāko attīstību. Pirmo statistikas datu vākšana attiecas uz tālu senatni un saistīti galvenokārt ar iedzīvotāju uzskaiti. Normandijas hercogs Vilhelms I Iekarotājs (1027-1087.g.) pēc Anglijas iekarošanas 1066. g. lika sastādīt "Briesmīgās tiesas grāmatu" - dokumentu, kurā bija ziņas par karaļa, baznīcas un feodāļu īpašumiem, to lielumu, lopu skaitu un inventāru. Šo dokumentu izmantoja zemes nodokļu uzlikšanai. Sistemātiski un plaši statistiski pētījumi sākas kapitalistisko attiecību veidošanās laikā, attīstoties tirdzniecībai un naudas operācijām, tai skaitā saistītām ar apdrošināšanu. 14. gs. Itālijā un Niderlandē nodibinājās pirmās jūras pārvadājumu apdrošināšanas sabiedrības, kurām bija jārisina uzdevumi, kas saistīti ar risku finansu operācijās. Renesanses laikmetā strauji attīstījās dabaszinātnes, radās jautājumi par novērojumu rezultātu apstrādi - gadījumkļūdu novērtēšanu. Visi šie notikumi veicināja varbūtību teorijas attīstību. Lielākā daļa pirmatnējo varbūtības teorijas uzdevumu saistīti ar azarta spēlēm. Daudzi autori uzskata, ka pirmie darbi, kuros veidojās varbūtību teorijas pamatjēdzieni 16-17. gs. saistīti ar centieniem izveidot azarta spēļu teoriju. Patiesībā azarta spēles veicināja varbūtību teorijas attīstību kā piemērots modelis ar savu terminoloģiju, kas derēja daudzu parādību un uzdevumu aprakstam. Varbūtību teorijas pamatjēdzieni ir izmēģinājums un notikums. Izmēģinājums ir noteiktu apstākļu kompleksa radīšana. Notikums ir izmēģinājuma rezultāts. Piemēram, basketbolists metot bumbu groza virzienā piešķir tai noteikta virziena paātrinājumu. Šī izmēģinājuma rezultātā iespējami divi notikumi: 1) bumba trāpa grozā; 2) bumba netrāpa grozā. Abi notikumi saistīti ar zināmu nenoteiktību - nav garantijas, kurš no notikumiem noteikti iestāsies. Notikumus, kuri izmēģinājuma rezultātā var iestāties vai neiestāties sauc par gadījumnotikumiem. Ja izmēģinājuma rezultāts ir iepriekš zināms, notikumu sauc par determinētu notikumu. Ja izmēģinājuma rezultātā mūs interesējošais notikums nekad neiestājas, to sauc par neiespējamu notikumu, bet, ja izdarot kaut vienu izmēģinājumu, notikums vienmēr iestājas, tas ir nenovēršams notikums. Piemēram, sportistam metot rīku - disku, šķēpu vai granātu, nenovēršams notikums ir rīka piezemēšanās pēc lidojuma, bet neiespējams notikums, ka rīks tam piešķirtā paātrinājuma dēļ pārvarēs zemes pievilkšanas spēku un sāks riņķot ap zemi kaut kādā orbitā. Kas attiecas uz gadījumnotikumu iespējamību, tad nevar paredzēt viena izmēģinājuma rezultātu, bet, atkārtojot izmēģinājumus daudz reižu un reģistrējot notikuma iestāšanās reižu skaitu, var novērtēt šāda notikuma iestāšanās iespēju. Tādējādi esam nonākuši līdz varbūtību teorijas galvenajam jēdzienam -

notikuma varbūtībai, kas ir notikuma iespējamības mērs - skaitlis, kas atrodas intervalā no 0 līdz 1 un raksturo notikuma iespējamību. Neiespējama notikuma varbūtības vērtība ir 0, nenovēršama notikuma - 1. Gadījumnotikuma varbūtības vērtība atrodas starp šīm divām robežvērtībām. Varbūtības vērtībai tuvojoties 1, notikuma iespējamība palielinās, bet tuvojoties 0 - samazinās. Notikuma varbūtības klasiskā definīcija: notikuma ***A*** varbūtība ***P(A)*** ir šim notikumam labvēlīgo gadījumu skaita ***m*** attiecība, pret visu iespējamo gadījumu kopskaitu ***n***, pie nosacījuma, ka visos gadījumos saglabājas vienādi apstākļi un gadījumi ir nesavienojami, vienādi iespējami un vienīgi iespējami:

$$P(A) = \frac{m}{n} \tag{1.1}$$

Frāze "visu iespējamo gadījumu kopskaitu" šai definīcijā dara mūs uzmanīgus, jo ar līdzīgu sastapāmies iepazīstot jēdzienus "ģenerālkopa" un "paraugkopa", kad nonācām pie slēdziena, ka "visi iespējamie" bieži vien nozīmē bezgalīgi lielu skaitli un praktiski nav realizējami. Pēc klasiskās definīcijas notikuma varbūtības vērtību var noteikt tikai elementāriem notikumiem izmantojot noteiktas aksiomas. Piemēram, spēļu kauliņš mūsu izpratnē ir kubveida, homogens un simetrisks ķermenis. Tādēļ var uzskatīt, ka nokrītot uz līdzenas virsmas jebkuras kuba skaldnes atrašanās virspusē ir vienādi iespējama (aksioma). Protams virspusē parādās tikai viena no visām sešām skaldnēm, līdz ar to izslēdzot pārējo skaldņu parādīšanos (arī aksioma). Ja vēlamies aprēķināt "sešnieka" uzmešanas varbūtību, tad zinot, ka "sešnieks" ir viens no iespējamiem sešiem iznākumiem (kuba 6 skaldnes), izskaitļojam attiecību 1 : 6 = 0,166666... Praktiskajā dzīvē pārsvarā sastopamies ar sarežģītiem notikumiem, kad visi iespējamie iznākumi nav zināmi. Šādu notikumu iespējamības vērtēšanai izmanto klasiskajai varbūtībai līdzīgu jēdzienu - "statistisko varbūtību" jeb notikuma relatīvo biežumu (frekvenci), kuru aprēķina pēc praktisku novērojumu rezultātiem, kuros fiksēts kopējais izmēģinājumu skaits un notikuma iestāšanās reižu skaits. Notikuma ***A*** statistiskā varbūtība ***ω(A)*** ir to izmēģinājumu skaita *m*, kuros notikums A iestājies, attiecība pret veikto izmēģinājumu skaitu ***n***:

$$\omega(A) = \frac{m}{n} \tag{1.2}$$

Notikuma statistiskā varbūtība ir skaitlis, kurš palielinoties kopējam izmēģinājumu skaitam arvien vairāk tuvojas notikuma teorētiskās varbūtības vērtībai. To pierāda angļu statistiķa Kārļa Pirsona mēģinājums. Ģērboņa parādīšanās teorētiskā varbūtība metot monētu ir $1 : 2 = 0.5$ (1 labvēlīgais iznākums no 2 iespējamiem). K.Pirsona eksperimentā monētu meta divās sērijās pa 12000 reižu katrā (1. tabula).

1. tabula

K. Pirsona mēģinājums

Mēģinājumu skaits	Ģerboņa parādīšanās reižu skaits	Ģerboņa parādīšanās statistiskā varbūtība
12000	6019	0,5016
24000	12012	0,5005

Tabulā redzamie rezultāti apliecina, ka liela izmēģinājumu skaita gadījumā ģerboņa parādīšanās statistiskās varbūtības vērtība maz atšķiras no teorētiskās un divkāršojot mēģinājumu skaitu vēl vairāk tuvojas tai. Pamatojoties uz šādiem apsvērumiem praksē notikumu prognozēšanai izmanto statistisko varbūtību (relatīvo frekvenci), ko aprēķina pēc praktisku novērojumu rezultātiem. Novērojumu rezultāti var uzkrāties daudzu gadu laikā. Piemēram, laika prognoze ir viens no dotajā gadalaikā, datumā un situācijā iespējamiem variantiem ar vislielāko statistisko varbūtību. Prognozes pamatā ir ilggadīgu ikdienas meteoroloģisko novērojumu rezultāti. Vairāk neiedziļinoties varbūtību teorijā pievērsīsimies diviem jēdzieniem, kuri labi jāizprot izmantojot matemātiskās statistikas metodes. Paraugkopas analīzes rezultātus nevar absolūti pārnest uz ģenerālkopu. Vispārināšana notiek aptuveni t.i. ar kļūdu, kura nedrīkst pārsniegt kādu iepriekš pieņemtu vērtību, kas raksturo vispārinošo secinājumu precizitāti. Šo secinājumu neapstrīdamības varbūtību P sauc par rezultātu ticamības līmeni, bet kļūdas varbūtību α - par rezultātu būtiskuma līmeni. Tās ir divu pretēju notikumu varbūtības, ko saista sakarība $\alpha = 1 - P$, un kā vienu tā otru var izmantot secinājumu precizitātes raksturošanai. Rezultātu ticamības līmeņa vai būtiskuma līmeņa skaitlisko vērtību izvēlas atkarībā no risināmā uzdevuma satura - pēc tā, cik precīziem jābūt secinājumiem, lai pēc tiem varētu vadīties praksē. Pieņemtās ticamības un atbilstošās būtiskuma līmeņu vērtības dotas 2. tabulā.

2. tabula

Pieņemtie rezultātu ticamības un būtiskuma līmeņi

Ticamības līmeņi P	0,9	0,95	0,99	0,999
Būtiskuma līmeņi $\alpha = 1 - P$	0,1	0,05	0,01	0,001

Atsevišķos gadījumos pat visaugstākais ticamības līmenis nav pietiekošs, piemēram, ja runa ir par to, vai jaunais ārstniecības līdzeklis nav nāvējošs cilvēkam. Pedagoģiskos un bioloģiskos pētījumos pietiekošs secinājumu ticamības līmenis ir 0,95. Vienkāršoti var teikt, ka secinājumi realizēsies praksē 95 gadījumos no 100, ja vien praksē ievēros tos pašus apstākļus, kādi bija novērojumu vai eksperimenta laikā. Jāatgādina, ka katru attiecību var izteikt procentos, ja to reizina ar 100. Tādēļ lasītājam nevajadzētu samulst, sastopot atsevišķos literatūras avotos norādes par 95% ticamības vai 5% būtiskuma līmeni. Rakstot secinājumus, iekavās jānorāda pieņemtais ticamības vai būtiskuma līmenis, piemēram, ($P = 0,95$) vai ($\alpha = 0,05$).

1.3. Kad izmanto matemātiskās statistikas metodes?

Sporta zinātnē un praksē sastopami galvenokārt sekojoši statistiskās analīzes varianti:

1) Analizējot sportistu grupas sacensību vai testēšanas rezultātus (paraugkopu), aprēķina vidējo rādītāju (vidējo aritmētisko, mediānu vai modu), kurš kopumā raksturo dotās grupas sagatavotības līmeni. Aprēķina variēšanas rādītājus - standartnovirzi un variācijas koeficientu, kuri raksturo rezultātu blīvumu (izkliedi ap vidējo aritmētisko) jeb sportistu sagatavotības vienveidību. Variācijas koeficientam ir praktiska nozīme - pēc tā var vadīties organizējot turpmāko treniņa procesu - izlemt vai grupas sportistiem piemērojama vienāda treniņa slodze (blīvu rezultātu gadījumā) vai arī tā jāplāno individuāli. Slodzes lielumu savukārt izvēlas atbilstoši sagatavotības līmenim u.t.t. Pēc vidējā aritmētiskā standartkļūdas var spriest par vidējā aritmētiskā precizitāti attiecībā uz visiem dotās kategorijas sportistiem (ģenerālkopu).

2) Treniņu metodes efektivitāti var vērtēt pēc sportista sasniegumu pieauguma noteiktā laika periodā, tomēr jāņem vērā mērīšanas objekta nemitīgais mainīgums, kura ietekmē indivīda rezultātu pieaugums ne vienmēr atbilst kopējai tendencei treniņa grupā. Tāpēc aprēķina grupas vidējo rezultātu pieaugumu un novērtē tā ticamību, izmantojot noteiktu iepriekš pieņemtu ticamības līmeni. Šai procedūrā lieto statistikas metodes, ko sauc par atšķirību kritērijiem. Līdzīgi rīkojas, ja vēlas noskaidrot, vai kopumā ņemot viena sportistu grupa sagatavota labāk nekā otra - novērtē grupu vidējo rezultātu starpības ticamību.

3) Sporta treniņā sastopamies ar parādību, ko sauc par trenētības pārnešanu - trenējoties vienā fiziskajā vingrinājumā bieži novērojam rezultātu pieaugumu arī kādā citā vingrinājumā, kuram līdzīga kustību struktūra un tas pats fizioloģiskais mehanisms. Lai novērtētu, kā sasniegumi vienā vingrinājumā ietekmē sasniegumus citā vingrinājumā, izmanto korelācijas analīzi. Aprēķinātais korelācijas koeficients ir sasniegumu savstarpējās atkarības mērs. Ar tā palīdzību izvēlas efektīvākos treniņa un kontroles vingrinājumus - testus.

4) Lai prognozētu, piemēram, gaidāmo sportista sasniegumu sacensībās pēc kontroles vingrinājuma rezultāta vai arī pasaules rekordu nākošo Olimpisko spēļu gadā izmanto regresijas analīzē iegūto vienādojumu - izteiksmi, kura aptuveni apraksta šo lielumu savstarpējo atkarību.

5) Sacensību rezultātu vai kontrolvingrinājumu rezultātu pārvēršanai punktos izmanto dažādas vērtēšanas skalas (vienkāršākais šādas skalas piemērs ir daudzcīņas punktu tabula). Tā saukto standarta skalu gadījumā punktu aprēķināšanai par atskaites sākuma punktu izmanto rezultātu vidējo aritmētisko, bet par mērvienību - standartnovirzi. Matemātiskās statistikas metožu arsenāls ir ļoti plašs, tās izmanto arī eksperimenta plānošanai, kuras pamatuzdevums noteikt minimālo novērojamu skaitu, kas garantē ticamus rezultātus.

2. DATU ANALĪZE

2.1. Novērojumu rezultātu pirmapstrāde

Novērojumos iegūto skaitļu rindu sakārto, lai tā būtu pārskatāma un rastos priekšstats par pētamās pazīmes variēšanas īpatnībām. Datu pirmapstrādes paņēmieni atkarīgi no paraugkopas apjoma un pazīmes rakstura (kvantitatīva vai kvalitatīva). Maza apjoma paraugkopas variantes ranžē. Ar šo procedūru vienmēr sastopamies noformējot sacensību protokolu, kurā sportistu rezultātus ieraksta sasniegumu labuma secībā. Ranžēšana ir rezultātu sakārtošana pieaugošā vai dilstošā kārtībā. **Matemātiskajā statistikā variantes ranžē pieaugošā kārtībā (neatkarīgi no sporta sasnieguma labuma).** Tas jāievēro, lai novērstu kļūdainus slēdzienus, kad vienlaicīgi pētot vairākas pazīmes, aprēķinos izmanto varianšu rangus. Pēc sakārtošanas katrai variantei pieraksta tās rangu - skaitli, kas atbilst variantes vietai ranžētajā rindā. Vienādām variantēm dod vienādus rangus - varianšu vietu vidējo aritmētisko. Piemēram, 100 m skrējienā reģistrēti šādi rezultāti: 11,6 10,5 11,0 11,6 10,8 11,1 11,6 10,8 11,7. Šīs paraugkopas pirmapstrādes rezultāti attēloti 3. tabulā.

3. tabula

Ranžēšana

Numurs pēc kārtas	Variante	Rangs
1	10,5	1
2	10,8	2,5
3	10,8	2,5
4	11,0	4
5	11,1	5
6	11,6	7
7	11,6	7
8	11,6	7
9	11,7	9

Lielas paraugkopas variantes grupē. Pazīmes diskrētās un nepārtrauktās variēšanas gadījumā izmanto atšķirīgus grupēšanas paņēmienus. Apskatīsim grupēšanas operācijas pazīmes diskrētas variēšanas gadījumā. LSPA 1.kursa studentu-vīriešu vispārējās fiziskās sagtavotības sacensībās vingrinājumā "pievilkšanās pie stieņa" (reižu skaits) reģistrēti šādi rezultāti: 11, 11, 9, 11, 13, 11, 11, 9, 12, 13, 11, 8, 11, 11, 9, 10, 11, 5, 11, 11, 8, 11, 12, 10, 11, 11, 7, 9, 7, 11, 8, 11, 10, 11, 11, 8, 11, 7, 10, 11, 9, 11, 10, 12, 11, 9, 10, 11, 7, 11, 13, 12, 13, 12, 12, 11, 10, 12, 11, 10, 11, 6, 12, 13, 12, 12, 13, 12, 11, 11, 8, 7, 9. Saskaitām variantes - $\boldsymbol{n} = 73$. Izveidojam darba tabulu (4. tabula), kuras otrajā stabiņā pieaugošā kārtībā ierakstām varianšu vērtības bez atkārtošanās - katru tikai vienu reizi. Tādējādi katrai no tām tabulā atbilst viena rinda. Izskata novērojumu rezultātus un saskaita, cik reižu katra variante atkārtojas. Skaitli, kas rāda, cik reižu atkārtojas dotā variante, sauc par variantes frekvenci - $\boldsymbol{n_i}$. To ieraksta atbilstošajā rindā tabulas 3.stabiņā.

4. tabula

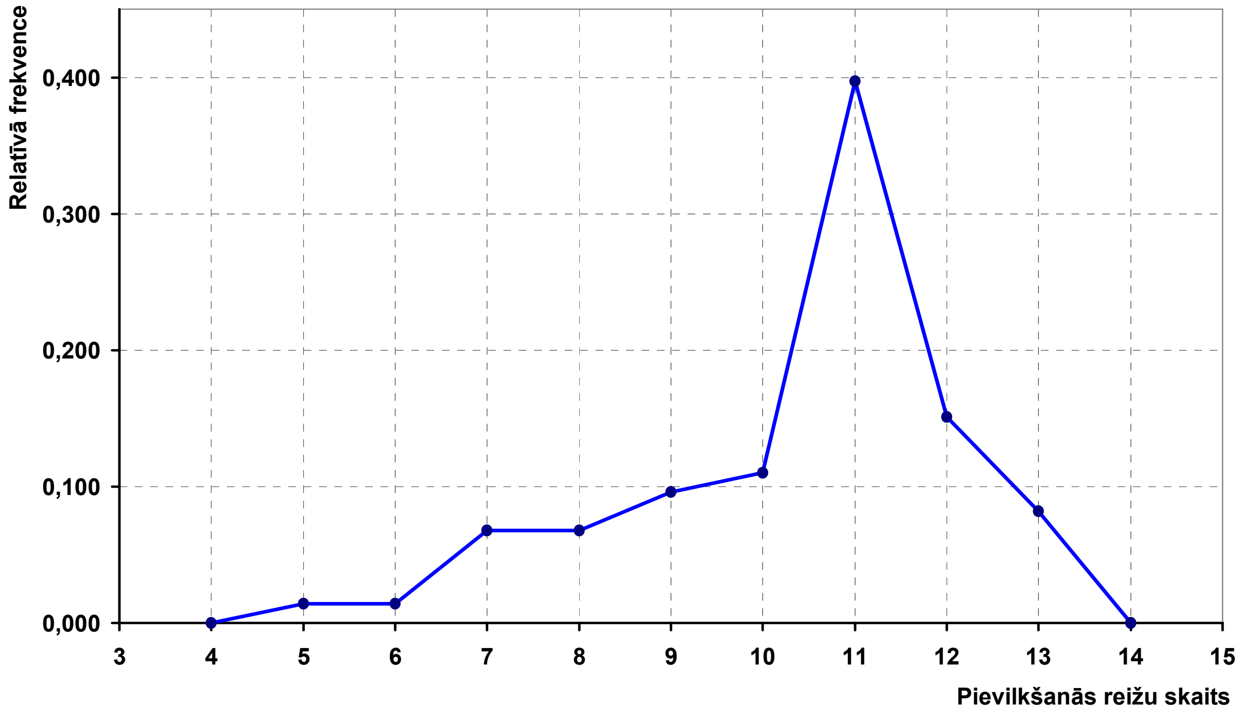
Empīriskā sadalījuma sastādīšana

Variante x_i	Variantes frekvence n_i	Relatīvā frekvence $\omega_{\varkappa}$	Kumulatīvā relatīvā frekvence
5	1	0.014	0.014
6	1	0.014	0.028
7	5	0.068	0.096
8	5	0.068	0.164
9	7	0.096	0.260
10	8	0.110	0.370
11	29	0.397	0.767
12	11	0.151	0.918
13	6	0.082	1.000
	$\Sigma n_i = 73$	$\Sigma\omega_{\varkappa}=1$	

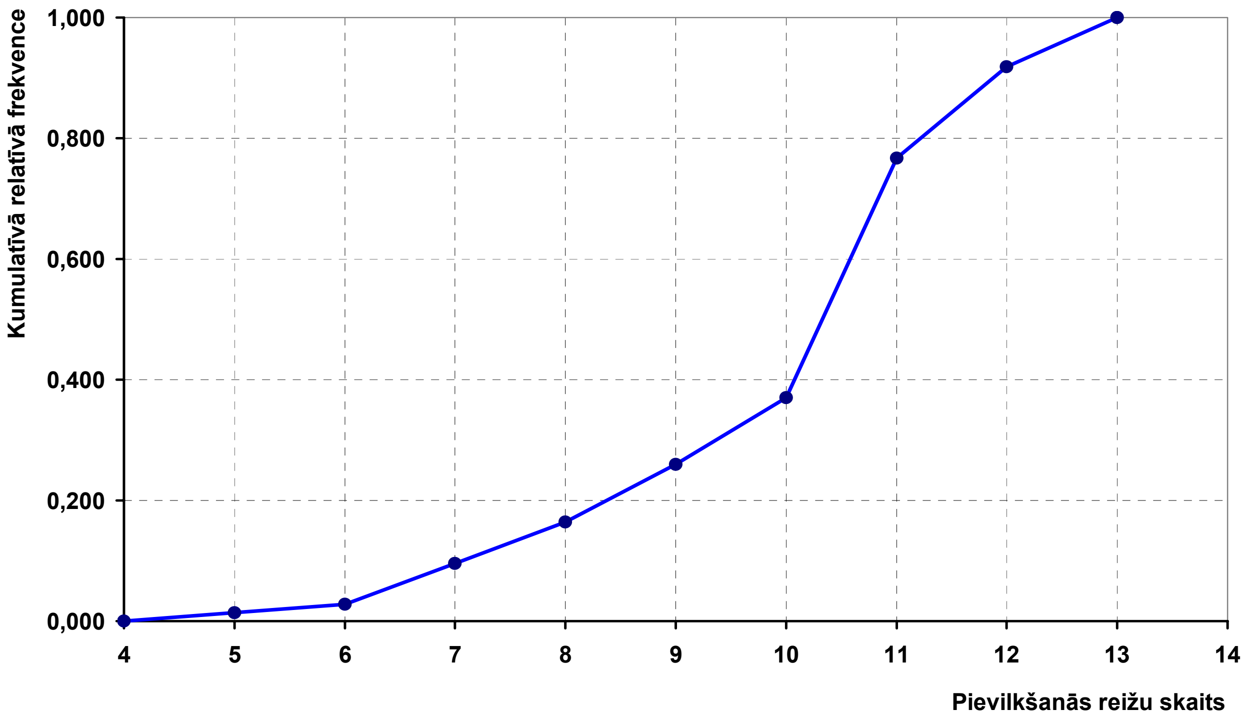
Lai gūtu priekšstatu par, katra rezultāta parādīšanās iespēju atsevišķā izmēģinājumā un noteiktu, kādu paraugkopas daļu veido dotās variantes vērtības, aprēķina variantes relatīvo frekvenci jeb statistisko varbūtību:

$$\omega_i = \frac{n_i}{n} \qquad (2.3)$$

un ieraksta tabulas 4.stabiņā. Lai būtu redzams, kā uzkrājas novērojumu skaits pieaugot varianšu vērtībām, aprēķina kumulatīvo relatīvo frekvenci. Tā ir vienāda ar dotās un par to mazāko varianšu relatīvo frekvenču summu, un rāda, kādu paraugkopas daļu veido dotā un par to mazākas variantes. Aprēķināto vērtību ieraksta tabulas 5. stabiņā. Varianšu frekvenču summa Σn_i ir vienāda ar paraugkopas apjomu $\boldsymbol{n}$, bet relatīvo frekvenču summa un lielākās variantes kumulatīvā relatīvā frekvence vienāda ar 1 (aprēķinot relatīvās frekvences noapaļošanas rezultātā var veidoties kļūda-novirze 0,001-0,002). Šīs sakarības izmanto pārbaudei, vai saskaitot varianšu frekvences un veicot iepriekšējos aprēķinus nav pielaistas kļūdas. Pa pāriem saistīto skaitļu - varianšu un to frekvenču virkni sauc par variāciju rindu jeb sakārtotu empīrisko sadalījumu. Empīriskā sadalījuma uzskatāmai attēlošanai var izmantot divus grafiskos attēlus: empīriskā sadalījuma poligonu (1. att.) un empīriskā sadalījuma kumulātu (2. att.).



1. att. Empīriskā sadalījuma poligons



2. att. Empīriskā sadalījuma kumulāta

Zīmējot abus grafikus uz abscisu ass attēlo variančšu vērtības, bet uz ordinātas poligonam - variančšu relatīvās frekvences, kumulātai - kumulatīvās relatīvās frekvences. Koordinātu plaknē atliek pa pāriem saistīto skaitļu (variančšu un frekvenču) punktus. Blakus esošām variančšu vērtībām atbilstošos punktus savieno ar taisnes nogriežņiem. Uzzīmētos grafikus sauc par sadalījuma poligonu un kumulātu.
Pazīmei variējot nepārtraukti, variantes grupē klasēs. Grupēsim 1. kursa studentu-vīriešu vieglatlētikas sacensību rezultātus lodes grūšanā:

9,73	10,68	9,36	9,50	9,20	10,27
10,00	9,63	11,40	9,83	8,80	9,30
9,04	10,08	10,22	9,80	9,34	9,45
9,32	10,00	8,40	8,95	9,35	9,50
6,99	10,45	11,14	11,00	9,33	9,00
9,20	10,00	10,00	11,10	9,78	8,80
9,00	9,30	8,80	10,00	9,20	7,88
9,00	9,80	9,21	9,10	8,90	11,00
9,10	8,90	11,43	10,48	9,49	8,20
9,87	8,80	7,80	9,50	9,48	8,97
9,80	9,00	10,55	9,24	9,45	9,00

Paraugkopas apjoms $\boldsymbol{n} = 66$, $\boldsymbol{x}_{min} = 6{,}99$, $\boldsymbol{x}_{max} = 11{,}43$. Variēšanas intervālu starp $\boldsymbol{x}_{min}$ un $\boldsymbol{x}_{max}$ sadala vairākos vienādos intervālos - klasēs, kuru skaitu nosaka pēc formulas:

$$k = 1 + 3{,}32 \lg n = 1 + 3{,}32 \lg 67 \approx 7 \qquad (2.4)$$

Klases intervāla garumu aprēķina, noapaļojot uz augšu - uz pāra skaitli:

$$c = \frac{x_{\max} - x_{\min}}{k - 1} = \frac{12{,}17 - 6{,}99}{7 - 1} \approx 0{,}74 \qquad (2.5)$$

Aprēķina 1. klases minimālo robežu, pieņemot, ka $\boldsymbol{x}_{\min}$ atrodas šī intervāla vidū:

$$t_0 = x_{\min} - \frac{c}{2} = 6{,}99 - \frac{0{,}74}{2} \approx 6{,}62 \qquad (2.6)$$

Ik reizi pieskaitot pa klases intervālam aprēķina pārējās klašu robežas $t_1 = 6{,}62 + 0{,}74 = 7{,}36$; $t_2 = 8{,}10$; $t_3 = 8{,}84$; $t_4 = 9{,}58$; $t_5 = 10{,}32$; $t_6 = 11{,}06$; $t_7 = 11{,}80$ un ieraksta darba tabulas (5. tabula) 2. stabiņā. Pirmās klases maksimālā robeža ir otrās klases minimālā robeža, otrās klases maksimālā robeža ir trešās klases minimālā robeža u.t.t. Pieņemot, ka 1. klases intervāla vidus jeb vidējā vērtība ir $x_1 = x_{min}$, aprēķina pārējo klašu vidējās vērtības, ik reizi pieskaitot pa klases intervālam: $x_2 = x_1 + c = 6{,}99 + 0{,}74 = 7{,}73$; $t_3 = 8{,}47$; $x_4 = 9{,}21$; $x_5 = 9{,}95$; $x_6 = 10{,}69$; $x_7 = 11{,}43$ un ieraksta darba tabulas 4. stabiņā. Izskatot novērojumu rezultātus saskaita, cik varianšu ietilpst katrā klasē. Iegūto skaitli sauc par klases frekvenci. To ieraksta darba tabulas (26. tabula) 3. stabiņā. Varianti, kura ir uz klases robežas, iedala klasē ar zemākām vērtībām. Turpmāk darba tabulu izpilda tāpat kā grupējot variantes diskrēti variējošas pazīmes gadījumā: aprēķina relatīvās un kumulatīvās relatīvās klašu frekvences. Pirmapstrādes rezultātu pareizības pārbaudei izmanto klašu frekvenču summu un klašu relatīvo frekvenču summu $\sum n_i = n$ $\sum \omega_i = 1{,}000$.

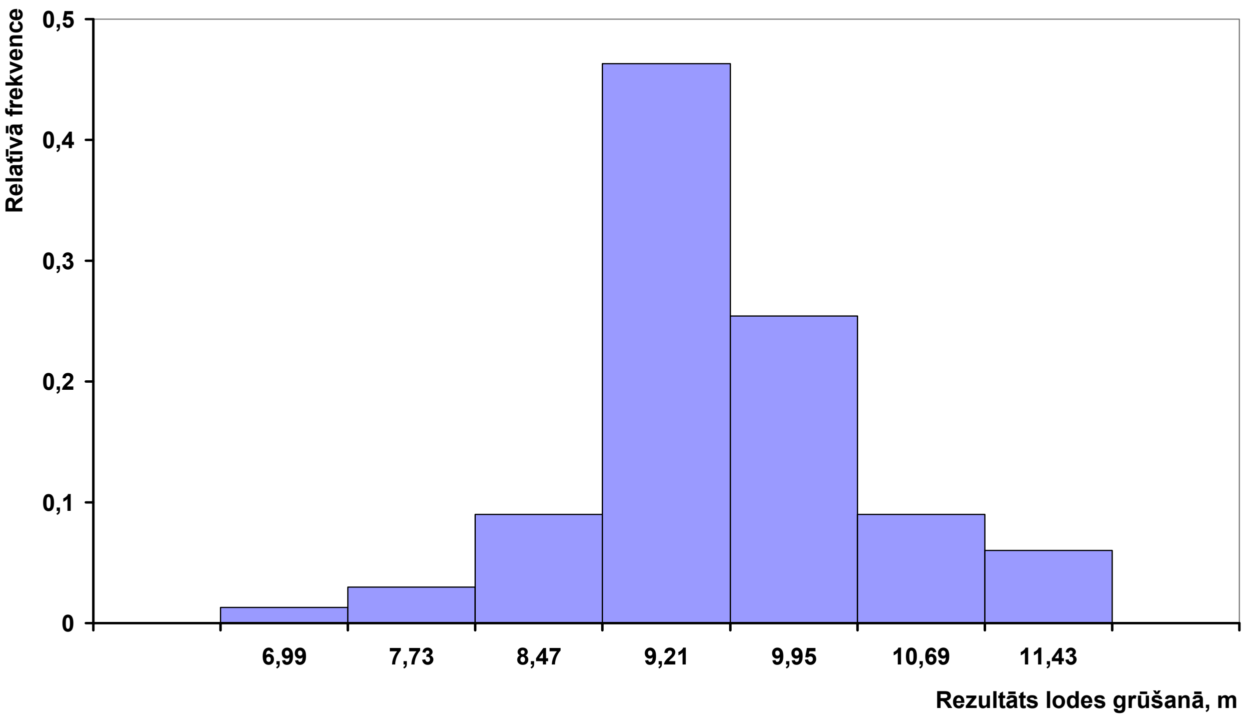
Pirmapstrādē iegūtās klašu vidējās vērtības un klašu frekvences veido intervālu variāciju rindu. Šāda empīriskā sadalījuma grafiskai attēlošanai izmanto stabiņveida diagrammu, ko sauc par histogrammu (3. zīm.). Uz abscisas attēlo pazīmes skaitliskās vērtības - klašu robežas, bet uz ordinātas - statistiskās varbūtības - klašu relatīvās frekvences.

5. tabula

Intervālu variāciju rindas sastādīšana

Nr	Klases robe˛as	Klases videja vertiba	Klases frekvence	Klases relativa frekvence	Kumulativa frekvence
k	x_i	x_k	n_i	ω_i	ω_i'
1	6,62÷7,36	6,99	1	0,013	0,013
2	7,36÷8,10	7,73	2	0,030	0,043
3	8,10÷8,84	8,47	6	0,090	0,133
4	8,84÷9,58	9,21	31	0,463	0,596
5	9,58÷10,32	9,95	17	0,254	0,850
6	10,32÷11,06	10,69	6	0,090	0,940
7	11,06÷11,80	11,43	4	0,060	1,000
			$\sum n_i = 67$	$\sum \omega_i = 1{,}000$	

Histogrammā visu stabiņu pamatnes ir vienādas un atbilst klases intervāla garumam, bet stabiņa augstums - klases relatīvai frekvencei. Intervālu variāciju rindā katras klases variantes tiek pielīdzinātas klases vidējai vērtībai, tādēļ zīmējot empīriskā sadalījuma kumulātu uz abscisu ass attēlo klašu vidējās vērtības, bet uz ordinātas - kumulatīvās relatīvās frekvences. Paraugkopas pirmapstrādes rezultātus, kas iegūti grupējot variantes, turpmāk var izmantot aprēķinot empīriskā sadalījuma parametrus - statistiskos rādītājus. Lielu paraugkopu gadījumā tas atvieglo aprēķinus, ja nav pieejams dators. Visas aprakstītās novērojumu rezultātu pirmapstrādes operācijas var automatizēt izmantojot datoru un kādu no tabulu procesoriem, no kuriem šodien neapšaubāmi populārākais it MICROSOFT EXCEL.



3. att. Empīriskā sadalījuma histogramma

2.2. Aprakstošā statistika

Ranžēta paraugkopa un grupējot izveidotais sakārtotais empīriskais sadalījums ir diezgan grūti pārskatāmas skaitļu rindas, to vienīgais kompaktais raksturojums ir paraugkopas apjoms. Statistiskās analīzes uzdevums - raksturot pētāmās pazīmes vidējo vērtību un variēšanu. Bez tam jāzin, cik ticami ir šie raksturpjumi - cik labi paraugkopa reprezentē ģenerālkopu. Uz šiem jautājumiem var atbildēt aprēķinot trīs skaitļus, kuri koncentrētā veidā satur nepieciešamo, derīgo informāciju. Tos sauc par statistiskajiem rādītājiem jeb empīriskā sadalījuma parametriem. Ir trīs statistisko rādītāju grupas: vidējie rādītāji, variēšanas jeb izkliedes rādītāji un reprezentācijas jeb standartkļūdas. Vidējie rādītāji: vidējais aritmētiskais, vidējais ģeometriskais, vidējais kvadrātiskais, vidējais harmoniskais, mediāna, moda raksturo pazīmes vidējo vērtību. Katrā konkrētā uzdevumā izmanto vienu - piemērotāko no šiem rādītājiem - kvantitatīvām pazīmēm visbiežāk vidējo aritmētisko. Variēšanas rādītāji raksturo varianšu izkliedi ap vidējo aritmētisko. Tie ir: vidējā absolūtā novirze, dispersija, stadartnovirze (vidējā kvadrātiskā novirze), variācijas koeficients, normētā novirze. Visbiežāk izmanto standartnovirzi un variācijas koeficientu. Reprezentācijas kļūdas raksturo neprecizitāti, kas rodas aizstājot ģenerālkopas raksturojumus ar paraugkopas statistiskajiem rādītājiem. Tās novērtēšanai katram vidējam un katram izkliedes rādītājam aprēķina standartkļūdu.

2.2.1. Vidējie rādītāji

Vidējais aritmētiskais $\bar{x}$, vidējais ģeometriskais $\bar{x}_g$, vidējais kvadrātiskais $\bar{x}_{kv}$, vidējais harmoniskais $\bar{x}_h$, vidējais kubiskais $\bar{x}_q$, mediana (Me) un moda (Mo) pēc vērtības savstarpēji atšķiras. Vidējo rādītāju novietojums variēšanas intervālā ir šāds:

$$x_{\min} < \bar{x}_h < \bar{x}_g < \bar{x} < \bar{x}_{kv} < \bar{x}_q < x_{\max} \qquad (2.7)$$

Mediāna un moda šai rindā var atrasties dažādās vietās, atkarībā no pazīmes variēšanas īpatnībām. Tādēļ aprakstot pētīšanas metodes jānorāda, kurš no rādītājiem izmantots. Visbiežāk izmanto vidējo aritmētisko, bet kvalitatīvu pazīmju raksturošanai - mediānu vai modu. Pārējie rādītāji noderīgi atsevišķos gadījumos, par ko var smelties informāciju speciālajā literatūrā. Iepazīstoties ar vidējo aritmētisko, mediānu un modu, piemēram aplūkosim šādu paraugkopu:

4 4 5 7 10

Moda ir variante ar vislielāko frekvenci jeb pazīmes vērtība, kas sastopama visbiežāk. Mūsu piemērā:

Mo = 4

Mediāna ir variante, kas atrodas ranžētas paraugkopas vidū. Mūsu piemērā:

Me = 5

Vidējais aritmētiskais ir paraugkopas visu variansu summa dalīta ar variansu skaitu:

$$\bar{x} = \frac{4+4+5+7+10}{5} = 6$$

Ģenerālkopai šis rādītājs nav izskaitļojams, jo nevar iegūt un apstrādāt visas šīs kopas variantes. Paraugkopas vidējam aritmētiskam atbilstošo lielumu ģenerālkopā sauc par vidējo ģenerālo jeb pazīmes vidējo vērtību un apzīmē ar grieķu burtu μ (mī). Mēdz teikt, ka vidējais ģenerālais ir vidējā aritmētiskā matemātiskā cerība. Dotajā piemērā vidējais aritmētiskais ir svarīgākais no aplūkotajiem rādītājiem, jo ir "visjūtīgākais" - tā vērtība mainās, mainoties jebkurai variantei. Turpretī mediānas vai modas vērtība mainās tikai dažreiz. Piemēram, ja aplūkotajā kopā varianti 7 aizvieto ar 17, ranžētā rinda izskatās šādi:

4 4 5 10 17

Mediānas un modas vērtības nemainās, bet vidējā aritmētiskā vērtība ir 8. Mediāna iegūtu citu vērtību tikai tad, ja centrālās variantes 5 vietā būtu cits skaitlis. Mediānu bieži izmanto sporta sacensību tiesneši. Ja nav nav elektroniskās laika mērīšanas sistēmas, distances veikšanas laiku reģistrē ar trim rokas hronometriem un par sacensību rezultātu pieņem mediānu t.i. atmet labāko un sliktāko rezultātu (riteņbraukšanas hitos, kalnu slēpošanā, vieglatlētikas sprintā, peldēšanā, ātrslidošanā). Vidējais aritmētiskais - labi raksturo pazīmes vidējo vērtību - sportistu grupas vidējo sacensību vai kontroles vingrinājuma rezultātu, funkcionālās sagatavotības vidējo līmeni u.c. Vidējo aritmētisko aprēķina pēc formulas:

$$\bar{x} = \frac{\sum x_i}{n} \quad (2.8),$$

kur x_i - variante; n - varianšu skaits. Ja veicot paraugkopas pirmapstrādi sastādīta variāciju rinda, vidējo aritmētisko var aprēķināt izmantojot formulu:

$$\bar{x} = \frac{\sum x_k n_i}{n} \quad (2.9),$$

kur x_i - variantes vērtība vai klases vidējā vērtība; n_i - variantes vai klases frekvence; n - varianšu skaits. Ja grupējot sastādīta intervālu variācijas rinda, tad pēc formulas (1.9) aprēķināto vidējo sauc par svērto vidējo aritmētisko, jo visas vienā klasē ietilpstošās variantes tiek pielīdzinātas klases vidējai vērtībai. Vidējais aritmētiskais tāpat kā pārējie vidējie rādītāji ir nosaukts skaitlis - tam ir tā pati mērvienība, kas atsevišķai variantei. Noapaļojot aprēķināto vidējo, jāņem vērā, ka nevar precīzāk aprēķināt nekā mērīts - noapaļo līdz tādai precizitātei, ar kādu dotas paraugkopas variantes vai arī atstājot aiz komata vienu decimālzīmi vairāk, ja skaitlis tiks ievietots formulās turpmākiem aprēķiniem. Svarīga vidējā aritmētiskā īpašība ir tā, ka visu variānšu centrālo noviržu (variantes novirze no vidējā) summa ir 0:

$$\sum \left(x_i - \bar{x}\right) = 0 \quad (2.10)$$

Nejaušu cēloņu radīto individuālo gadījumu skaitliskās dažādības savstarpējās līdzsvarošanās (dzēšanās) procesu statistikā sauc par lielā skaita likumu. Neierobežoti palielinot pazīmes (gadījumlieluma) savstarpēji neatkarīgu novērojumu skaitu, iegūto rezultātu vidējā vērtība tuvojas noteiktam konstantam lielumam . Proti, ja x_1, x_2 ,...x_n ir neatkarīgi gadījummainīgie ar vienu un to pašu sadalījumu, un $n \rightarrow \infty$, tad

$\frac{x_1 + x_2 + \ldots + x_n}{n}$ tuvojas $\bar{x}$ matemātiskai cerībai μ. Tādējādi vidējais aritmētiskais it kā ir paraugkopas smaguma centrs.

2.2.2. Variēšanas rādītāji

Vidējie rādītāji nav universāli, jo pazīmes ar vienādiem vidējiem var atšķirties pēc variēšanas lieluma un rakstura. Variēšanas sinonīmi no statistikas viedokļa ir jēdzieni: rezultātu izkliede, blīvums, vienveidība, mainīgums. Tie visi, katrs savā veidā raksturo vienu parādību - pazīmes variēšanu kā varianšu jeb novērojumu rezultātu atšķirības: jo mazākas ir paraugkopas variantu savstarpējās atšķirības, jo pazīme variē mazāk, jo variantes vairāk savstarpēji atšķiras, jo pazīme variē stiprāk. Arī sportista meistarību raksturo ne tikai atkārtotos mēģinājumos sasniegtais vidējais vai augstākais rezultāts, bet arī rezultātu stabilitāte atkārtotos mēģinājumos. Tādēļ, kopā ar vidējo rādītāju pazīmes raksturošanai izmanto arī variēšanas rādītājus. Tie var būt variēšanas intervāla robežas - minimālā un maksimālā variante: x_{min} un x_{max} , variācijas intervāla garums (amplitūda) $\boldsymbol{l = x_{max} - x_{min}}$. Jo stiprāk variē pazīme, jo lielāka variācijas amplitūda un otrādi. Tie ir mazinformatīvi variēšanas rādītāji, jo vienāda intervāla robežās, dažādu paraugkopu variantes var būt sadalītas dažādi, bet vidējie aritmētiskie tai pat laikā var būt vienādi. Pilnīgāk pazīmes variēšanu var raksturot izpētot variantu izkliedi ap vidējo aritmētisko, kurš, kā iepriekš noskaidrojām, ir it kā paraugkopas smaguma centrs. Par pamatu šādam variēšanas vērtējumam, var izmantot variantu centrālās novirzes $x_i - \bar{x}$. Viens no iespējamiem izkliedes rādītājiem var būt vidējā absolūtā novirze:

$$\delta = \frac{\sum |(x_i - \bar{x})|}{n} \tag{2.11}$$

Tomēr šo rādītāju izmanto reti, ērtāka un informatīvāka ir vidējā kvadrātiskā novirze jeb standartnovirze, kuru aprēķina kā pozitīvu kvadrātsakni no dispersijas - s . Dispersija ir variantu centrālo noviržu kvadrātu summas vidējais aritmētiskais:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} \tag{2.12}$$

Dispersijai piemīt vairākas vērtīgas īpašības, kuras izmanto statistiskās analīzes metodēs, kas pazīstamas ar nosaukumu - dispersijas analīze. Pirmkārt, ar dispersiju var novērtēt kā grupas tā individuālo variāciju. Otrkārt, ja pazīme atkarīga no vairākiem ārējiem faktoriem, dispersiju var sadalīt komponentēs, no kurām katra raksturo viena faktora ietekmes svaru. Kā patstāvīgu izkliedes rādītāju dispersiju neizmanto, šim nolūkam lieto standartnovirzi:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \tag{2.13}$$

Aprēķinot mazas paraugkopas standartnovirzi, vērtējuma precizitātes paaugstināšanas nolūkā *n* aizvieto ar *n - 1*. Praktiskiem aprēķiniem formula ir neērta daudzo skaitļošanas operāciju dēļ, un to izmantojot parādās papildkļūda - vidējā aritmētiskā noapaļošanas rezultātā. Tādēļ aprēķiniem izmanto formulas pārveidoto t.i. darba variantu:

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} \tag{2.14}$$

Palielinoties pazīmes variēšanai pieaug standartnovirzes vērtība, savukārt mazāka standartnovirze atbilst blīvākiem, vienveidīgākiem rezultātiem. Standartnovirze ir nosaukts skaitlis, tai ir tā pati mērvienība, kas variantēm. Aprēķināto standartnovirzes vērtību noapaļo līdz precizitātei, ar kādu dotas variantes. Standartnovirzei ir divi trūkumi, kuru dēļ to nevar vienmēr izmantot. Pēc formulas (1.13) redzams, ka standartnovirzes vērtība atkarīga no vidējā aritmētiskā. Tādēļ, ja vēlas salīdzināt rezultātu variēšanu divām sportistu grupām, standartnovirzi var izmantot šim nolūkam tikai tad, ja abu grupu vidējie (aritmētiskie) rezultāti ir vienādi. Ja vēlas salīdzināt divu dažādu pazīmju variēšanu, traucē arī mērvienība - standartnovirzes, kas izteiktas dažādās mērvienībās nav salīdzināmas. Šo iemeslu dēļ variēšanas vērtēšanai parasti izmanto standartnovirzes relatīvo vērtību - variācijas koeficientu. Tas ir standartnovirzes attiecība pret vidējo aritmētisko procentos:

$$s\% = \frac{s}{\bar{x}} \cdot 100\% \tag{2.15}$$

Variācijas koeficients ir universāls izkliedes rādītājs, tādēļ, ka ir nenosaukts skaitlis. Aprēķināto variācijas koeficientu noapaļo līdz vienai decimālzīmei aiz komata. Svarīgi ir atcerēties variācijas koeficienta robežvērtību - 10%. Ja s% ≤ 10%, novērojumu rezultāti ir vienveidīgi, pretējā gadījumā tos par vienveidīgiem uzskatīt nevar un jānoskaidro lielās variācijas cēloņi. Tie var būt rupjas mērīšanas kļūdas vai arī dotai grupai netipiska objekta klātbūtne novērojumos. Šai gadījumā pārbauda ekstremālās variantes - x_{min} un x_{max} un, ja neapstiprinās to piederība ģenerālkopai (skat."Ekstremālo rezultātu pārbaude") atmet kā analīzei nederīgas. Variācijas koeficienta izmantošana ir ierobežota - tas nav derīgs, ja mērīšana notiek intervālu skalā, piemēram, mērot leņķus vai temperatūru.

Normētā novirze ir rādītājs, kurš raksturo atsevišķu varianti, parādot, cik standartnoviržu attālumā no vidējā aritmētiskā tā atrodas:

$$t = \frac{x_i - \overline{x}}{s} \qquad (2.16)$$

Normētā novirze ir nenosaukts skaitlis, tādēļ šo rādītāju izmanto ļoti dažādu uzdevumu risināšanai.

Pazīmes variēšanas īpatnības - empīriskā sadalījuma nobīdi uz abscisu ass no variēšanas intervāla vidus raksturo asimetrijas rādītājs:

$$A = \frac{\sum x_i^3 - \frac{3\sum x_i \sum x_i^2}{n} + \frac{2\left(\sum x_i\right)^2}{n^2}}{ns^3} \qquad (2.17)$$

Jo lielāka sadalījuma asimetrija, jo lielāka - centrālo noviržu kubu summa. Asimetrijas rādītājs ir nenosaukts pozitīvs vai negatīvs skaitlis, pilnīgi simetriska sadalījuma gadījumā tas ir 0. Labās (negatīvās) asimetrijas gadījumā lielākā daļa variānšu grupējas sadalījuma labajā pusē, $A < 0$ un Mo > Me. Ja sadalījums nobīdīts pa kreisi (pozitīvā asimetrija), $A > 0$ un Mo < Me < $\overline{x}$. Kā simetriski, tā asimetriski rādītāji var būt ekscesīvi. Pozitīva ekscesa gadījumā variānšu lielākā daļa sablīvējas ap vidējo aritmētisko, bet pie negatīva ekscesa - variantes blīvāk novietotas pie abām paraugkopas robežām.

Ekscesu raksturo ekscesa raditajs:

$$E = \frac{\sum x_i^4 - \frac{4\sum x_i \sum x_i^3}{n} + \frac{6\sum x_i^2\left(\sum x_i\right)^2}{n^2} - \frac{3\left(\sum x_i\right)^4}{n^3}}{ns^4} \qquad (2.18)$$

Ekscesa rādītājs ir nenosaukts pozitīvs, vai negatīvs skaitlis. Ja sadalījumā ekscess nav novērojams, tas vienāds ar 0.

2.2.3. Reprezentācijas kļūdas

Reprezentācijas jeb standartkļūdu aprēķina katram vidējam un izkliedes rādītājam. Tā raksturo neprecizitāti, kas rodas vispārinot paraugkopas rādītāju uz ģenerālkopu. Standartkļūdas apzīmē ar burtu ***s***, pievienojot indeksā rādītāja apzīmējumu, kuram kļūda aprēķināta.

Vidējā aritmētiskā standartkļūda:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{2.19}$$

Standartnovirzes standartkļūda:

$$s_s = \frac{s}{\sqrt{2n}} \tag{2.20}$$

Variācijas koeficienta standartkļūda:

$$s_{s\%} = \frac{s\%}{\sqrt{2n}}\sqrt{1 + 2\left(\frac{s\%}{100}\right)} \tag{2.21}$$

Vidējā aritmētiskā un standartnovirzes standartkļūdas ir nosaukti skaitļi, ar to mērvienību, kas variantēm. Standartkļūdas noapaļo līdz precizitātei, ar kādu dotas variantes. Secinājumos, tekstā un tabulās parasti standartkļūdu raksta kopā ar atbilstošo rādītāju, kuram tā aprēķināta: $\bar{x} \pm s_{\bar{x}}$, $s \pm s_s$, $s\% \pm s_{s\%}$.

3. NORMĀLAIS SADALĪJUMS

3.1. Teorētiskie sadalījumi

Katrai pazīmei raksturīgas noteiktas variēšanas īpatnības - likumsakarības. Daļēji tās atspoguļo sakārtots empīriskais sadalījums - pazīmes vērtības (variantes) un to statistiskās varbūtības (relatīvās frekvences):

$$\begin{array}{lllll} x_1, & x_2, & x_3, \ldots & x_{n-1}, & x_n \\ \omega_1, & \omega_2, & \omega_3, \ldots & \omega_{n-1}, & \omega_n \end{array}$$

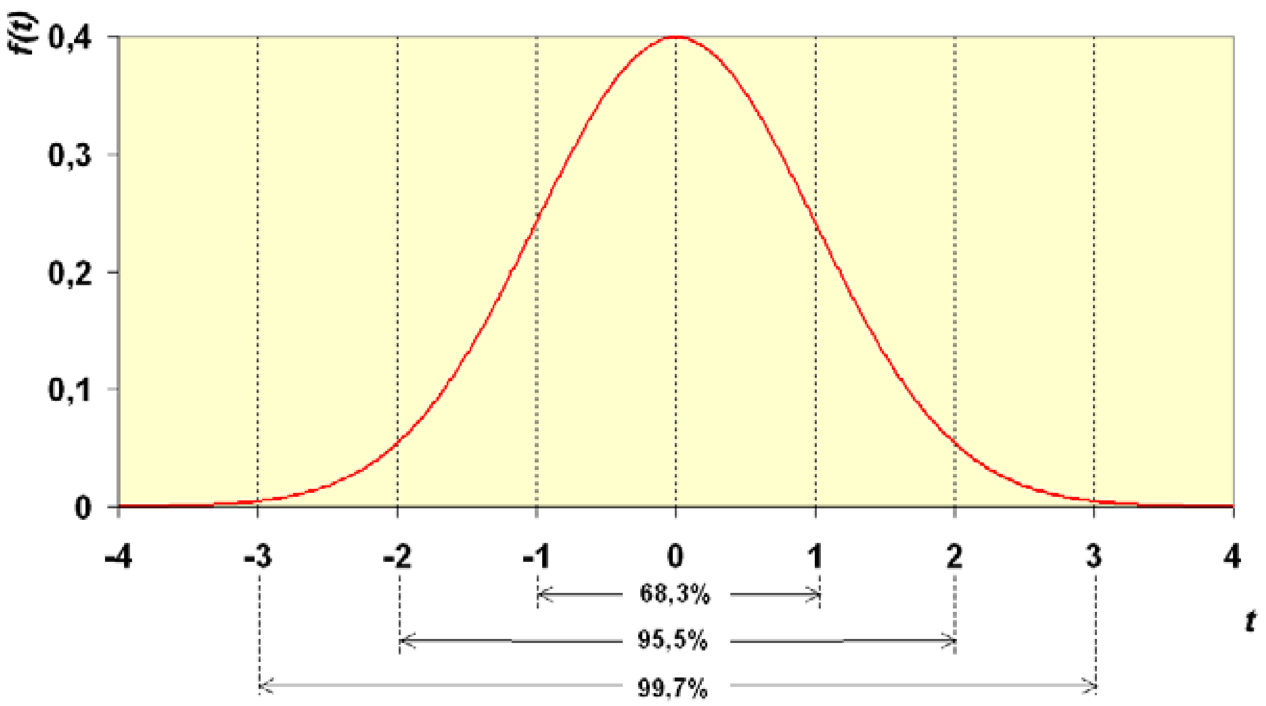
Pieņemsim, ka mēs neierobežoti palielinām novērojumu skaitu - $n \to \infty$, tad varianšu statistisko varbūtību vērtības tuvojas teorētisko varbūtību vērtībām, kas būtībā ir pazīmes skaitliskās vērtības funkcija - $\varphi(x)$. Šādi iegūtās skaitļu rindas veido tā sauktā teorētiskā sadalījuma rindas un vispārīgā veidā attēlo dotās pazīmes variēšanas likumsakarības:

$$\begin{array}{lllll} x_1, & x_2, & x_3, \ldots & x_{n-1}, & x_n \\ \varphi_1, & \varphi_2, & \varphi_3, \ldots & \varphi_{n-1}, & \varphi_n \end{array}$$

Aprakstot šo sakarību ar atbilstošu vienādojumu iegūstam teorētisko sadalījumu - dotās pazīmes variēšanas vispārīgo likumu jeb matemātisko modeli. Zināmi vairāki teorētiskie sadalījumi, plašāk pazīstami un bioloģijā izmantotie ir binomiālais, normālais un Stjūdenta sadalījumi. Bioloģiskās pazīmes bieži variē abilstoši normālā sadalījuma likumam. Tas ir viens no pazīmes nepārtrauktas variēšanas vispārīgajiem likumiem. Pirmo reizi 1733.g. to formulējis A.Muaurs, pēc tam neatkarīgi viens no otra Laplass un K.Gauss. Universālas likumsakarības formulēšanai izmanto standartizētu normālo sadalījumu. Tajā pazīmes vērtības aizstāj ar to normētajām novirzēm, tādējādi notiek atbrīvošanās no mērvienības. Tad atbilstošās varbūtību vērtības paliek nemainīgas, standartnovirze kļūst par mērvienību - $\sigma = 1$, bet vidējā vērtība kļūst par atskaites sākuma punktu - $\mu = 0$. Standartizēto normālo sadalījumu apraksta vienādojums:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \qquad (3.22),$$

kur $\varphi(t)$ - vērtības x normētās novirzes t varbūtība;
t - pazīmes vērtības normētā novirze;
e - naturālā logaritma bāze (2,718281).



4. att. Normālā sadalījuma poligons

Sadalījuma poligons ir simetriska, zvanveida līkne (9.3.zīm). Ja pazīme variē atbilstoši normālajam likumam, tad varianšu izvietojumu ap vidējo vērtību raksturo daži zīmīgi skaitļi. Vienas standartnovirzes intervālā ap vidējo vērtību atrodas 68,3% visu pazīmes vērtību, divu standartnoviržu intervālā - 95,5%, bet trīs standartnovižu intervālā - 99,7%.

Šo parādību sauc arī par triju standartnoviržu likumu, jo praktiski visas variantes izņemot 3 no tūkstoša atrodas trīs standartnoviržu intervālā ap vidējo vērtību. Varbūtību, kas raksturo iespēju, ka mērot iegūtās variante iekļūs noteiktā intervālā ap vidējo vērtību, sauc par varbūtību integrāli un apzīmē ar $\Phi(t)$. Triju standartnoviržu likumu raksturo šādas varbūtību integrāļa vērtības: $\Phi(1) = 0{,}683$, $\Phi(2) = 0{,}955$, $\Phi(3) = 0{,}997$. **Ņemot par pamatu normālā sadalījuma likumu, ir izstrādātas daudzas statistiskās analīzes metodes. To izmantošana konkrētā gadījumā pieļaujama tikai tad, ja nav šaubu, ka pētamā pazīme variē atbilstoši normālajam likumam.** Lai par to pārliecinātos izmanto dažādas pārbaudes metodes. Bioloģiskajos un sporta pētījumos bieži sastopamies ar neliela apjoma paraugkopām. Ja pazīme variē atbilstoši normālajam likumam, šādā paraugkopā biežāk sastopamas variantes, kas tuvāk vidējai vērtībai. Ņemot vērā šo parādību, angļu statistiķis V.Gossets izstrādāja normālā sadalījuma modeli (variantu) mazām kopām, ko publicējot nosauca autora pseidonīma vārdā par Stjūdenta sadalījumu. Tā svarīgākais lielums ir Stjūdenta sadalījuma normētā novirze - t , kuras vērtība atkarīga no brīvības pakāpju skaita un pieņemtā ticamības līmeņa. Stjūdenta sadalījuma normēto novirzi izmanto kā palīglielumu dažādās tā paša autora izstrādātās statistikas metodēs. Šī rādītāja atbilstošo vērtību var nolasīt tā sauktajā Stjūdenta tabulā (1.pielikums).

3.2. Ticamības intervâls

Kā zināms no iepriekšējā materiāla, paraugkopas analīzē iegūtie statistiskie rādītāji vispārināmi uz ģenerālkopu. Tieša šo rādītāju vērtību izmantošana būtu nekorekta. Proti, ja no vienas ģenerālkopas ir sastādītas vairākas paraugkopas, tad šo paraugkopu statistiskie rādītāji, piemēram, vidējie aritmētiskie variē - tiem ir dažādas vērtības. Kura tad būtu pareizā? Nevienu no šīm vērtībām nevar viennozīmīgi izmantot, kā vidējo ģenerālo, bet katra no tām ir derīga, lai atbilstoši pieņemtajam ticamības līmenim aprēķinātu intervāla robežas, kurā šis rādītājs atrodas. Intervālu, kurā ar pieņemto ticamības līmeni atrodas ģenerālkopas statistiskais rādītājs sauc par ticamības jeb reprezentācijas intervālu. Ticamības intervāla robežas aprēķina, lai raksturotu mūs interesējošo ģenerālkopas rādītāju - dotu tā vērtējumu. Vidējā aritmētiskā ticamības intervāla robežas aprēķina izmantojot izteiksmi:

$$\bar{x} - s_{\bar{x}} \cdot t_{\alpha;\nu} < \mu < \bar{x} + s_{\bar{x}} \cdot t_{\alpha;\nu} \qquad (3.23),$$

kur μ - vidējais ģenerālais;

$\bar{x}$ - paraugkopas vidējais aritmētiskais;

$t_{\alpha;\nu}$ - Stjūdenta sadalījuma normētā novirze, kuru nolasa tabulā (1. pielikums) pēc būtiskuma līmeņa α = 0,05 un brīvības pakāpju skaita $\boldsymbol{n = n - 1}$.

Aplūkojot šo izteiksmi un vidējā aritmētiskā standartkļūdas aprēķināšanas formulu varam secināt, ka ticamības intervāls saīsinās t.i. vērtējuma precizitāte paaugstinās pieaugot novērojumu skaitam t.i. paraugkopas apjomam. Standartnovirzes, variācijas koeficienta un citu rādītāju ticamības intervāla robežu aprēķināšanai izmantojamās metodes šeit neapskatīsim, tas var iepazīt izmantojot speciālo literatūru.

4. PARAUGKOPAS STATISTISKĀ ANALĪZE

4.1. Kvantitatîvas paraugkopas analîze

Piemērs. Aprēķināt LVFKI 2. kursa studentu - vīriešu vienas mācību grupas vidējo rezultātu tāllēkšanā, novērtēt rezultātu izkliedi (27. tabula). Novērtēt visu 2. kursa studentu - vīriešu vidējo rezultātu.

6. tabula

Darba tabula

x_i	x_i^2	x_i^3	x_i^4
5	25	125	625
5,05	25,5025	128,787625	650,3775063
5,06	25,6036	129,554216	655,544333
5,1	26,01	132,651	676,5201
5,22	27,2484	142,236648	742,4753026
5,29	27,9841	148,035889	783,1098528
5,42	29,3764	159,220088	862,972877
5,43	29,4849	160,103007	869,359328
5,5	30,25	166,375	915,0625
5,62	31,5844	177,504328	997,5743234
5,7	32,49	185,193	1055,6001
5,7	32,49	185,193	1055,6001
5,74	32,9476	189,119224	1085,544346
5,8	33,64	195,112	1131,6496
6	36	216	1296
6,04	36,4816	220,348864	1330,907139
6,05	36,6025	221,445125	1339,743006
6,15	37,8225	232,608375	1430,541506
6,27	39,3129	246,491883	1545,504106
$\sum x_i =$ =106,14	$\sum x_i^2 =$ =595,8314	$\sum x_i^3 =$ =3360,979272	$\sum x_i^4 =$ =19049,08602614

Analizējot uzdevuma saturu, varam secināt, ka dotajā situācija ģenerālkopas apjoms ir ierobežots, jo visi iespējamie vienveidīgie pētījuma objekti ir LSPA 2. kursa studenti - vīrieši. Tādējādi novērtējamais ģenerālkopas rādītājs - vidējais ģenerālais ir visu LSPA 2.kursa studentu vidējais rezultāts. Sacensību rezultātus ierakstām darba tabulas (6.tabula) 1. stabiņā, aprēķinām un ierakstām atbilstošajos stabiņos varianšu kvadrātus, kubus un ceturtās pakāpes, bet pēc tam - summas katrā stabiņā - $\sum x_i$; $\sum x_i^2$; $\sum x_i^3$; $\sum x_i^4$. Ar darba tabulas palīdzību iegūtās summas izmanto turpmākiem aprēķiniem pēc formulām.

1) Izmantojot formulu (1.8) aprēķina vidējo aritmētisko:

$$\bar{x} = \frac{106{,}14}{19} = 5{,}5863157 \approx 5{,}59$$

2) Aprēķina standartnovirzi (1.14):

$$\sigma = \sqrt{\frac{595 - \frac{(106{,}14)^2}{19}}{18}} = 0{,}4013776 \approx 0{,}40$$

3) Aprēķina variācijas koeficientu (1.15):

$$s\% = \frac{0{,}401}{5{,}586} \cdot 100\% = 7{,}178661 \approx 7{,}2\%$$

4) Aprēķina vidējā aritmētiskā standartkļūdu (1.19):

$$S_{\bar{x}} = \frac{0{,}401}{\sqrt{19}} = 0{,}09199571 \approx 0{,}09$$

5) Aprēķina vidējā aritmētiskā ticamības intervāla robežas (1.23):

$$5{,}586 - 0{,}092 \cdot 2{,}101 \le \mu \le 5{,}586 + 0{,}092 \cdot 2{,}101$$

$$5{,}39 \le \mu \le 5{,}78$$

6) Lai būtu iespēja novērtēt empīriskā sadalījuma atbilstību normālajam sadalījumam, aprēķina asimetrijas rādītāju, izmantojot formulas (1.17) darba variantu:

$$A = \frac{\sum x_i^3 - \frac{3\sum x_i \cdot \sum x_i^2}{n} + \frac{2(\sum x_i)^3}{n^2}}{ns^3} \qquad (4.24)$$

$$A = \frac{3360{,}98 - \frac{189719{,}09}{19} + \frac{2391482{,}71}{361}}{1{,}228607} = 0{,}063525 \approx 0{,}06$$

7) Izmantojot formulas (1.18) darba variantu, aprēķina ekscesa rādītāju, kurš arī izmantojams sadalījuma novērtēšanai:

$$E = \frac{\sum x_i^4 - \frac{4\sum x_i \sum x_i^3}{n} + \frac{6\sum x_i^2 (\sum x_i)^2}{n^2} - \frac{3(\sum x_i)^4}{n^3}}{ns^4} \qquad (4.25)$$

$$E = \frac{19049{,}09 - \frac{1426937{,}36}{19} + \frac{40274745{,}39}{361} - \frac{380747962{,}43}{6859}}{0{,}4931357} =$$

$$= 1{,}624322 \approx 1{,}6$$

Secinājumi:

1) LVFKI 2. kursa studentu - vīriešu mācību grupas vidējais rezultāts tāllēkšanā ir (5,59 ± 0,09) m.

2) Grupas sagatavotību tāllēkšanā var uzskatīt par vienveidīgu ($s\%$ < 10%).

3) LVFKI 2. kursa visu studentu - vīriešu vidējais rezultāts ir robežās no 5,39 līdz 5,78 m (P = 0,95).

4.2. Kvalitatīvas paraugkopas analīze

Analizējot kvalitatīvu paraugkopu, nosaka, kādai pētāmo objektu daļai piemīt vai nepiemīt mūs interesējošā pazīme (alternatīvas kopas) vai arī kāda paraugkopas daļa atbilst tai vai citai pētāmās pazīmes gradācijas pakāpei vai kvalifikācijas grupai (nealternatvās kopas). Šo paraugkopas daļu izteiktu ar decimāldaļskaitli vai procentos sauc par pazīmes īpatsvaru. Citiem vārdiem runājot, šeit ir darīšana ar procentu rēķiniem. Tipisks šādas analīzes veids ir aptaujas anketu apstrāde, kad tiek apkopotas daudzu respondentu domas par vieniem un tiem pašiem jautājumiem.
Alternatīvas kvalitatīvas paraugkopas pazīmes īpatsvars:

$$p = \frac{n_1}{n} \qquad (4.26)$$

kur n_1 - pazīmei atbilstošais novērojumu skaits;
n - kopējais novērojumu skaits jeb paraugkopas apjoms.
Pretējās pazīmes īpatsvars:

$$q = 1 - p \qquad (4.27)$$

Alternatīvas kvalitatīvas paraugkopas pazīmes standartnovirze:

$$s = \sqrt{p(1-p)} \qquad (4.28)$$

Nealternatīvas kvalitatīvas paraugkopas atsevišķas klases īpatsvaru aprēķina pēc formulas:

$$p_i = \frac{n_i}{n} \qquad (4.29)$$

kur $i = 1, 2, 3, \ldots\ldots, k$. Šai gadījumā k ir pazīmes gradācijas klasu skaits.

5. ATŠĶIRĪBU NOVĒRTĒŠANA

5.1. Nulles hipotēze un tās pārbaude

Pieņēmumu par kādu ģenerālkopas īpašību sauc par statistisku hipotēzi, kuru pārbauda sastādot un analizējot paraugkopu. Statistiskās hipotēzes izvirza par atsevišķu kopas parametru vai par visu sadalījumu. Veicot novērojumu rezultātu apstrādi bieži jārisina šādi jautājumi:

1) vai variante, kura ievērojami atšķiras no pārējām, pieder pētāmai kopai vai arī jāatmet kā nederīga;

2) vai empīriskais sadalījums būtiski atšķiras no teorētiskā sadalījuma;

3) vai starp divām paraugkopām ir būtiskas atšķirības?

Minētos jautājumus risina pārbaudot nulles hipotēzi. Tā ir pieņēmums, ka divu ģenerālkopu rādītāju starpība ir nulle t.i., bezgalīgi palielinot salīdzināmo paraugkopu apjomus, iegūst vienu un to pašu ģenerālkopu. Pārbaudes rezultātā nulles hipotēzi pieņem vai noraida. Lēmumu pieņem nevis kā absolūtu patiesību, bet gan ar izvēlēto ticamības līmeni (mūsu vajadzībām P = 0,95). Tātad 5% gadījumu iespējama kļūda. Attiecībā uz nulles hipitēzes pārbaudi iespējamas divu veidu kļūdas. Ja hipotēze ir pareiza, bet statistiskās pārbaudes rezultātā to noraida, sastopamies ar pirmā veida kļūdu. Ja hipotēze nepareiza, bet statistiskās pārbaudes rezultātā to pieņem, radusies otrā veida kļūda. Nulles hipotēzi būtu jāformulē tā, lai nepareiza lēmuma gadījumā tiktu pieļauta 1.veida kļūda, kas ir mazāk bīstami - noraidīt pareizu hipotēzi un nevis pieņemt nepareizu.

Pārbaudi veic, izmantojot kādu parametrisku vai neparametrisku metodi, kuras ideja ir sekojoša. Salīdzinot divu paraugkopu rādītājus, to starpība vairāk vai mazāk atšķiras no nulles arī tad, ja kopas pieder vienai ģenerālkopai. Novērojamās starpības cēlonis šai gadījumā ir pazīmes variēšana - starpībai ir gadījuma jeb nejaušības raksturs. Turpretī, ja salīdzināmās paraugkopas ir sastādītas no dažādām ģenerālkopām, tad starpība ir ticama. Lai to novērtētu starpību standartizē - ar skaitļošanas operāciju palīdzību atbrīvojas no mērvienības un salīdzina ar pieņemtajam ticamības līmenim pieļaujamo starpību. Atšķirību kritērijs ir nepieciešamā un pietiekamā starpība, lai noraidītu nulles hipotēzi. Empīrisko starpību uzskata par ticamu, ja tās vērtība lielāka par pieļaujamo - tad nulles hipotēzi noraida. Pretējā gadījumā nulles hipotēze paliek spēkā. Pieļaujamo jeb kritisko starpību - kritērija teorētisko vērtību nolasa speciālā tabulā atkarībā no paraugkopas apjoma un ticamības līmeņa. Šāda tabula sastādīta katrai metodei un atrodama attiecīgās statistikas mācību grāmatas vai rokasgrāmatas pielikumā. Metodes, ko izmanto nulles hipotēzes pārbaudei, parasti sauc par atšķirību kritērijiem. Ja novērtē nulles hipotēzi par divu kopu parametriem - statistiskiem rādītājiem, tad tie ir parametriski kritēriji. Tie pamatojas uz normālā sadalījuma likumu, tādēļ to izmantošanas iespējas ir ierobežotas. Metodes ar kurām novērtē atšķirības kopumā sauc par neparametriskajiem kritērijiem. To izmantošanai nav ierobežojumu. Ja empīriskais sadalījums neatšķiras no normālā sadalījuma, uzdevuma risināšanai var izmantot kā parametriskas tā neparametriskas metodes. Šai sakarībā jāatceras, ka skaitļošanas operāciju darbietilpības ziņā neparametriskie kritēriji ir vienkāršāki. Turpretī ar parametrisku metodi iegūts slēdziens ir precīzāks, jo parametrisko kritēriju izšķiršanas spēja jeb jūtība ir lielāka. Var gadīties, ka neparametriskais kritērijs nulles hipotēzi pieņem, bet izmantojot šai pašā uzdevumā parametrisko metodi izrādās, ka tā noraidāma. Ņemot vērā šīs īpatnības jāapsver, ar kādām metodēm visracionālāk veikt nulles hipotēzes pārbaudi.

5.1.1. Ekstremālo rezultātu pārbaude

Ekstremālās variantes ir x_{min} un x_{max}, kuras var ievērojami atšķirties no pārējām. Ja šādas novirzes cēlonis ir rupja mērīšanas kļūda vai arī nepiederoša objekta klātbūtne apsekotajā grupā, variante jāatmet kā nederīga. "Izlecošās" variantes pārbauda galvenokārt maza apjoma paraugkopām (n<30). Ja novērojumu skaits ir liels, atsevišķa apšaubāma rezultāta ietekme uz statistiskajiem rādītājiem ir niecīga. Variantes piederību mazai paraugkopai var pārbaudīt ar Diksona kritēriju. Vispirms variantes ranžē, iegūstot virkni: $x_1, x_2, x_3,..., x_{n-1}, x_n$, kur $x_1 = x_{min}$ un $x_n = x_{max}$. Pārbaudot minimālo varianti, kritērija empīrisko vērtību aprēķina pēc formulas:

$$\tau'' = \frac{x_2 - x_i}{x_{n-1} - x_i} \qquad (5.30)$$

Maksimālās variantes pārbaudei izmanto formulu:

$$\tau' = \frac{x_n - x_{n-1}}{x_n - x_2} \qquad (5.31)$$

Aprēķināto kritērija vērtību τ‘ vai τ" salīdzina ar teorētisko vērtību τ_α, kuru nolasa tabulā (10. pielikums) pēc paraugkopas apjoma *n* um būtiskuma līmeņa $\alpha = 0{,}05$. Ja $\tau > \tau_\alpha$, tad nulles hipotēzi noraida, tas nozīmē ka pārbaudamā variante pieder citai ģenerālkopai un jāatmet kā nederīga.

Piemērs. Doti LSPA 2. kursa studentu-vīriešu grupas rezultāti distanču slēpošanā - 10 km (min:s). Pārbaudīt, vai visi rezultāti izmantojami novērtējot dotās grupas studentu vidējo sagatavotības līmeni. Uzrakstām rezultātus ranžētā rindā: 38:08, 42:04, 42:09, 43:09, 43:36, 44:03, 44:19, 45:03, 45:06, 45:25, 45:40, 46:11, 46:59, 47:20, 47:23, 48:11, 49:20, 49:25, 49:26, 49:31, 51:57

Pārbaudām labāko rezultātu - 38:08.

$$\tau' = \frac{42{:}04 - 38{:}08}{49{:}31 - 38{:}08} = \frac{3{:}56}{11{:}23} = \frac{236}{683} \approx 0{,}346$$

Tabulā atrodam $\tau_{0,05} = 0{,}327$. Aprēķinātais $\tau' = 0{,}346$, bet $\tau_{0,05} = 0{,}327$, tātad rezultāts jāatmet. Pārbaudām nākošo labāko rezultātu - 42:04 (izsvītrojot no virknes 38:08, paraugkopas apjoms kļūst *n* - 1).

$$\tau'' = \frac{42{:}09 - 42{:}04}{49{:}31 - 42{:}04} = \frac{0{:}06}{7{:}27} = \frac{5}{447} \approx 0{,}011$$

Tabulā atrodam $\tau_{0,05} = 0{,}334$. $\tau' = 0{,}011 < \tau_{0,05} = 0{,}334$, tātad rezultāts 42:04 un līdz ar to tam sekojošie pieder kopai. Pārbaudām vājāko rezultātu 51:57 :

$$\tau'' = \frac{51{:}57 - 49{:}31}{51{:}57 - 42{:}09} = \frac{2{:}26}{9{:}48} = \frac{146}{588} \approx 0{,}248$$

$\tau'' = 0{,}248 < \tau_{0,05} = 0{,}334$, tātad rezultāts 51:57 pieder kopai.

Secinājums. Rezultāts 38:08 būtiski atšķiras no pārējiem (P = 0,95), tādēļ to nevar izmantot novērtējot grupas vidējo sagatavotības līmeni šai slēpošanas distancē. Visticamāk, ka šo rezultātu sasniedzis students, kurš atšķirībā no pārējiem specializējas slēpošanā.

5.1.2. Empīriskā un teoretiskā sadaîījuma atbilsîības pārbaude

Statistiskās metodes, kas pamatojas uz normālā sadalījuma likuma var izmantot tikai tad, ja pierādīta empīriskā sadalījuma atbilstība normālajam. Nulles hipotēzi par empīriskā un teorētiskā - normālā sadalījuma atšķirībām var pārbaudīt ar vairākām metodēm. Vienkāršākā no tām izmanto asimetrijas (A) un ekscesa (E) rādītājus, aprēķinot normēto novirzi:

$$t_A = \frac{A}{s_A} \tag{5.32}$$

$$t_E = \frac{E}{s_E} \tag{5.33}$$

kur s_A un s_E ir attiecīgi asimetrijas un ekscesa rādītāju standartkļūdas. Ja abos gadījumos $t < 3$, nulles hipotēzi par empīriskā un normālā sadalījuma atšķirībām pieņem. Izmantojot iepriekšējā nodaļā veiktās viendimensijas paraugkopas analīzes datus, pārbaudīsim šīs kopas atbilstību normālajam sadalījumam. Iepriekš aprēķinātie $A = 0{,}06$ un $E = 1{,}62$, paraugkopas apjoms $n = 19$. Aprēķināsim asimetrijas un ekscesa rādītāju standartkļūdas s_A un s_E .

$$s_A = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{5.34}$$

$$s_E = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \tag{5.35}$$

$$s_A = \sqrt{\frac{6 \cdot 19 \cdot 18}{17 \cdot 20 \cdot 22}} \approx 0{,}5237$$

$$s_E = \sqrt{\frac{24 \cdot 19 \cdot 18^2}{16 \cdot 17 \cdot 22 \cdot 24}} = 1{,}0142$$

$$t_A = \frac{0{,}06}{0{,}52} \approx 0{,}121$$

$$t_E = \frac{1{,}62}{1{,}01} \approx 1{,}601$$

$t_A = 0{,}121 < 3$ un $t_E = 1.601 < 3$, tādēļ nav pamata uzskatīt, ka dotās paraugkopas sadalījums atšķiras no normālā.

Mazliet sarežģītāka, bet precīzāka metode ir Šapiro-Vilkija *W*-kritērijs paraugkopām, kuru apjoms nepārsniedz 50 variantes. Sastāda darba tabulu, kuras 2. un 3. stabiņā attiecīgi ieraksta ranžētās kopas variantes x_i un to kvadrātus x_i^2 (7. tabula), katrā stabiņā izskaitļo attiecīgās summas: Σx_i un Σx_i^2, kas vajadzīgas statistisko rādītāju aprēķināšanai.

7. tabula

Palīglielumu aprēķināšana *W*-kritērijam

Nr. p.k.	Variante x_i	x_i^2	Variansu starpība d_k	Starpības koeficients $a_{n;k}$	Reizinājums $d_k a_{n;k}$
1	x_1	x_1^2	$d_1 = x_n - x_1$	$a_{n;1}$	$d_1 a_{n;1}$
2	x_2	x_2^2	$d_2 = x_{n-1} - x_2$	$a_{n;2}$	$d_2 a_{n;2}$
3	x_3	x_3^2	$d_3 = x_5 - x_3$	$a_{n;3}$	$d_3 a_{n;3}$
4	x_4	x_4^2			$\Sigma d_k a_{n;k}$
5	x_5	x_5^2			
6	x_6	x_6^2			
7	x_7	x_7^2			
	Σx_i	Σx_i^2			

Aprēķina variansu pāru starpības d_k , no rindas pēdējās variantes atņemot pirmo, no priekšpēdējās otro u.t.t. Starpības ieraksta tabulas 4. stabiņā. Ja paraugkopas apjoms ir pāru skaitlis tad starpību skaits $k = n/2$, bet, ja nepāru - $k = (n - 1)/2$ (centrālo varianti starpību aprēķināšanai neizmanto). Tabulā (8. pielikums) katrai starpībai atbilstoši paraugkopas apjomam un starpības numuram atrod koeficientu $a_{n;k}$ un ieraksta darba tabulas 5. stabiņā. Aprēķina starpību un tām atbilstošo koeficientu reizinājumus $d_k a_{n;k}$, kurus ieraksta darba tabulas 6. stabiņā. Aprēķina reizinājumu summu $\Sigma d_k a_{n;k}$. W-kritērija empīrisko vērtību aprēķina izmantojot formulu:

$$W = \frac{\left(\sum d_k a_{n;k}\right)^2}{s^2(n-1)} \tag{5.36}$$

kur s^2 - paraugkopas dispersija;
n - paraugkopas apjoms.

Aprēķināto kritērija vērtību W salīdzina ar teorētisko W_α , kuru atrod tabulā (9. pielikums) pēc paraugkopas apjoma un būtiskuma līmeņa $\alpha = 0{,}05$. Ja $W > W_\alpha$, nulles hipotēze paliek spēkā t.i. empīriskā un normālā sadalījuma atšķirībām ir gadījuma raksturs. Pārbaudot lielu paraugkopu, W-kritērija vietā var izmanto citas metodes.

Piemērs. Doti Murjāņu sporta ģimnāzijas jauno riteņbraucēju aerobo darba spēju rādītāji - maksimālais skābekļa patēriņa līmenis (ml.min^{-1}.kg^{-1}). Novērtēt rezultātu atbilstību normālajam sadalījumam. 29. tabulā aprēķina starprezultātus - summas:

8. tabula

Darba tabula

Nr. p.k.	Variante x_i	x_i^2	Variančšu starpība d_k	Starpības koeficients $a_{n;k}$	Reizinājum s $d_k a_{n;k}$
1	57,7	3329,29	12,8	0,5601	7,16928
2	60,0	3600,00	9,0	0,3315	2,9835
3	61,4	3769,96	5,6	0,2260	1,2556
4	62,3	3881,29	3,1	0,1429	0,44299
5	63,2	3994,24	1,3	0,0695	0,09035
6	63,4	4019,56		$\sum d_k a_{n;k} =$	11,95172
7	64,5	4160,25			
8	65,5	4277,16			
9	67,0	4489,00			
10	69,0	4761,00			
11	70,5	4970,25			
	$\sum x_i =$ = 704,4	$\sum x_i^2 =$ = 45252,0			

$$s^2(n - 1) = 45252 - \frac{(704{,}4)^2}{11} = 144{,}786$$

$$W = \frac{(11{,}95172)^2}{144{,}786} = 0{,}9865844 \approx 0{,}987$$

Tabulā (9. pielikums) pēc $n = 11$ un $\alpha = 0{,}05$ atrodam $W_\alpha = 0{,}850$. $W = 0{,}987 > W_{0,05}$, tātad nulles hipotēze paliek spēkā. **Secinājums.** Nulles hipotēze par empīriskā un normālā sadalījuma atšķirībām paliek spēkā. Jauno riteņbraucēju aerobo darba spēju rādītāji (maksimālais skābekļa patēriņa līmenis) variē atbilstoši normālajam likumam. Tāpēc rezultātu analīzē drīkst izmantot parametriskās metodes, kuras pamatojas uz šo likumu.

5.1.3. Saistītu paraugkopu atšķirību novērtēšana ar Stjūdenta t-kritēriju

Bieži sastopams statistiskās analīzes uzdevums ir nulles hipotēzes pārbaude par paraugkopu vidējiem aritmētiskiem. No parametriskām metodēm izmanto Stjūdenta t-kritērija divus variantus - saistītām un neatkarīgām paraugkopām. Par saistītām sauc paraugkopas, kas iegūtas apsekojot vienu pētāmo objektu grupu atkārtoti pēc noteikta laika intervāla. Piemēram, piecas saistītas paraugkopas ir vienas un tās pašas studentu grupas sasniegumi kādā testā iestājeksāmenos, 1., 2., 3. un 4. kursā. Neatkarīgas kopas veido variantes, kas iegūtas apsekojot divas vai vairākas pētāmo objektu grupas (piemēram, 3^A. un 3^B. klases zēnu sasniegumi bumbiņas mešanā).

Uzmanību! Turpmāk aprakstītos Stjūdenta kritērija un Fišera kritērijus drīkst izmantot, ja pierādīta novērojumu rezultātu atbilstība normālajam sadalījumam. Pretējā gadījumā jāizvēlas kāda neparametriska metode. Sportistu grupas vidējā sacensību vai testa rezultāta izmaiņu ticamību noteiktā laika periodā (no mēģinājuma uz mēģinājumu, dažādos treniņa periodos u.t.t.) novērtē, izmantojot Stjūdenta kritēriju saistītām kopām. Tabulas (9. tabula) 1. ailē ieraksta beigu rezultātu $x_{2.i}$, 2. ailē - sākuma rezultātu $x_{1.i}$, 3. ailē - rezultātu starpību - d_i , 4. ailē starpības kvadrātu. Aprēķina summas Σd_i ; Σd_i^2.

9. tabula

Palīglielumu aprēķināšana Stjūdenta *t*-kritērijam

Gala rezultāts $x_{2.i}$	Sākuma rezultāts $x_{1.i}$	Starpība $d_i = x_{2.i} - x_{1.i}$	d_i^2
$x_{2.1}$	$x_{1.1}$	d_1	d_1^2
$x_{2.2}$	$x_{1.2}$	d_2	d_2^2
$x_{2.3}$	$x_{1.3}$	d_3	d_3^2
$x_{2.n}$	$x_{1.n}$	d_n	d_n^2
		Σd_i	Σd_i^2

Aprēķina vidējo starpību:

$$\bar{d} = \frac{\sum d_i}{n} \qquad (5.37)$$

kur n - rezultātu pāru skaits.
Aprēķina vidējās starpības standartkļūdu:

$$s_{\bar{d}} = \sqrt{\frac{\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n(n-1)}} \quad (5.38)$$

Aprēķina Stjūdenta kritērija empīrisko vērtību:

$$t = \frac{|\bar{d}|}{s_{\bar{d}}} \quad (5.39)$$

Stjūdenta tabulā (1. pielikums) pēc $\alpha = 0{,}05$ un $\nu = n - 1$ nolasa Stjūdenta sadalījuma normēto novirzi $t_{\alpha;\nu}$. Ja $t \geq t_{\alpha;\nu}$, tad pie pieņemtā rezultātu ticamības līmeņa rezultātu izmaiņas ir neapstrīdamas - būtiskas. Ja $t < t_{\alpha;\nu}$, tad rezultātu starpībai ir gadījuma raksturs - tā radusies pazīmes dabīgās variēšanas rezultātā.

Piemērs. Doti 22. Olimpisko spēļu vieglatlētikas sacensību finālistu rezultāti 400 m skrējienā sievietēm. Novērtēt rezultātu izmaiņas finālā, salīdzinot ar pusfinālu. Aprēķina palīglielumus izpildot darba tabulu.

10. tabula

Darba tabula

Sportiste	Komanda	Rezultāts finālā	Rezultāts pusfinālā	d_i	d_i^2
Koha	VDR	48,88	50,57	-1,69	2,85671
Kratohvilova	ČSR	49,46	50,79	-1,33	1,7689
Latana	VDR	49,66	50,16	-0,50	0,25
Nazarova	PSRS	50,07	50,18	-0,11	0,0121
Zjuskova	PSRS	50,17	51,12	-0,95	0,9025
Leve	VDR	51,33	50,85	0,48	0,2304
Hegmane	Somija	51,35	51,02	0,33	0,1089
Makdonalda	L-brit.	52,40	51,60	0,80	0,6400
				-2,97	6,7689

$$\bar{d} = \frac{-2{,}97}{8} = -0{,}37125 \approx -0{,}37$$

$$s_{\bar{d}} = \sqrt{\frac{6{,}7689 - \frac{(-2{,}97)^2}{8}}{8(8-1)}} = -0{,}31809385 \approx 0{,}318$$

$$t = \frac{|-0{,}371|}{0{,}318} = 1{,}1671083 \approx 1{,}167$$

$$t_{0,05;7} = 2{,}365$$

$$t < t_{0,05;7}$$

Secinājums: Rezultātu izmaiņas 400 m skrējiena finālā salīdzinot ar pusfinālu nav statistiski ticamas ($\alpha > 0{,}05$).

5.1.4. Saistītu paraugkopu atšķirību novērtēšana ar Vilkoksona kritēriju

Pārbaude ar Vilkoksona kritēriju ir neparametriska metode. To izmanto saistītu paraugkopu salīdzināšanai, ja nedrīkst lietot Stjūdenta kritēriju (novērojumu rezultāti neatbilst normālajam sadalījumam). Darba tabulas (11. tabula) 1. ailē ieraksta beigu rezultātu $x_{2.i}$, 2. ailē - sākuma rezultātu $x_{1.i}$, 3. ailē - rezultātu starpību - d_i , 4. ailē katrai starpībai pēc absolūtās vērtības piešķir rangu g_i , kuram pieraksta attiecīgās starpības zīmi. Atsevišķi aprēķina pozitīvo rangu summu $T_{(+)} = \Sigma g_{(+)}$ un negatīvo rangu summu $T_{(-)} = \Sigma g_{(+)}$. Pēc absolūtās vērtības mazāko no šīm summām izmanto kā kritērija empīrisko vērtību.

11. tabula

Palīglielumu aprēķināšana Vilkoksona kritērijam

Gala rezultāts $x_{2.i}$	Sākuma rezultāts $x_{1.i}$	Starpība $d_i = x_{2.i} - x_{1.i}$	$\|d_i\|$ rangs g_i
$x_{2.1}$	$x_{1.1}$	d_1	g_1
$x_{2.2}$	$x_{1.2}$	d_2	g_2
$x_{2.3}$	$x_{1.3}$	d_3	g_3
$x_{2.n}$	$x_{1.n}$	d_n	g_n
			Σg_i

Speciālā tabulā (2. pielikums) pēc rezultātu pāru skaita n (ja rezultātu starpība ir nulle, pāri neskaita) un $\alpha = 0{,}05$ nolasa $T_{\alpha;n}$ - kritērija robežvērtību. Ja $T \leq T_{\alpha;n}$ rezultātu izmaiņas ir būtiskas, pretējā gadījumā ($T > T_{\alpha;n}$) atšķirībām ir gadījuma raksturs.

Piemērs. Doti LSPA studentu - sieviešu grupas rezultāti tāllēkšanā 1. kursā un pēc gada 2. kursā. Novērtēt rezultātu izmaiņas. Sastādam darba tabulu (12. tabula) un izpildām pārējos aprēķinus.

12. tabula

Darba tabula

Rezultāts 2. kursā $x_{2.i}$	Rezultāts 1. kursā $x_{1.i}$	Rezultātu starpība $d_i = x_{2.i} - x_{1.i}$	$\lvert d_i \rvert$ rangs g_i
5,29	5,31	-0,02	-2
3,85	3,79	0,06	2
4,2	4,31	-0,11	-7
3,86	3,60	0,26	10
4,15	3,68	0,47	13
4,19	4,28	-0,09	-5
3,88	3,80	0,08	3,5
3,99	3,74	0,24	9
3,83	3,83	0	-
4,45	4,35	0,10	6
4,20	4,28	-0,08	-3,5
3,93	3,75	0,18	8
4,30	3,85	0,45	12
4,31	4,00	0,31	11

$\Sigma d_i = 1{,}85$

$$T_{(-)} = |\Sigma g_{(-)}| = 16{,}5$$

$$T_{(+)} = |\Sigma g_{(+)}| = 74{,}5$$

$$\bar{d} = \frac{1{,}85}{14} \approx 0{,}13$$

Tabulā (2. pielikums) pēc $n = 13$ un $\alpha = 0{,}05$ atrodam $T_{0,05;15} = 17$.

$T_{(-)} < T_{0,05;15}$, tātad rezultātu izmaiņas ir būtiskas.

Secinājums. LVFKI studentu - sieviešu grupas rezultāti tāllēkšanā gada laikā vidēji uzlabojušies par 0,13 m. Šīs izmaiņas ir statistiski ticamas (P = 0,95).

5.1.5. Neatkarīgu paraugkopu salīdzināšana ar Stjūdenta kritēriju

Šeit būs runa par to, kā novērtēt dažādu sportistu grupu vidējā sagatavotības līmeņa atšķirības. **Šo paņēmienu izmanto salīdzinot kontroles un eksperimentālās grupas rezultātus pedagoģiskā eksperimentā.** Vispirms veic abu salīdzināmo paraugkopu statistisko analīzi, aprēķinot vidējos aritmētiskos, standartnovirzes u.c. Pēc tam aprēķina Stjūdenta kritērija empīrisko vērtību:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{\bar{x}_1 - \bar{x}_2}} \quad (5.40),$$

`kur $\bar{x}_1$ un $\bar{x}_2$ - attiecīgi pirmās un otrās paraugkopas vidējie aritmētiskie; $s_{\bar{x}_1 - \bar{x}_2}$ - vidējo aritmētisko starpības standartkļūda. Šī lieluma aprēķināšanas formulu izvēlas atkarībā no novērojumu skaita. Ja mazās paraugkopās (n < 30) varianšu skaits dažāds ($n_1 \neq n_2$), tad vidējo aritmētisko starpības standartkļūdu aprēķina pēc formulas:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}} \quad (5.41)$$

Pārējos gadījumos, salīdzinot kā lielas tā mazas paraugkopas, var izmantot formulu:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{s_{\bar{x}_1} + s_{\bar{x}_2}} \quad (5.42)$$

Stjūdenta tabulā (1.pielikums) pēc α = 0,05 un $\nu = n_1 + n_2 - 2$ nolasa kritērija robežvērtību $t_{\alpha;\nu}$. Ja $t \geq t_{\alpha;\nu}$, tad atsķirības ir būtiskas - statistiski ticamas, turpretī pie $t < t_{\alpha;\nu}$ rezultātu atšķirības nevar uzskatīt par pierādītām, jāpieņem, ka tām ir gadījuma raksturs.

Piemērs. Doti šosejas riteņbraucēju - Latvijas 1974. un 1978. gada izlases komandu dalībnieku aerobo darba spēju rādītāji (maksimālais O_2 -patēriņa līmenis, ml.min^{-1}.kg^{-1}) treniņa sagatavošanās periodā (sportistu kvalifikācija atbilstoši tā laika klasifikācijai: sporta meistars, starptautiskās klases sporta meistars, meistarkandidāts). Novērtēt abu komandu aerobo darba spēju vidējo lielumu atšķirību.

13. tabula

Darba tabula

1974. gada izlases komanda			1978. gada izlases komanda		
Sportists	Kvalifikācija	$_{max}VO_2$ $ml.min^{-1}.kg^{-1}$	Sportists	Kvalifikācija	$_{max}VO_2$ $ml.min^{-1}.kg^{-1}$
V.G.	SKSM	77,1	J.K.	SM	79,1
R.K.	SKSM	72,2	A.J.	SKSM	80,5
V.F.	SM	75,1	A.R.	SM	75,0
V.B.	SM	73,4	V.M.	SM	75,4
A.N.	SM	71,1	V.F.	SM	84,8
A.S.	SM	74,8	A.B.	SM	83,4
A.E.	SKSM	40,8	A.A.	SMK	75,0
A.B.	SMK	70,5			

Veicot abu paraugkopu analīzi iegūti šādi raksturojumi:

$\bar{x}_1 = 73{,}12$ $\bar{x}_2 = 79{,}03$

$s_1^2 = 5{,}69$ $s_2^2 = 16{,}70$

$s_1\% = 3{,}3\%$ $s_2\% = 5{,}2$

$s_{\bar{x}_1} = 2{,}4$ $s_{\bar{x}_2} = 1{,}5$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{5{,}69 \cdot 7 + 16{,}7 \cdot 6}{8 + 7 - 2} \cdot \frac{8 + 7}{8 \cdot 7}} \approx 1{,}70$$

$$t = \frac{|73{,}12 - 79{,}03|}{1{,}70} = 3{,}47647 \approx 3{,}46$$

$$t_{0,05;13} = 2{,}160$$

$t > t_{0,05;13}$, tātad vidējo rezulātu starpība ir būtiska.

Secinājums. 1978. gada Latvijas šosejas riteņbraucēju izlases komandas dalībnieku aerobās darba spējas salīdzinot ar 1974. gada komandu vidēji ir augstākas - maksimālā skābekļa patēriņa līmenis ir par 5,9 $ml.min^{-1}.kg^{-1}$ lielāks. Šīs atšķirības ir statistiski ticamas (P > 0,95).

5.1.6. Neatkarīgu paraugkopu saīdzināšana ar Van der Vardena kritēriju

Van der Vardena kritērijs ir viena no precīzākajām, bet arī darbietilpīgākajām neparametriskām metodēm. **To izmanto Stjūdenta kritērija vietā neatkarīgu paraugkopu salīdzināšanai, ja novērojumu rezultāti neatbilst normālajam sadalījumam.** Atrisināsim iepriekšējo uzdevumu par riteņbraucēju aerobajām darba spējām izmantojot šo metodi.

14. tabula

Darba tabula van der Vardena kritērija aprēķināšanai

1974. g. $_{max}VO_2$ ml.min^{-1}.kg^{-1} $x_{1.i}$	1978. g. $_{max}VO_2$ ml.min^{-1}.kg^{-1} $x_{2.i}$	g_i	$\frac{g_i}{n_1+n_2+1}$	$\Psi(\frac{g_i}{n_1+n_2+1})$
1	2	3	4	5
70.5		1		
70.8		2		
71.1		3		
72.2		4		
73.4		5		
74.8		6		
	75.0	7,5	0,46875	-0,08
	75.0	7,5	0,46875	-0,08
75.1		9		
	75.4	10	0.625	0,32
77.1		11		
	79.1	12	0.75	0,67
	80.5	13	0.8125	0,89
	83.4	14	0.875	1,15
	84.8	15	0.9375	1,535
				4,405

Abu paraugkopu variantes uzraksta vienā ranžētā rindā, bet katru kopu savā stabiņā - $x_{1.i}$ vai $x_{2.i}$ (35. tabula). 3. kolonā ieraksta varianšu rangus g_i (ievērojot ranžēšanas noteikumus). 4. kolonā **mazākās** kopas variantēm aprēķina $\frac{g_i}{n_1+n_2+1}$ (n_1, n_2 - paraugkopu apjomi). 11. pielikumā katram dalījumam nolasa $\Psi\left(\frac{g_i}{n_1+n_2+1}\right)$ vērtību un ieraksta 5. kolonā. Ja dalījums mazāks par 0,5, tad nolasa pēc $1-\dfrac{g_i}{n_1+n_2+1}$ un nolasīto vērtību raksta ar mīnus zīmi. Aprēķina summu 5. stabiņā - Van der Vardena kritērija empīrisko vērtību:

$$X=\sum\Psi\left(\frac{g_i}{n_1+n_2+1}\right)=4{,}405 \qquad (5.43)$$

Pēc $n_1+n_2=15$, $|n_1-n_2|=1$ un $\alpha=0{,}05$ 3. pielikuma tabulā nolasa van der Vardena kritērija teorētisko vērtību $X_\alpha=3{,}24$. Nulles hipotēzi pieņem, ja $X\leq X_\alpha$. Mūsu gadījumā $X>X_\alpha$, tātad - nulles hipotēze jānoraida - atšķirības ir ticamas.

5.1.7. Paraugkopu dispersiju salīdzināšana ar Fišera kritēriju

Lai novērtētu vai rezultātu izkliedes (blīvuma) atšķirības divās sportistu grupās, aprēķina kritērija empīrisko vērtību:

$$F=\frac{s_1^2}{s_2^2} \qquad (5.44)$$

kur s_1^2 un s_2^2 ir attiecīgi 1. un 2. paraugkopas dispersijas. Jāievēro nosacījums $s_1^2>s_2^2$, skaitītājā liekot lielāko vērtību. Tabulā (5. pielikums) pēc $\alpha=0{,}05$ un $\nu_1=n_1-1$; $\nu_2=n_2-1$ nolasa kritērija robežvērtību $F_{\alpha;\nu1\nu2}$. Dispersiju atšķirība ir būtiska, ja $F>F_{\alpha;\nu1\nu2}$.

Novērtēsim aerobo darba spēju testa rezultātu vienveidību abās, iepriekšējā uzdevumā minētās riteņbraucēju komandās, salīdzinot dispersijas: $s_1^2=5{,}69$; $s_2^2=16{,}70$.

$$F=\frac{16{,}7}{5{,}69}=2{,}9349736\approx 2{,}93$$

Tabulā nolasītā Fišera kritērija teorētiskā vērtība $F_{0{,}05;6;7}=3{,}87$.

Secinājums. Var uzskatīt, ka no funkcionālās sagatavotības vienveidības viedokļa starp abām komandām nav būtisku atšķirību ($\alpha<0{,}05$).

5.1.8. Kvalitatīvu paraugkopu salīdzināšana

Kvalitatīvu paraugkopu pazīmes īpatsvaru salīdzināšanai aprēķina Stjūdenta t-kritērija empīrisko vērtību:

$$t = \frac{\left| \frac{n_{1x}}{n_1} - \frac{n_{2x}}{n_2} \right|}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{|p_1 - p_2|}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (5.45),$$

kur p_1 un p_2 - pētāmās pazīmes īpatsvari attiecīgi pirmā un otrā paraugkopā;
n_1 un n_2 - pirmās un otrās paraugkopas apjoms;

$$p = \frac{n_{1x} + n_{2x}}{n_1 + n_2} \qquad (5.46);$$

n_{1x} un n_{2x} - pētāmai pazīmei atbilstošais novērojumu skaits pirmā un otrā paraugkopā. Rezultātu ticamības koeficientu t salīdzina ar Stjūdenta tabulā nolasīto $t_{\alpha;\infty}$, bet vienpusēja ierobežojuma testa pielietošanas gadījumā (pārbaudot vai pazīmes īpatsvars nepārsniedz kādu noteiktu nominālo vērtību) ar - $t_{2\alpha;\infty}$. Ja $t > t_{\alpha;\infty}$ vai $t > t_{2\alpha;\infty}$ nulles hipotēzi noraida - atšķirības ir statistiski ticamas. Pretējā gadījumā pieņem, ka pētāmās pazīmes īpatsvari būtiski neatšķiras.

5.2. Dispersijas analīze

Dispersijas analīzi izmanto faktoreksperimentā iegūto datu apstrādei un analīzei. Tā pamatojas uz skaidrojumu, ka pētāmās pazīmes variēšanu rada visu kontrolējamo un nekontrolējamo faktoru summārā ietekme. Analīzes uzdevums ir, noskaidrot, vai variējamais faktors būtiski ietekmē pētāmās pazīmes variēšanu un ja tā, tad noteikt tā ietekmes īpatsvaru. Atkarībā no vienlaikus pētāmo faktoru skaita izšķir viena, divu faktoru un daudzfaktoru dispersijas analīzi. Katru pētāmo faktoru dala vairākās gradācijas klasēs. Gradācijas klases var būt kvalitatīvas un kvantitatīvas.

5.2.1. Viena faktora dispersijas anaīīze

Piemērs. Pēc speciālas apmācības, kuras laikā tika apgūti trīs spēku sadales varianti 4 km individuālajā iedzīšanas braucienā, seši riteņbraucēji piedalījās faktoreksperimentā. Eksperimenta uzdevums bija, novērtēt taktikas varianta ietekmi uz sacensību rezultātu un noskaidrot optimālo variantu augsta rezultāta sasniegšanai. Tika salīdzinātas faktora **A** - "Taktikas (spēka sadales) variants" gradācijas klases: **A1** - mērens starts un vienmērīgs ātrums visā distancē; **A2** - ass starts un cenšanās saglabāt maksimāli iespējamo ātrumu līdz finišam; **A3** - ātrs starts, ātruma samazināšana un spēcīgs finišs. Šādi raksturotas faktora gradācijas klases sauc par kvalitatīvām (kaut gan katru no variantiem, var raksturot arī ar konkretu ātruma līkni). Atbilstoši eksperimenta plānam sportistus izlozes veidā iedalīja 3 grupās, pa 2 cilvēkiem katrā. Kontroles sacensības notika standarta apstākļos - slēgtā velotrekā, 3 dienas pēc kārtas, ik dienas vienā un tai pašā laikā. Veicot distanci vienu reizi dienā, katra grupa izmantoja savu spēku sadales variantu. Pirmajā dienā variants tika izlozēts, bet turpmāko secību noteica eksperimenta plāns (15. tabula). Uzdevums bija, ik reizi sasniegt maksimāli iespējamo rezultātu.

15. tabula

Viena faktora eksperimenta plāns

Faktora A “Spēku sadale distancē” **gradācijas klases**	Eksperimentālās grupas		
	X1	**X2**	**X3**
A1	1	3	2
A2	2	1	3
A3	3	2	1

Dispersijas analīzes procedūras sākas ar statistiskā kompleksa jeb darba tabulas (17. tabula) sastādīšanu atbilstoši šablonam (16. tabula).

16. tabula

Viena faktora statistiskais komplekss

Gradācijas klases	**Variantes x_{ij}**	n_i	S_i	$\bar{x}_i$
A_1	$x_{1.1}$ $x_{1.2}$ $x_{1.3}$ $x_{1.4}$ $x_{1.5}$ $x_{1.6}$	n_1	S_1	x_1
A_2	$x_{2.1}$ $x_{2.2}$ $x_{2.3}$ $x_{2.4}$ $x_{2.5}$ $x_{2.6}$	n_2	S_2	x_2
A_3	$x_{3.1}$ $x_{3.2}$ $x_{3.3}$ $x_{3.4}$ $x_{3.5}$ $x_{3.6}$	n_3	S_3	x_3

Darba tabulu var izmantot aprēķiniem ar kalkulātoru pēc zemāk aprakstītajām formulām, bet personālais dators un elektroniskā tabula EXCEL ļauj veikt dispersijas analīzi automātiski.

Visa kompleksa vidējais aritmētiskais:

$$\bar{x} = \frac{S}{N} = \frac{\sum S_i}{\sum n_i} \qquad (5.47)$$

Viena faktora statistiskajā kompleksā izšķir 3 veidu variācijas.

1. Visu varianšu izkliede ap statistiskā kompleksa vidējo aritmētisko neatkarīgi no piederības gradācijas klasēm.
2. Gradācijas klašu vidējo vērtību x_i izkliede ap kompleksa vidējo aritmētisko.

3. Dotās gradācijas klases varianšu izkliede ap gradācijas klases vidējo vērtību.

Katra izkliedes veida raksturošanai aprēķina tam atbilstošo noviržu kvadrātu summas un vidējos noviržu kvadrātus.

Visa kompleksa noviržu kvadrātu summa:

$$Q = \sum x_{ij}^2 - \frac{S^2}{N} \tag{5.48}$$

Faktoram A atbilstošo noviržu kvadrātu summa:

$$Q_A = \sum \frac{S_i^2}{n_i} - \frac{S^2}{N}$$

Nekontrolēto faktoru jeb atlikuma noviržu suma:

$$Q_Z = Q - Q_A \tag{5.49}$$

Aprēķina brīvības pakāpju skaitus.

Kopējais brīvības pakāpju skaits:

$$\nu = N - 1$$

Faktora A brīvības pakāpju skaits:

$$\nu_A = k - 1$$

Atlikuma brīvības pakāpju skaits:

$$\nu_Z = \nu - \nu_A$$

Atlikuma brīvības pakāpju skaits nedrīkst būt mazāks par 10. Šis noteikums jāievēro plānojot eksperimentu - tas limitē minimālo eksperimentālo personu skaitu.

Vidējos noviržu kvadrātus jeb dispersijas aprēķina, dalot noviržu kvadrātu summas ar attiecīgo brīvības pakāpju skaitu.

Kopējā dispersija:

$$\overline{Q} = \frac{Q}{\nu} \tag{5.50}$$

Faktora A dispersija:

$$\overline{Q}_A = \frac{Q_A}{\nu_A} \tag{5.51}$$

Atlikuma dispersija:

$$\overline{Q}_Z = \frac{Q_Z}{\nu_Z} \tag{5.52}$$

Aprēķinātās kvadrātu summas, brīvības pakāpju skaitus un vidējos kvadrātus ieraksta darba tabulā (17. tabula).
Faktora ietekmes īpatsvaru var noteikt ar vairākām metodēm. Vienkāršākā no tām - noviržu kvadrātu summu attiecību metode.
Faktora A ietekmes īpatsvars:

$$\eta_A^2 = \frac{Q_A}{Q} \tag{5.53}$$

Atlikuma īpatsvars:

$$\eta_Z^2 = \frac{Q_Z}{Q} \tag{5.54}$$

Lai faktoru īpatsvarus izteiktu procentos, šādi izskaitļotos lielumus reizina ar 100. Tādējādi faktora īpatsvars var būt intervālā no 0 līdz 100%. Jo lielāks īpatsvars, jo lielāka šī faktora ietekme (ja pārbaudot tā atzīta par ticamu).
Dotajā eksperimentā sacensību rezultāti veidoja sekojošu statistisku kompleksu (17. tabula).

17. tabula

4 km rezultātu dispersijas analīzes statistiskais komplekss

Spēku sadales varianti	Variantes - rezultāti 4 km, s x_i	n_i	S_i	$\overline{x}_i$
A1	292.1 290.0 291.0 281.5 291.0 290.0	6	1735.6	289.27
A2	305.0 300.0 305.0 295.0 300.0 301.0	6	1806.0	301.00
A3	293.5 293.0 294.0 285.0 294.0 293.0	6	1752.5	292.08
		18	5294.1	
	$\Sigma x_{ij}^2=1557740$	N	S_{ij}	

Variančšu summas S_i katram spēku sadales režīmam aprēķinātas pa horizontāli, kopējais kompleksa variančšu skaits N un variančšu summa S_{ij} - pa vertikāli. Izskaitļota visu kompleksa variančšu kvadrātu summa Σx_{ij}^2 un katrā režīmā sasniegtie vidējie rezultāti $\bar{x}_i$.

Visa kompleksa vidējais aritmētiskais:

$$\bar{x} = \frac{1735.6 + 1806.0 + 1752.5}{6 + 6 + 6} = 294.1167$$

Noviržu kvadrātu summas:

$$Q = 1557740 - \frac{5294.1^2}{18} = 656.22$$

$$Q_A = \frac{1735.6^2 + 1806^2 + 1752.5^2}{6} - \frac{5294.1^2}{18} = 450.2233$$

$$Q_Z = 656.22 - 450.2233 = 206.6417$$

Vidējie kvadrāti:

$$\overline{Q} = \frac{Q}{\nu} = \frac{656.86}{18 - 1} = 38.63912$$

$$\overline{Q}_A = \frac{Q_A}{\nu_A} = \frac{450.22}{3 - 1} = 225.1117$$

$$\overline{Q}_Z = \frac{Q_Z}{\nu_Z} = \frac{206.6417}{17 - 2} = 13.77611$$

Faktoru ietekmes īpatsvari:

$$\eta_A^2 = \frac{450.865}{656.865} \cdot 100\% = 68.54123\%$$

$$\eta_Z^2 = \frac{206.6417}{656.865} \cdot 100\% = 31.45877\%$$

Faktora A ietekmes īpatsvara ticamības novērtēšanai aprēķina Fišera kritērija empīrisko vērtību, kā faktoru A un Z vidējo kvadrātu attiecību:

$$F = \frac{\overline{Q}_A}{\overline{Q}_Z} = \frac{225.1117}{13.77611} = 16.34073$$

Fišera kritērija teorētisko vērtību F = 3,68 nolasa tabulā (5. pielikums) pēc α = 0,05, $\nu_1 = 2$, $\nu_2 = 15$. Tādējādi nulles hipotēze par faktoru A un Z dispersijām tiek noraidīta un faktora A (spēku sadales variants) ietekme uz sacensību rezultātu ir būtiska. Dispersijas analīzes rezultāti apkopoti 39. tabulā.

18. tabula

Vienfaktora dispersijas analīzes rezultāti

Izkliede	Noviržu kvadrātu summa Q	Ietekmes īpatsvars η^2 procenti	Brīvības pakāpju skaits ν	Vidējais kvadrāts $\overline{Q}$	F	$F_{0,05}$
Kopējā	656.86	100	17	38.63		
Faktora A	450.22	68.5	2	225.11	16.34	3.68
Atlikuma	206.64	31.5	15	13.78		

Ja pētāmā faktora - "spēku sadales distancē" ietekme ir būtiska, tad analīze jāturpina, novērtējot atšķirības starp gradācijas klašu vidējiem aritmētiskiem. Citiem vārdiem - mūs interesē, kurš spēku sadales variants ļauj sasniegt augstāku rezultātu. Vidējo rezultātu starpību ticamības novērtēšanai aprēķina minimālo ticamo starpību. Ja divu vidējo rezultātu starpība sabsolūtā vērtība pārsniedz šo skaitli, novērojamā starpība ir ticama, tātad viens no abiem spēku sadales variantiem veicina augstāka rezultāta sasniegšanu.

Šo pārbaudi var veikt izmantojot Stjūdenta kritēriju. Kritisko starpību aprēķina pēc formulas:

$$\gamma_{\alpha} = t_{\alpha;\nu} \cdot s_D \quad (5.55)$$

$t_{\alpha;\nu} = 2,131$ nolasa Stjūdenta tabulā pēc $\alpha = 0,05$ un $\nu_Z = 15$. Pieņem, ka atsevišķu gradācijas klašu vidējo aritmētisko standartkļūdas ir vienādas un līdz ar to novērtējamās starpības standartkļūdu s_D aprēķina izmantojot formulu:

$$s_D = 1,41 \cdot \sqrt{\frac{\overline{Q}_Z}{n}} = 1,41 \cdot \sqrt{\frac{13,77611}{6}} = 1,5152618$$

tādējādi:

$$\gamma_{\alpha} = 2,131 \cdot 1,5152618 = 3,22902$$

Faktora A gradācijas klašu vidējo aritmētisko starpību analīzes rezultāti attēloti 19. tabulā. Ticamās starpības pasvītrotas.

19. tabula

Faktora A starpību analīzes tabula

Faktora "Spēku sadale" gradācijas klases	Vidējais rezultāts, s $\bar{x}_i$	Starpības D		Kritiskā starpība S_D
		A2, s	A3, s	
A1	289.27	<u>11.73</u>	2.81	3.23
A2	301.00		<u>8.92</u>	3.23
A3	292.08			3.23

Secinājumi. Faktora eksperimentā konstatēts, ka spēku sadale distancē būtiski ietekmē rezultātu 4 km individuālajā iedzīšanas braucienā. Augstāku vidējo rezultātu - 289,27 s sasniedz sportisti, kuri distanci veic relatīvi vienmērīgā atrumā (A1) vai arī (A3) pēc ātra starta "izslēdzas" un veic kāpinājumu distances otrajā pusē - 292,08 s. Starpība starp šiem variantiem nav statistiski ticama. Sākot distanci ļoti spēcīgi un cenšoties visu laiku saglābāt lielu ātrumu, tiek sasniegti attiecīgi par 11,73 s un 8,92 s sliktāki rezultāti ($\alpha < 0,05$).

5.2.2. Divu faktoru dispersijas analīze

Divu faktoru dispersijas analīze pēta divu variējamo faktoru un to mijiedarbības ietekmes īpatsvaru, ietekmes ticamību un novērtē gradācijas klašu vidējo aritmētisko starpības.

Paplašināsim iepriekš aprakstīto viena faktora eksperimentu. No prakses zināms, ka katrā sacensību distancē trekā atkarībā no apstākļiem un sagatavotības riteņbraucējs izmanto optimālu pārnesumu. Tas var būt viens no rezultātu ietekmējošiem faktoriem. Fiziski "pārnesums" izpaužas kā divriteņa priekšējā un pakaļējā zobratu attiecība reizināta ar pakaļējā riteņa apkārtmēru - citiem vārdiem tas ir divriteņa pārvietošanās attālums, pedālim veicot pilnu apli. Šo attālumu sauc par metrāžu, un tas ir identisks soļa garumam skriešanā. Šī faktora gradācijas klases ir kvantitatīvas: 7,32 m un 7,18 m - visbiežāk praksē izmantotie pārnesumi. Ar divu faktoru dispersijas analīzes palīdzību novērtēsm abu faktoru ietekmi uz sacensību rezultātu un atšķirības starp abu faktoru gradācijas klašu kombinācijām.

Piemērs. Pēc speciāla apmācības perioda, kura laikā tika apgūti 3 spēku sadales varianti 4 km individuālajā iedzīšanas braucienā, 6 riteņbraucēji piedalījās faktoreksperimentā. Eksperimenta uzdevums bija, novērtēt taktikas varianta (3 gradācijas klases) un izvēlētā pārnesuma (divas gradācijas klases) ietekmi uz sacensību rezultātu. Eksperimentālajā grupā 6 sportisti. Atbilstoši eksperimenta plānam sportisti piedalījās kontroles sacensībās, kas notika standarta apstākļos - slēgtā velotrekā, 6 dienas pēc kārtas, ik dienas vienā un tai pašā laikā. Uzdevums bija, ik reizi sasniegt maksimāli iespējamo rezultātu. Veicot distanci vienu reizi dienā, katrs sportists izmantoja savu spēku sadales un pārnesuma kombināciju. Pirmajā dienā šī kombinācija tika izlozēta, bet turpmāko secību noteica eksperimenta plāns (skat. 1. daļa "Divu faktoru eksperiments"). Sacensību rezultāti veidoja sekojošu statistisku kompleksu (20. tabula).

20. tabula

Divu faktoru dispersijas analīzes statistiskais komplekss

Faktora A grad.klases	Faktora B grad.klases		n_i	S_i	$\bar{x}_i$
	B1 7,32 m	**B2** 7,18 m			
A1 Relatīvi vienmērīgs ātrums	292.1	291.1	2		
	290	289	2		
	291	292	2		
	281.5	283.4	2		
	291	290.1	2		
	290	289.4	2		
Summas	1735.6	1735	12	3470.6	289.2166667

A2 Ļoti ātrs starts	305	304.3	2		
	300	299.1	2		
	305	305.9	2		
	295	297.2	2		
	300	298.1	2		
	301	300.4	2		
Summas	1806	1805	12	3611	300.9166667

A3 Ātrs starts, izslēgšanās, finiša kāpinājums	293	291.9	2		
	293	291.2	2		
	294	292.1	2		
	285	284.3	2		
	294	292.3	2		
	293	291.5	2		
Summas	1752	1743.3	12	3495.3	291.275

n_j	18	18	36		
S_j	5293.6	5283.3		10576.9	
$\bar{x}_j$	294.0888889	293.5166667			293.8027778
			N	S	$\bar{x}$

$\sum x_{ijp}^2 = 3108826.01$

$C = 3107522.6$

Aprēķina visa kompleksa variančsu skaitu - N, variančsu summu - S un variančsu kvadrātu summu - $\sum x_{ijp}^2$, kur i - faktora B gradācijas klases numurs, j - faktora A gradācijas klases numurs, p - variantes numurs attiecīgajā grupā. Izskaitļo korekciju - C:

$$C = \frac{S^2}{N} = \frac{10576{,}9^2}{36} = 3107523 \qquad (5.56)$$

Aprēķina noviržu kvadrātu summas.

Kopējo noviržu kvadrātu summa:

$$Q = \sum x_{ijp}^2 - C = 3108826 - 3107523 = 1303{,}41 \qquad (5.57)$$

Faktoram A atbilstošo noviržu kvadrātu summa:

$$Q_A = \frac{\sum S_i^2}{mn} - C = \frac{3470{,}6^2 + 3611^2 + 3495{,}3^2}{2 \cdot 6} = 936{,}3539 \qquad (5.58)$$

Faktoram A atbilstošo noviržu kvadrātu summa:

$$Q_B = \frac{\sum S_j^2}{kn} - C = \frac{5293{,}6^2 + 5283{,}3^2}{3 \cdot 6} - 3107523 = 2{,}946944 \qquad (5.59)$$

Faktoru A un B mijiedarbības noviržu kvadrātu summa:

$$Q_{AB} = \frac{\sum S_{ij}^2}{n} - \frac{\sum S_i^2}{mn} - \frac{\sum S_j^2}{kn} + C \qquad (5.60)$$

$$Q_{AB} = \frac{1735{,}6^2 + 1735^2 + 1806^2 + 1805^2 + 1752^2 + 1743{,}2^2}{6} -$$
$$- \frac{3740{,}6^2 + 3611^2 + 3495{,}3^2}{2 \cdot 6} - \frac{5293{,}6^2 + 5283{,}3^2}{3 \cdot 6} + 3107523 =$$
$$= 3{,}473889$$

Apzīmējumi formulās: n - varianšu skaits vienā kopā (atbilst vienai faktoru kombinācijai), k - faktora A grādācijas klašu skaits, m - faktora B gradācijas klašu skaits.

Nekontrolēto faktoru jeb atlikuma noviržu suma:

$$Q_Z = Q - (Q_A + Q_B + Q_{AB}) \qquad (5.61)$$

$$Q_Z = 1303{,}41 - (936{,}3539 + 2{,}946944 + 3{,}473889) = 360{,}635$$

Vidējo noviržu kvadrātu jeb dispersiju aprēķināšanai un ticamības pārbaudei izskaitļo faktoru, to mijiedarbības un atlikuma brīvības pakāpju skaitus.

Kopējais brīvības pakāpju skaits:

$$\nu = N - 1 = 36 - 1 = 35$$

Faktora A brīvības pakāpju skaits:

$$\nu_A = k - 1 = 3 - 1 = 2$$

Faktora B brīvības pakāpju skaits:

$$\nu_B = m - 1 = 2 - 1 = 1$$

Faktoru A un B mijiedarbības brīvības pakāpju skaits:

$$\nu_{AB} = \nu_A \cdot \nu_B = 2 \cdot 1 = 2$$

Atlikuma brīvības pakāpju skaits:

$$\nu_Z = \nu - (\nu_A + \nu_B + \nu_{AB}) = 35 - (2 + 1 + 2) = 30$$

Faktora A dispersija:

$$\overline{Q}_A = \frac{Q_A}{\nu_A} = \frac{936{,}3539}{2} = 468{,}1769 \qquad (5.62)$$

Faktora B dispersija:

$$\overline{Q}_B = \frac{Q_B}{\nu_B} = 2{,}946944 \qquad (5.63)$$

Faktoru A un B mijiedarbības dispersija:

$$\overline{Q}_{AB} = \frac{Q_{AB}}{\nu_{AB}} = \frac{3{,}473889}{2} = 1{,}736944 \qquad (5.64)$$

Atlikuma Z dispersija:

$$\overline{Q}_Z = \frac{Q_Z}{\nu_Z} = \frac{360{,}635}{30} = 12{,}02117 \qquad (5.65)$$

Kvadrātu summas, brīvības pakāpju skaitus un vidējos kvadrātus (dispersijas) ieraksta analīzes tabulā. Aprēķina īpatsvarus.

Faktora A ietekmes īpatsvars:

$$\eta_A^2 = \frac{Q_A}{Q} \cdot 100\% = \frac{936{,}3539}{1303{,}41} \cdot 100\% = 71{,}8\% \qquad (5.66)$$

Faktora B ietekmes īpatsvars:

$$\eta_B^2 = \frac{Q_B}{Q} \cdot 100\% = \frac{2{,}946944}{1303{,}41} \cdot 100\% = 0{,}2\% \qquad (5.67)$$

Faktoru A un B mijiedarbības ietekmes īpatsvars:

$$\eta_{AB}^2 = \frac{Q_{AB}}{Q} \cdot 100\% = \frac{3{,}473889}{1303{,}41} \cdot 100\% = 0{,}266523\% \qquad (5.68)$$

Atlikuma īpatsvars:

$$\eta_Z^2 = \frac{Q_Z}{Q} \cdot 100\% = \frac{360{,}635}{1303{,}41} \cdot 100\% = 27{,}7\% \qquad (5.69)$$

Faktora ietekmes īpatsvara ticamības novērtēšanai aprēķina Fišera kritērija empīrisko vērtību, kā attiecīgā faktora un atlikuma vidējo kvadrātu attiecību.

Faktoram A:

$$F_A = \frac{\overline{Q}_A}{\overline{Q}_Z} = \frac{468{,}1769}{12{,}02117} = 38{,}94605 \qquad (5.70)$$

faktoram B:

$$F_B = \frac{\overline{Q}_B}{\overline{Q}_Z} = \frac{2{,}946944}{12{,}02117} = 0{,}245146 \qquad (5.71)$$

faktoru A un B mijiedarbība:

$$F_{AB} = \frac{\overline{Q}_{AB}}{\overline{Q}_Z} = \frac{1{,}736944}{12{,}02117} = 0{,}144491 \qquad (5.72)$$

Fišera kritērija teorētisko vērtību ***F*** nolasa tabulā (5. pielikums) pēc $\alpha = 0{,}05$, ν_1 - novērtējamo īpatsvaru (A, B, AB) brīvības pakāpju skaits, ν_2 - atlikuma brīvības pakāpju skaits.

21. tabula

Divu faktoru dispersijas analīzes rezultāti

Izkliede	Noviržu kvadrātu summa Q	Ietekmes īpatsvars η^2, procenti	Brīvības pakāpju skaits ν	Vidējais kvadrāts $\overline{Q}$	Fišera kritē-rijs F	$F_{0,05}$
Kopējā	1303.41	100	35	37.24		
Faktora A	936.9569	71.8	2	468.18	38.94	3.32
Faktora B	2.946944	0.2	1	2.95	0.24	4.17
A+B	3.43889	0.3	2	1.74	0.1	3.32
Atlikuma	360.635	27.7	30	12.02		

Nulles hipotēze par faktoru A un Z dispersijām tiek noraidīta ($F > F_{0,05;2;30}$) un faktora A (spēku sadales variants) ietekme uz sacensību rezultātu ir būtiska. Faktora B un faktoru A un B mijiedarbības ietekmes īpatsvari nav statistiski ticami. Tādēļ gradācijas klasu vidējo starpību analīze attiecināma tikai uz faktoru A.
Starpību ticamību pārbauda Stjūdenta kritēriju. Kritisko starpību aprēķina pēc formulas:

$$\gamma_\alpha = t_{\alpha;\nu} \cdot s_D \qquad (5.73)$$

$t_{\alpha;\nu} = 1{,}96$ nolasa Stjūdenta tabulā pēc $\alpha = 0{,}05$ un $\nu_Z = 30$. Pieņem, ka atsevišķu gradācijas klašu vidējo aritmētisko standartkļūdas ir vienādas un līdz ar to novērtējamās starpības standartkļūdu s_D aprēķina izmantojot formulu:

$$s_D = 1{,}41 \cdot \sqrt{\frac{\overline{Q}_Z}{mn}} = 1{,}41 \cdot \sqrt{\frac{12{,}02117}{2 \cdot 6}} = 1{,}41124319 \qquad (5.74)$$

tādējādi:

$$\gamma_\alpha = 1{,}96 \cdot 1{,}41124319 = 2{,}76603$$

Faktora A gradācijas klašu vidējo aritmētisko starpību analīzes rezultāti attēloti 43. tabulā. Ticamās starpības pasvītrotas.

22. tabula

Faktora A starpību analīzes tabula

Spēku sadales distancē, gradācijas klases	Vidējais rezultāts, s $\bar{x}_i$	Starpības D		Kritiskā starpība S_D
		Ātrs sākums, s	Ātrs-vid-ātrs, s	
A1	289,22	11,7	2,05	2,77
A2	300,92		9,65	
A3	291,27			

Secinājumi. Faktora eksperimentā konstatēts, ka spēku sadale distancē būtiski ietekmē rezultātu 4 km individuālajā iedzīšanas braucienā. Pārnesuma variācijas - 7,32 m un 7, 18 m ietekmi uz rezultātu šajos spēka sadales variantos neizdevās pierādīt. Sportisti saniedz augstāku vidējo rezultātu, veicot distanci relatīvi vienmērīgā atrumā - 289,22 s vai arī pēc ātra starta "izslēdzoties" un kāpinot ātrumu distances pēdējā trešdaļā - 291,27 s. Starpība starp šiem variantiem nav statistiski ticama. Sākot distanci ļoti spēcīgi un cenšoties visu laiku saglābāt lielu ātrumu, tiek sasniegti par attiecīgi 11,7 s un 9,65 s sliktāki rezultāti ($\alpha < 0,05$).

6. KORELĀCIJAS UN REGRESIJAS ANALĪZE

6.1. Funkcionālā atkarība un korelācija

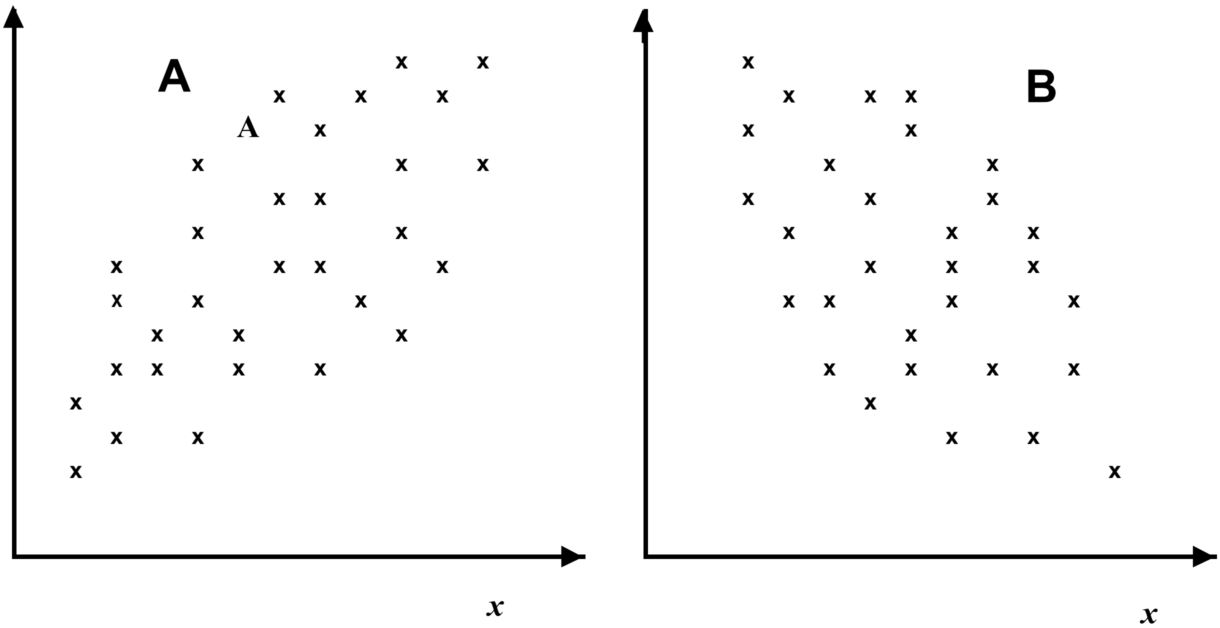
Vairums parādību dabā ir savstarpēji saistītas. Noteikta saikne ir arī starp pazīmēm, kuras raksturo mūsu pētamo objektu - sportistu. Teorētiski varētu pieņemt, ka nedalāma, dzīva organisma funkcijas un līdz ar to arī visas sportista īpašības ir savstarpēji atkarīgas. Būtībā šī dažādo pazīmju savstarpējā ietekme var būt stiprāk vai vājāk izteikta un bieži arēji nav pamanāma. Tādēļ bioloģiskam objektam - sportistam piemītošās īpašības nosacīti iedala divās grupās: savstarpēji atkarīgas un neatkarīgas pazīmes (anatomiski morfoloģiskie rādītāji, elpošanas un sirds-asinsvadu sistēmas īpašības, nervu sistēmas īpatnības, sportista fiziskās īpašības: spēks, ātrums, izturība, veiklība, lokanība, gribas īpašības, tehniskās un taktiskās iemaņas u.c.). Vieglatlēti - metēji treniņā izmanto vingrinājumus ar svaru stieni, jo pieaugot noteiktu muskuļu grupu spēkam uzlabojas rezultāti lodes grūšanā, diska, šķēpa un vesera mešanā. Turpretī maratonistam šādi vingrinājumi, kuri saistīti ar muskuļu masas pieaugumu var izraisīt nevēlamu efektu, tādēļ arī tos neizmanto. Tātad absolūtais spēks un rezultāti mešanas disciplīnās vai maratonskrējienā ir savstarpēji atkarīgas pazīmes. Muskuļu spēka pieaugums neietekmē šahista meistarību, tāpat kā peldēt prasme - šāvēja precizitāti. Neatkarīgas pazīmes ir arī sprintera maksimālais ātrums distancē un spēja īsā laikā sasniegt šo ātrumu starta paātrinājumā. Ja katrai pazīmes ***X*** vērtībai atbilst viena noteikta pazīmes ***Y*** vērtība, tad starp pazīmēm ir funkcionāla atkarība. Ar tādu sastopamies eksaktajās zinātnēs - matemātikā, fizikā, astronomijā. Dabā starp bioloģiskajām pazīmēm šādas stingri definētas atkarības nav. Parasti katrai faktorālās pazīmes vērtībai atbilst vairākas rezultatīvās pazīmes vērtības. Šādu atkarības veidu sauc par korelāciju. Katru korelāciju ar pieņemto ticamības līmeni var tuvināt kādai matemātiskai funkcijai. Tad faktorālā pazīme ***X*** (variantes - x_i) ir neatkarīgais mainīgais - arguments, bet rezultatīvā pazīme ***Y*** (variantes - y_i) - atkarīgais mainīgais jeb funkcija. Kura no pazīmēm ir faktorālā, kura - rezultatīvā, to nosaka cēloņsakarību loģiskās analīzes ceļā.

6.2. Korelâcijas veidi

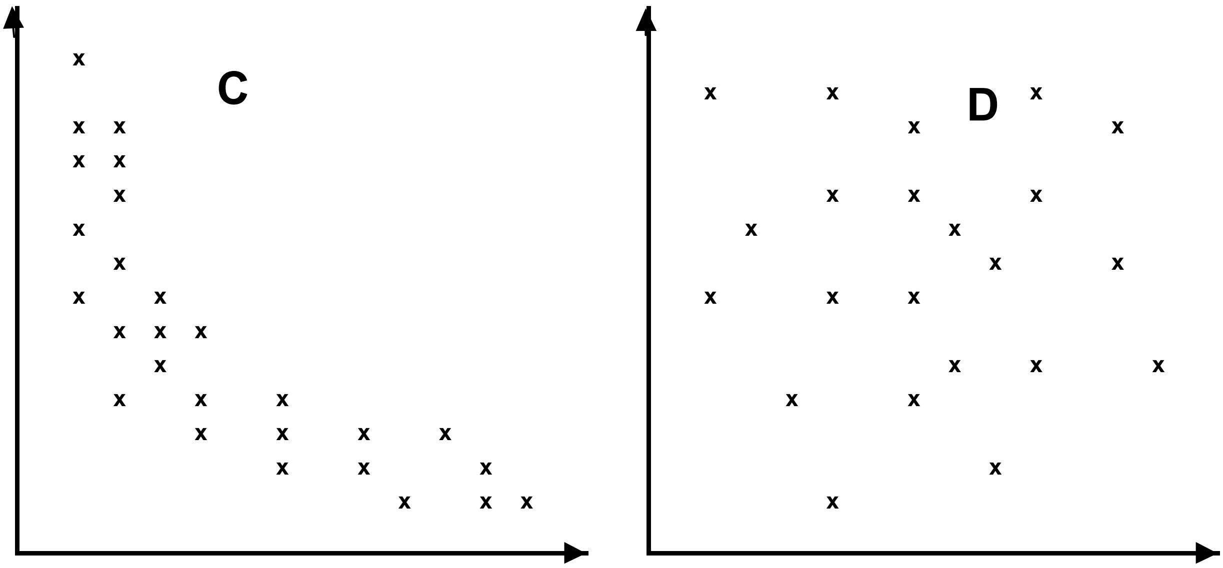
Korelācijas analīzes uzdevums - kvantitatīvi novērtēt atkarību starp divām vai vairākām pazīmēm. Atkarībā no pētīšanas uzdevumiem un analīzes metodēm izšķir pāru korelāciju, parciālo korelāciju un daudzfaktoru korelāciju. Visos gadījumos, veicot analīzi, aprēķina atbilstoša nosaukuma korelācijas rādītāju - koeficientu vai korelācijas attiecību. Šis rādītājs kvantitatīvi raksturo pazīmju savstarpējās ietekmes stiprumu un virzienu. Pāru korelācija ir atkarība starp divām pazīmēm, kuru pēta neņemot vērā varbūtēju citu pazīmju ietekmi uz tām. Tāpēc, ja abas pētamās pazīmes atkarīgas no kādas trešās pazīmes, sastopamies ar šķietamo korelāciju. Piemēram, šķietamā korelācija novērojama starp kāju muskuļu spēku un plaušu vitālo tilpumu - patiesībā savstarpēji neatkarīgām pazīmēm. Abas pazīmes savukārt ir atkarīgas no trešās pazīmes - ķermeņa masas, kas rada šķietamo korelāciju. Parciālā korelācija ir patiesā atkarība starp divām pazīmēm, kuru novērtē izslēdzot citu pazīmju varbūtējo ietekmi uz tām. Parciālās korelācijas koeficients rāda, ka kāju spēks un plaušu vitālais tilpums ir savstarpēji neatkarīgas pazīmes. Vienas pazīmes atkarību no vairāku citu pazīmju grupas sauc par daudzfaktoru jeb multiplo korelāciju. Rezultāts daudzcīņā, piemēram, ir atkarīgs no sasniegumiem katrā disciplīnā.

6.3. Lineārā korelācija

Pāru korelācija var būt lineāra vai nelineāra. Ja vienādām faktorālās pazīmes vērtību izmaiņām atbilst aptuveni vienādas rezultatīvās pazīmes vērtību izmaiņas, tad korelācija ir lineāra - tā tuvojas lineārai funkcijai. Nelineārā atkarība tuvojas kādai līklīnijas funkcijai. Atkarību starp divām pazīmēm var attēlot grafiski - uzzīmējot izkliedes diagrammu (9.1. zīm). Dekarta koordinātu sistēmā uz abscisas (x - ass) attēlo faktorālās pazīmes vērtības, uz ordinātas (y - ass) - rezultatīvās pazīmes vērtības. Diagrammu iegūst atliekot koordinātu plaknē pa pāriem saistīto rezultātu punktus. Ja atkarība ir lineāra, šie punkti novietojas elipsveida apgabalā ap elipses lielo asi - taisni, kas ir tās lineārās funkcijas grafiks, kurai tuvojas korelācija. Kā lineārās korelācijas piemēru var minēt cilvēka sirdsdarbības frekvences atkarību no veicamā mehaniskā darba jaudas (piemēram, uz veloergometra). Ja pulss ir no 110 līdz 180 sit/min, atkarība ir gandrīz funkcionāla. Lineārā korelācija pēc virziena var būt pozitīva (5. zīm. A) vai negatīva (5. zīm. B). Ja pieaugot faktorālās pazīmes vērtībām, vidēji pieaug rezultatīvās pazīmes vērtības, korelācija ir pozitīva - tā tuvojas augošai lineārai funkcijai. Ja pazīmju vērtības mainās pretējos virzienos, korelācija ir negatīva - tuvojas dilstošai lineārai funkcijai. Nelineārās korelācijas diagrammā faktorālās un rezultatīvās pazīmes variantšu pāru punkti grupējas ap kādu matemātisku līkni (6. zīm. C).



5. att. Korelācijas veidi: A - lineāra, pozitīva; B - lineāra, negatīva.



6. att. C - nelineāra korelācija; D - korelācijas pazīmju nav.

Jo stiprāk, vienas pazīmes izmaiņas ietekmē otru pazīmi, jo vairāk atkarība tuvojas funkcijai un variansu pāru punkti izkliedes diagrammā ciešāk grupējas ap nosacīto teorētisko līniju. Pazīmju savstarpējās ietekmes spēku sauc par korelācijas ciešumu.

Uzmanību! Korelācijas analīze ir korekta tikai tad, ja tiek vērtēts loģiski pamatotas pazīmju savstarpējās atkarības ciešums! **Korelācijas analīze ir formāla**

matemātiska metode, kuru nedrīkst izmantot atkarību meklēšanai! Ja neievēro šo noteikumu, tad analizējot "veiksmīgi" aprēķinātus korelācijas koeficientus viegli var nonākt pie sensacionāla secinājuma, ka rezultāts tāllēkšanā atkarīgs no lēcēja deguna garuma, bet sasniegumi maratonskrējienā - no matu daudzuma uz galvas un ķermeņa u.t.t.

6.4. Lineārās pāru korelācijas koeficients

Šo lielumu sauc arī par Pirsona korelācijas koeficientu, un tas raksturo lineārās atkarības virzienu un ciešumu. Pirsona korelācijas koeficientu var izmantot atkarības novērtēšanai, ja pierādīta novērojumu rezultātu atbilstība normālajam sadalījumam. Paraugkopas korelācijas koeficientu apzīmē ar latīņu burtu r, bet ģenerālo korelācijas koeficientu ar grieķu alfabeta burtu ρ (ro). Paraugkopas korelācijas koeficientu būtību raksturo formula:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}} \qquad (6.75)$$

kur x_i - faktorālās pazīmes variante; y_i - rezultatīvās pazīmes variante; $\bar{x}$ - faktorālās pazīmes vidējais aritmētiskais; $\bar{y}$ - rezultatīvās pazīmes vidējais aritmētiskais; n - varianšu pāru skaits jeb paraugkopas apjoms. Formula ir neērta aprēķiniem, jo satur daudz skaitļošanas operāciju un tai raksturīga vidējo aritmētisko noapaļošanas papilkļūda. Praktiskām vajadzībām izmanto formulas darba variantu, kuram šie trūkumi nepiemīt:

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \cdot \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} \qquad (6.76)$$

Korelācijas koeficientu aprēķina noapaļojot līdz 3 zīmēm aiz komata. Tas ir nenosaukts skaitlis un atrodas robežās no -1 līdz +1. Ja $|r| > 1$, tad aprēķinos ieviesusies rupja kļūda. Jo vairāk koeficienta modulis tuvojas 1, jo ciešāka korelācija - spēcīgāk viena pazīme ietekmē otru. Ja $|r| = 0$, pazīmes ir neatkarīgas, bet, ja $|r| = 1$, atkarība ir funkcionāla. Tā kā korelācijas koeficients aprēķināts paraugkopai, tad rodas pamatots jautājums, vai tas ir pietiekoši ticams atkarības novērtēšanai. Ja mūsu rīcībā būtu ģenerālais korelācijas koeficients, tad pazīmju savstarpējā atkarība pastāvētu visos gadījumos, kad šis koeficients nav nulle. Lai pārliecinātos, vai aprēķināto Pirsona korelācijas koeficientu var izmantot atkarības vērtēšanai, koeficienta moduli salīdzina ar kritisko vērtību $r_{\alpha;n}$, kuru nolasa tabulā (4. pielikums) atbilstoši paraugkopas apjomam n un rezultātu būtiskuma līmenim $\alpha = 0{,}05$. Ja koeficienta modulis | r | lielāks vai vienāds ar kritisko vērtību, korelācija ir ticama un var izdarīt secinājumus par tās virzienu un ciešumu. Ja koeficients nav ticams ($r < r_{\alpha;n}$), tad pie dotā novērojumu skaita atkarību nav izdevies pierādīt, un pazīmes jāuzskata par neatkarīgām. Vienmēr jāatceras, ka korelācijas ticamība tieši atkarīga no novērojumu skaita. Piemēram, lai pierādītu vājas korelācijas ticamību $|r| = 0{,}150$, jāizdara vismaz 200 novērojumu. Jāņem vērā, ka pēc korelācijas analīzes rezultātiem daudzmaz nopietnu slēdzienu var izdarīt, ja novērojumu skaits nav mazāks par 50. Tikai ticamas korelācijas gadījumā ($r \geq r_{\alpha;n}$), dod slēdzienu par tās virzienu un ciešumu. Ja $r > 0$, tad korelācija ir pozitīva - pieaugot faktorālās pazīmes X vērtībām attiecīgi palielinās rezultatīvās pazīmes Y vērtības. Negatīvas korelācijas gadījumā ($r < 0$) palielinoties vienas pazīmes vērtībām, otras pazīmes vērtības samazinās. Raksturojot pazīmju savstarpējās ietekmes spēku, korelāciju vērtē kā: vāju, ja $0{,}2 < |r| < 0{,}49$ vidēju, ja $0{,}5 < |r| < 0{,}69$ ciešu, ja $0{,}7 < |r| < 0{,}99$ Pirsona korelācijas koeficientu izmanto parciālās un daudzfaktoru korelācijas koeficientu aprēķināšanai. Papildvērtējumam var izmantot determinācijas koeficientu:

$$D = r^2 \qquad (6.77)$$

Determinācijas koeficients rāda, kāda rezultatīvās pazīmes variācijas daļa tieši atkarīga no faktorālās pazīmes variācijas. Jāiegaumē, ka korelācijas analīze izmantojama tikai palīginstruments loģiskā ceļā izjaustu cēloņsakarību ciešuma vērtēšanai. Akli paļaujoties uz matemātiskās statistikas metožu visspēcību zinātniskajos pētījumos, var nodarīt vairāk ļauna nekā laba.

Piemērs. Doti LSPA 1.kursa studentu ieskaites sacensību rezultāti svarcelšanā (grūšana) un vieglatlētikā - lodes grūšanā. Novērtēt lodes grūšanas rezultātu atkarību no spēka sagatavotības. Novērojuma rezultātus ierakstām darba tabulā (23. tabula), aprēķinām x_i^2, y_i^2, $x_i.y_i$, bet pēc tam summu katrā tabulas stabiņā: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum x_i.y_i$.

23. tabula

Darba tabula korelācijas koeficienta aprēķināšanai

Uzgrūstā stieņa masa, kg x_i	Rezultāts lodes grūšanā, m y_i	x_i^2	y_i^2	$x_i.y_i$
1	2	3	4	5
60	7,8	3600	60.84	468
80	9	6400	81	720
75	9,3	5625	86.49	697.5
90	9,36	8100	87.6096	842.4
83	9,63	6889	92.7369	799.29

23. tabulas turpinājums

1	2	3	4	5
83	10,27	6889	105.4729	852.41
65	8,97	4225	80.4609	583.05
100	11,43	10000	130.6449	1143
95	10	9025	100	950
85	9	7225	81	765
83	9,1	6889	82.81	755.3
80	10,25	6400	105.0625	820
75	9,5	5625	90.25	712.5
73	9,49	5329	90.0601	692.77
70	8,4	4900	70.56	588
70	8,8	4900	77.44	616
85	9,8	7225	96.04	833
90	9,73	8100	94.6729	875.7
95	11,4	9025	129.96	1083
85	10,55	7225	111.3025	896.75
75	8,9	5625	79.21	667.5
70	8,97	4900	80.4609	627.9
85	10	7225	100	850
68	8,9	4624	79.21	605.2
83	11	6889	121	913
75	8,8	5625	77.44	660
90	9,5	8100	90.25	855
95	10,75	9025	115.5625	1021.25
83	9,33	6889	87.0489	774.39
75	10	5625	100	750
90	11,1	8100	123.21	999
90	11	8100	121	990
83	7,88	6889	62.0944	654.04
83	9,48	6889	89.8704	786.84
90	10,5	8100	110.25	945
83	10,08	6889	101.6064	836.64
75	10,38	5625	107.7444	778.5
$\sum x_i =$ $= 3015$	$\sum y_i =$ $= 358.35$	$\sum x_i^2 =$ $= 248715$	$\sum y_i^2 =$ $= 3500.371$	$\sum x_i . y_i =$ $= 29407.93$

$$r = \frac{29407{,}93 - \dfrac{3015 \cdot 358{,}35}{37}}{\sqrt{248715 - \dfrac{3015^2}{37}} \cdot \sqrt{3500{,}371 - \dfrac{358{,}35^2}{37}}} = 0{,}690456 \approx 0{,}690$$

Tabulā (4. pielikums) atrodam $r_{0,05;37}$ = 0,332, tātad korelācija ticama, jo $| r | > r_{0,05;37}$.

Secinājums. LSPA 1.kursa studentu lodes grūšanas rezultātu atkarība no spēka sagatavotības, ko raksturo uzgrūstā stieņa masa ir vidēji cieša un pozitīva (r = 0,690). Ir pamats uzskatīt, ka paaugstinot spēka sagatavotību pieaugs rezultāti lodes grūšanā (P > 0,95). (Šis piemērs ir formāla korelācijas analīzes procedūras demonstrācija, jo novērojumu skaits pat netuvojas vajadzīgajam $n \geq 50$).

6.5. Lineārās regresijas analīze

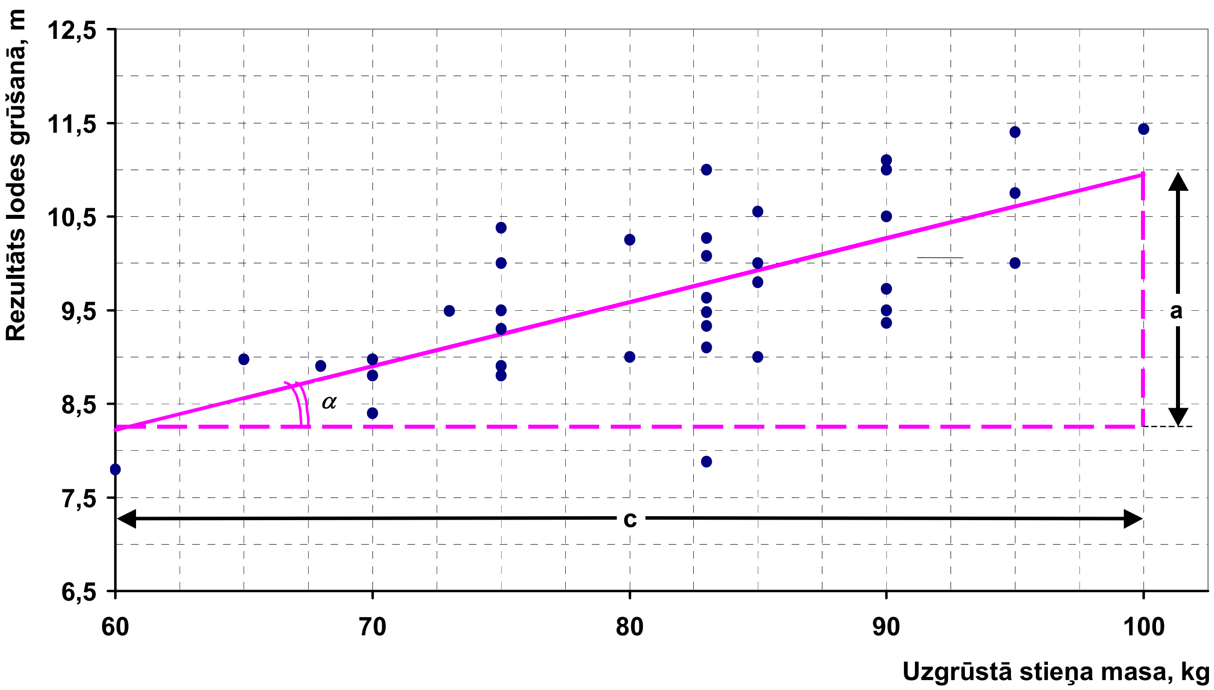
Regresijas analīzi var uzskatīt par korelācijas analīzes turpinājumu. Ar regresiju mēs saprotam likumsakarību, pēc kuras mainās savstarpēji atkarīgo pazīmju vērtības. Regresijas analīzes uzdevums ir šīs likumsakarības aprakstīšana ar izteiksmi, ko sauc par regresijas vienādojumu. Tas ir tās funkcijas vienādojums, kurai tuvojas korelācija. Regresijas vienādojumu izmanto rezultatīvās pazīmes vērtības prognozēšanai, ja zināma faktorālās pazīmes vērtība. Tādējādi katram korelācijas veidam atbilst tā paša nosaukuma regresijas veids: pāru, parciālā un daudzfaktoru regresija. Pēc formas izšķir lineāro un nelineāro regresiju. Regresijas analīzei raksturīgi trīs posmi:

- regresijas vienādojuma izvēle vispārīgā veidā;
- regresijas vienādojuma parametru aprēķināšana;
- regresijas ticamības pārbaude.

Piemērota regresijas vienādojuma izvēle no daudziem iespējamiem lielākajā daļā gadījumu ir sarežģīta procedūra izņemot variantu, kad korelācija ir lineāra un jāizmanto viens no diviem iespējamiem - dilstošas vai augošas lineāras funkcijas vienādojums. Vienādojumu piemeklēšanai izmanto darbietilpīgas matemātiskas metodes un tām atbilstošas datorprogrammas. Regresijas vienādojuma parametri, kuri jāaprēķina analīzes gaitā ir vienādojuma brīvais loceklis un viens vai vairāki regresijas koeficienti. Regrersijas ticamības pārbaude ir tikpat nepieciešama un formāla kā korelācijas koeficienta ticamības pārbaude. Vienkāršos uzdevumos, no regresijas ticamības pārbaudes var atteikties, ja iepriekš pierādīta korelācijas ticamība. Aplūkosim vienkāršāko regresijas analīzes veidu - lineāro pāru regresiju. Divu pazīmju lineārā korelācija tuvojas lineārai funkcijai. Likumsakarību, pēc kuras mainās savstarpēji atkarīgo pazīmju vērtības apraksta ar lineārās funkcijas vienādojumu:

$$\hat{y} = b_0 + b_1 x \tag{6.78}$$

kur $\hat{y}$ - rezultatīvās pazīmes vērtība; x - faktorālās pazīmes vērtība; b_0 - regresijas vienādojuma brīvais loceklis; b_1 - regresijaskoeficients. Analīzes uzdevums - noteikt b_0 un b_1 vērtības. To var izdarīt divējādi.



7. att. Regresijas analīzes grafiskā metode.

6.5.1. Grafiskā metode

Regresijas vienādojuma brīvo locekli b_0 un regresijas koeficientu b_1 var aprēķināt aptuveni, uzzīmējot uz milimetru papīra izkliedes diagrammu (7. att.). Koordinātu plaknē atliek papildpunktu, kas atbilst faktorālās un rezultatīvās pazīmes vidējiem aritmētiskiem. Caur šo punktu pēc acumēra velk taisni tā lai pārējo punktu novirzes no tās būtu aptuveni vienādas. Novilkto taisni sauc par teorētisko regresijas līniju un tā ir regresijas vienādojuma grafiks. Vienādojuma brīvo locekli nosaka izmērot nogriezni no ordinātu ass sākuma līdz tās krustpunktam ar teorētisko regresijas līniju (rezultatīvās pazīmes mērvienībās). Teorētiski b_0 ir rezultatīvās pazīmes vērtība, ja faktorālās pazīmes vērtība ir 0. Nav ieteicams praktiski interpretēt šo skaitli, jo ļoti bieži faktorālās pazīmes vērtība nevar būt 0. Koeficientu b_1 aprēķina kā teorētiskās regresijas līnijas un abscisu ass veidotā leņķa tangensu:

$$b_1 = tg\alpha = \frac{a}{c} \qquad (6.79)$$

Koeficients b_1 rāda par cik vienībām vidēji mainās rezultatīvās pazīmes vērtība, ja faktorālās pazīmes vērtība pieaug par vienu vienību. Izmaiņu virzienu norāda b_1 zīme, kura atbilst korelācijas koeficienta zīmei - negatīvas korelācijas gadījumā - mīnus. Koeficientam b_1 ir praktiska nozīme un secinājumos vienmēr jāpaskaidro tā būtība. Grafiskā regresijas analīzes metode ir vienkārša, bet neprecīza. To var izmantot tikai aptuvena priekšstata iegūšanai par pētāmo sakarību.

6.5.2. Vismazāko kvadrātu metode.

Precīzai regresijas vienādojuma parametru aprēķināšanai izmanto vismazāko kvadrātu metodi t.i. abstrakto regresijas līniju novelk tā, lai rezultātu pāru punktu noviržu no teorētiskās regresijas līnijas kvadrātu summa būtu vismazākā. Metodes darbietilpības dēļ ieteicams izmantot atbilstošu datorprogrammu. Aprēķinus veic bez zīmējuma un atrisinot vienādojumu sistēmu nonāk pie sekojošām formulām:

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad (6.80)$$

$$b_o = \bar{y} - b_1\bar{x} \qquad (6.81)$$

kur $\bar{y}$ un $\bar{x}$ ir attiecīgi rezultatīvās un faktorālās pazīmes vidējie aritmētiskie. b_0 un b_1 noapaļo atstājot vienu zīmi vairāk aiz komata nekā sākuma datiem. Tāpat kā viendimensijas paraugkopas analīzē aprēķina izkliedes rādītāju regresijas standartnovirzi:

$$s_{yx} = \sqrt{\frac{\sum y_i^2 - b_0\sum y_i - b_1\sum x_i y_i}{n-2}} \qquad (6.82)$$

Regresijas standatnovirze s_{yx} raksturo rezultātu pāru punktu izkliedi ap teorētisko regresijas līniju, līdzīgi tam, kā viendimensijas paraugkopas standartnovirze novērtē varianšu blīvumu ap vidējo aritmētisko. Jo ciešāka korelācija, jo tuvāk regresijas līnijai novietoti rezultātu pāru punkti un mazāka standartnovirzes vērtība. Funkcionālas atkarības gadījumā $s_{yx} = 0$. Regresijas standartnovirze ir nosaukts skaitlis (tam ir rezultatīvās pazīmes mērvienība).

Regresijas ticamību novērtē pārbaudot nulles hipotēzi par regresijas koeficientu. Izmanto Stjūdenta kritēriju. Šai nolūkā aprēķina koeficienta b standartkļūdu:

$$s_{b_1} = \frac{s_{yx}}{\sqrt{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}} \qquad (6.83)$$

un Stjūdenta kritērija empīrisko vērtību:

$$t = \frac{|b_1|}{s_{b_1}} \tag{6.84}$$

Tabulā (1. pielikums) pēc $\alpha = 0{,}05$ un $\nu = n - 2$ nolasa kritērija robežvērtību $t_{\alpha;\nu}$. Ja $t \geq t_{\alpha;\nu}$, regresija ir ticama, un regresijas vienādojumu var izmantot praksē rezultatīvās pazīmes vērtību prognozēšanai. Ja iepriekš aprēķinātais korelācijas koeficients ir ticams, un nav iespējams izmantot ESM , var pieņemt, ka regresija ir ticama un no pēdējās procedūras atteikties. Korelācijas un regresijas analīzes aprēķiniem lieliski var izmantot elektronisko tabulu EXCEL.

6.5.3. Prognozēšana

Regresijas analīzes rezultātus izmanto praktiskām vajadzībām. Ievietojot regresijas vienādojumā izvēlēto faktorālās pazīmes vērtību x, var aprēķināt tai atbilstošo rezultatīvās pazīmes vērtību. Šādu operāciju sauc par prognozēšanu. Ja atkarība ir funkcionāla, prognozēšana aprobežojas ar funkcijas aprēķināšanu, atbilstoši izvēlētai argumenta vērtībai. Turpretī korelācijas gadījumā tāda prognoze ir nekorekta, jo aprēķinātais gaidāmais rezultāts var atšķirties no patiesā, kuru iegūsim praksē. Tādēļ novērtējumu dod norādot intervāla robežas, kurā ar pieņemto ticamības līmeni var atrasties gaidāmais rezultāts. Pamatojoties uz normālā sadalījuma likumu, prognozēšanai var izmantot regresijas zonu, kura atbilst izvēlētajam ticamības līmenim (sk. korelācijas diagrammu). Lai aprēķinātu robežas, kurās atradīsies gaidāmais rezultāts, prognozēšanai izmanto izteiksmi:

$$b_0 + b_1 - t \cdot s_{yx} \leq \hat{y} \leq b_0 + b_1 + t \cdot s_{yx} \tag{6.85}$$

kur t - normētā novirze. Doto izteiksmi var izprast aplūkojot izkliedes diagrammu (8. att.). Atliekot uz ordinātas pa ±3 standartnovirzēm no teorētiskās līnijas krustpunkta un no iegūtajiem punktiem velkot regresijas līnijai paralēlas taisnes, var izveidot trīs regresijas zonas. Pēc normālā sadalījuma likuma vienas standartnovirzes zonā ap teorētisko regresijas līniju atrodas 68,3%, divu standartnoviržu - 95,5%, bet triju standartnoviržu intervālā - 99,7% visu rezultātu pāru punktu. Var pieņemt, ka prognozējamā rezultāta parādīšanās varbūtība vienā no šīm zonām ir attiecīgi 0,683, 0,955, 0,997. Teorētiski šis apgalvojums pareizs, ja $n \to \infty$. Tādēļ praktiski korelācijas un regresijas analīzes rezultātus var izmantot tikai tad, ja novērojumu skaits pietiekoši liels. Prognozēšanai izmantojamo zonu izvēlas atkarībā no rezultātu ticamības līmeņa (P = 0,95). Šāda regresijas zona attēlota 8. att. Lai vienkāršotu uzdevumu var izmantot 2 standartnoviržu zonu (P = 0,955). Tādā gadījumā prognozēšanas izteiksme ir:

$$b_0 + b_1 - 2s_{yx} \le \hat{y} \le b_0 + b_1 + 2s_{yx} \qquad (6.86)$$

Ja vēlas noteikt zonas robežas atbilstoši izvēlētajam ticamības līmenim, t atrod varbūtību integrāļu tabulā. Ja P = 0,95, tad t = 1,96. Tādējādi:

$$b_0 + b_1 - 1{,}96s_{yx} \le \hat{y} \le b_0 + b_1 + 1{,}96s_{yx} \qquad (6.87)$$

Ja faktorālās pazīmes vērtība atrodas empīriskā intervāla robežās ($x_{min} \le x \le x_{max}$), prognozēšanu sauc par interpolāciju. Turpretī, ja x vērtība ir ārpus empīriskā intervāla robežām ($x < x_{min}$ vai $x > x_{max}$), tad operāciju sauc par ekstrapolāciju. Lai izvairītos no rupjām kļūdām, ekstrapolējot pieļaujamas tikai nelielas novirzes no empīriskā intervāla robežām, jo nav zināms, vai ārpus tā nemainās atkarības forma.

Piemērs. Turpināsim korelācijas analīzē iesākto uzdevumu. Novērtēsim likumsakarību pēc kādas mainās rezultāts lodes grūšanā atkarībā no spēka sagatavotības, kuru vērtē pēc grūšanā paceltā stieņa masas, un sastādīsim izteiksmi lodes grūšanas rezultāta prognozēšanai. Aprēķiniem izmanto korelācijas analīzē sastādīto darba tabulu. Regresijas koeficients:

$$b_1 = \frac{37 \cdot 29407{,}93 - 3015 \cdot 358{,}35}{37 \cdot 248715 - 3015^2} = 0{,}068325 \approx 0{,}068$$

Faktorālās un rezultatīvās pazīmju vidējie aritmētiskie:

$$\bar{x} = \frac{3015}{37} = 81{,}486486 \approx 81{,}486$$

$$\bar{y} = \frac{358{,}35}{37} = 9685135 \approx 9{,}685$$

Regresijas vienādojuma brīvais loceklis:

$$b_0 = 9{,}685 - 0{,}068 \cdot 81{,}486 = 4{,}117538 \approx 4{,}12$$

Tātad regresijas vienādojums ir:

$$\hat{y} = 4{,}12 + 0{,}07x$$

Regresijas standartnovirze:

$$s_{yx} = \sqrt{\frac{3500{,}371 - 4{,}118 \cdot 358{,}35 - 0{,}068 \cdot 29407{,}93}{37-2}} \approx 0{,}666$$

Regresijas koeficienta standartkļūda:

$$s_{b_1} = \frac{0{,}67}{\sqrt{248715 - \frac{3015^2}{37}}} = 0{,}0121$$

Stjūdenta kritērija empīriskā vērtība:

$$t = \frac{0{,}068}{0{,}0121} = 5{,}646864 \approx 5{,}647$$

$\alpha = 0{,}05$ un $\nu = n - 2$ nolasa kritērija robežvērtību $t_{\alpha;\nu}$. Ja $t \geq t_{\alpha;\nu}$,

Pēc $\alpha = 0{,}05$ un $\nu = 37 - 2 = 35$ tabulā (1. pielikums) nolasām $t_{0,05;35} = 1{,}96$. Regresija ir ticama, jo $t \geq t_{0,05;35}$. Izteiksme lodes grūšanas rezultāta prognozēšanai (P = 0,95) ir:

$$4{,}117 + 0{,}068 - 1{,}96 \cdot 0{,}666 \leq \hat{y} \leq 4{,}117 + 0{,}068 + 1{,}96 \cdot 0{,}666$$

$$2{,}81 + 0.07x \leq \hat{y} \leq 5{,}42 + 0{,}07x$$

Abu pazīmju savstarpējo atkarību varam attēlot korelācijas diagrammā (8. att.).

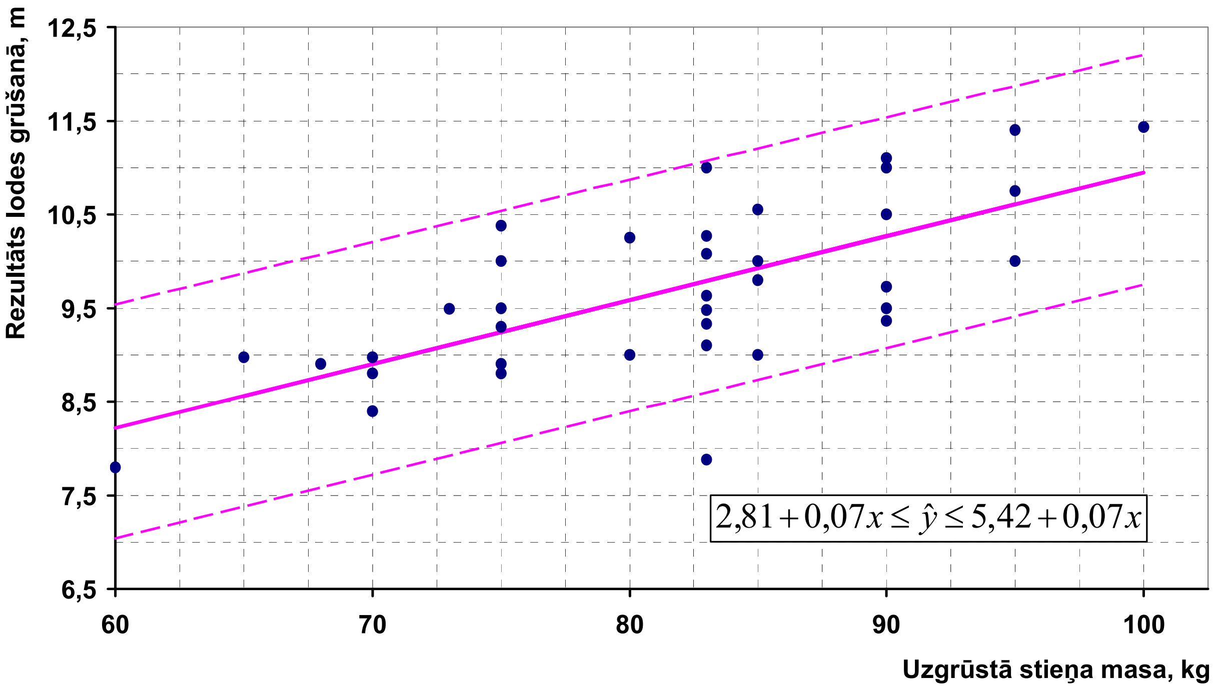
Secinājums. Likumsakarību, pēc kuras mainās rezultāts lodes grūšanā atkarībā no sasniegumiem svarcelšanā apraksta regresijas vienādojums:

$$\hat{y} = 4{,}12 + 0{,}07x$$

Regresija ir ticama (P > 0,95), lodes grūšanas prognozēšanai izmantojama izteiksme:

$$2{,}81 + 0.07x \leq \hat{y} \leq 5{,}42 + 0{,}07x$$

Ja vēlamies prognozēt iespējamo rezultātu lodes grūšanā, tad ievietojam x vietā skaitli, kas atbilst šī sportista sasniegumam svarcelšanā (grūšana).



8. att. Sakarība starp sasniegumiem svarcelšanā un vieglatlētikā.

6.6. Spīrmena rangu korelācijas koeficients

Šo rādītāju izmanto divu pazīmju savstarpējās atkarības novērtēšanai, ja nav pierādīta novērojumu rezultātu atbilstība normālajam sadalījumam kā arī tad, ja viena vai abas pazīmes novērtētas izmantojot rangu skalu. Sastāda darba tabulu (24. tabula), kuras pirmajās divās ailēs ieraksta attiecīgi pa pāriem saistītās faktorālās un rezultatīvās pazīmes variantes - x_i un y_i , 4. stabiņā ieraksta šo varianšu rangus $g_{x.i}$ un $g_{y.i}$. Aprēķina pa pāriem saistīto rangu starpības - $d_i = g_{x.i} - g_{y.i}$ un ieraksta tās 5. stabiņā, bet 6. stabiņā - šo starpību kvadrātus d_i^2. Aprēķina starpību kvadrātu summu $\sum d_i^2$.

Spirmena rangu korelācijas koeficientu aprēķina pēc formulas:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \qquad (6.88)$$

kur n - variančšu pāru skaits; d_i - varianšu rangu pāru starpība. Korelācija ir ticama, ja $|r| > r_\alpha$. Kritisko vērtību atrod tabulā. Spirmena rangu korelācijas koeficients tāpat kā Pirsona korelācijas koeficients raksturo atkarības virzienu un ciešumu.

24. tabula

Darba tabula rangu korelācijas koeficienta aprēķināšanai

Faktorālās pazīmes variante x_i	Rezultatīvās pazīmes variante y_i	x rangs $g_{x.i}$	y rangs $g_{y.i}$	Rangu starpība $d_i = g_{x.i} - g_{y.i}$	d_i^2
x_1	y_1	$g_{x.1}$	$g_{y.1}$	d_1	d_1^2
x_2	y_2	$g_{x.2}$	$g_{y.2}$	d_2	d_2^2
x_3	y_3	$g_{x.3}$	$g_{y.3}$	d_3	d_3^2
x_{n-1}	y_{n-1}	$g_{x.n-1}$	$g_{y.n-1}$	d_{n-1}	d_{n-1}^2
x_n	y_n	$g_{x.n}$	$g_{y.n}$	d_n	d_n^2
					$\sum d_i^2$

Piemērs. Doti jauno riteņbraucēju (11-12 g.v.) kontroles sacensību rezultāti tāllēkšanā no vietas un 10 km individuālajā braucienā. Novērtēt testa "tāllēkšana no vietas" informatīvumu, aprēķinot Spirmena rangu korelācijas koeficientu. Rezultātus x_i un y_i ieraksta darba tabulā (46. tabula), pieraksta tiem rangus, aprēķina rangu starpības d_i , to kvadrātus d_i^2 un aprēķina summu -$\sum d_i^2$, kuru ievieto Spirmena rangu korelācijas koeficienta r_S aprēķināšanai. Tabulā (6. pielikums) atrodam korelācijas koeficienta kritisko vērtību $r_S = 0{,}44$. $|r| > r_{0,05}$, tātad korelācija starp rezultātiem tāllēkšanā no vietas un 10 km individuālajā braucienā ir ticama. Varam secināt, ka 11-12 g.v. zēniem-iesācējiem, kuri ir labāki tāllēkšanā no vietas, ir lielākas perspektīvas sasniegt augstākus rezultātus 10 km individuālajā braucienā.

25. tabula

Darba tabula

Tāllēkšana no vietas, m	Ind.brauciens 10 km, min:s	Rangs tāllēkšanā $g_{x.i}$	Rangs ind.braucienā $g_{y.i}$	$\lvert d_i \rvert$	d_i^2
1	2	3	4	5	6
26.04	18:34	16.5	13	3.5	12.25
1.64	18:52	2	16	14	196
1.72	18:57	3	17	14	196
2.07	17:41	19	3	16	256
1.90	21:02	9.5	21	11.5	132.25
1.53	20:58	1	20	19	361
1.95	18:13	12	8	4	16
1.78	19:25	5	19	14	196

25. tabulas turpinājums

1	2	3	4	5	6
1.98	17:40	15	2	13	169
1.84	18:39	6	14	8	64
1.92	18:17	11	10	1	1
1.87	18:29	8	12	4	16
1.90	17:57	9.5	5	4.5	20.25
1.85	18:16	7	9	2	4
1.97	17:33	14	1	13	169
2.06	18:10	18	6	12	144
1.96	18:12	13	7	6	36
1.77	19:13	4	18	14	196
2.15	17:56	21	4	17	289
2.12	18:45	20	15	5	25
2.04	18:25	16.5	11	5.5	30.25
					$\sum d_i^2 =$ $= 2529$

$$r_s = 1 - \frac{6 \cdot 2529}{21(21^2 - 1)} = -0,64216233 \approx -0,642$$

Noslēdzot apskatu par matemātiskās statistikas metodēm jāatzīmē, ka minējām tikai vispopulārākās. To metožu skaits kuras var izmantot sporta zinātnē, ir ievērojami lielāks. Plašāk ar tām var iepazīties izmantojot speciālo literatūru [5, 6, 8, 12, 13].

PIELIKUMI

1. Stjūdenta kritērija teorētiskās vērtības $t_{\alpha;\nu}$

ν	α			ν	α		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6.314	12,706	63.657	17	1.740	2,110	2.898
2	2.920	4,303	9.925	18	1.734	2,101	2.878
3	2.353	3,182	5.841	19	1.729	2,093	2.861
4	2.132	2,776	4.604	20	1.725	2,086	2.845
5	2.015	2,571	4.032	21	1.721	2,080	2.831
6	1.943	2,447	3.707	22	1.717	2,074	2.819
7	1.895	2,365	3.499	23	1.714	2,069	2.807
8	1.860	2,306	3.355	24	1.711	2,064	2.797
9	1.833	2,262	3.250	25	1.708	2,060	2.787
9	1.812	2,228	3.169	26	1.706	2,056	2.779
11	1.796	2,201	3.106	27	1.703	2,052	2.771
12	1.782	2,179	3.055	28	1.701	2,048	2.763
13	1.771	2,160	3.012	29	1.699	2.045	2.756
14	1.761	2,145	2.977	30	1.697	2,042	2.750
15	1.753	2,131	2.947	>30	1.645	1,960	2.576
16	1.746	2,120	2.921				

Nulles hipotēzi noraida t.i. vērtējamā starpība ir ticama, ja $t \geq t_{\alpha;\nu}$

2. Vilkoksona kritērija teorētiskās vērtības

$T_{\alpha;n}$ ($\alpha = 0,05$)

n	$T_{\alpha;n}$	n	$T_{\alpha;n}$	n	$T_{\alpha;n}$
6	1	13	17	20	52
7	2	14	21	21	59
8	4	15	25	22	66
9	6	16	30	23	73
10	8	17	35	24	81
11	11	18	40	25	89
12	14	19	46		

Nulles hipotēzi noraida, ja $T \leq T_{\alpha;n}$

3. Van der Vardena kritērija teorētiskās vērtības

$X_\alpha (\alpha = 0{,}05)$

n	$n_1 - n_2$			n	$n_1 - n_2$		
	0 | 1	2 | 3	4 | 5		0 | 1	2 | 3	4 | 5
8	2,40	2,30	-	30	4,88	4,87	4,84
9	2,38	2,20	-	31	4,97	4,95	4,91
10	2,60	2,49	2,30	32	5,07	5,06	5,03
11	2,72	2,58	2,40	33	5,15	5,13	5,10
12	2,86	2,79	2,68	34	5,25	5,24	5,21
13	2,96	2,91	2,78	35	5,33	5,31	5,28
14	3,11	3,06	3,00	36	5,42	5,41	5,38
15	3,24	3,19	3,06	37	5,50	5,48	5,45
16	3,39	3,36	3,28	38	5,59	5,58	5,55
17	3,49	3,44	3,36	39	5,67	5,65	5,62
18	3,63	3,60	3,53	40	5,75	5,74	5,72
19	3,73	3,69	3,61	41	5,83	5,81	5,79
20	3,86	3,84	3,78	42	5,91	5,90	5,88
21	3,96	3,92	3,85	43	5,99	5,97	5,95
22	4,08	4,06	4,01	44	6,04	6,06	6,04
23	4,18	4,15	4,08	45	6,14	6,12	6,10
24	4,29	4,27	4,23	46	6,21	6,21	6,19
25	4,39	4,36	4,30	47	6,29	6,27	6,25
26	4,50	4,48	4,44	48	6,36	6,35	6,34
27	4,59	4,56	4,51	49	6,43	6,42	6,39
28	4,68	4,68	4,64	50	6,50	6,51	6,48
29	4,78	4,76	4,72				

Nulles hipotēzi noraida, ja $X > X_\alpha$

4. Korelācijas koeficienta kritiskās vērtības

$r_{\alpha;n}$ ($\alpha = 0{,}05$)

n	$r_{\alpha;n}$	n	$r_{\alpha;n}$	n	$r_{\alpha;n}$	n	$r_{\alpha;n}$
4	0,950	15	0,514	26	0,388	80	0,219
5	0,875	16	0,497	27	0,381	90	0,206
6	0,811	17	0,487	28	0,374	100	0,196
7	0,754	18	0,468	29	0,367	125	0,175
8	0,707	19	0,456	30	0,361	150	0,160
9	0,666	20	0,444	35	0,332	200	0,135
10	0,632	21	0,433	40	0,310	250	0,124
11	0,602	22	0,423	45	0,292	300	0,113
12	0,576	23	0,413	50	0,277	400	0,098
13	0,533	24	0,404	60	0,253	500	0,088
14	0,532	25	0,396	70	0,234	1000	0,063

Korelācija ir ticama, ja $r_{\alpha;n} < r$

5. Fišera kritērija teorētiskās vērtības

$F_{\alpha;\nu1;\nu2}$ ($s_1^2 > s_1^2$; $\alpha = 0{,}05$)

ν_2	ν_1							
	8	9	10	12	15	20	30	>30
8	3,44	3,39	3,35	3,28	3,22	3,15	3,08	2,93
9	3,23	3,18	3,14	3,07	3,01	2,94	2,86	2,71
10	3,07	3,02	2,98	2,91	2,85	2,77	2,70	2,54
11	2,93	2,90	2,85	2,79	2,72	2,65	2,57	2,40
12	2,85	2,80	2,75	2,69	2,62	2,54	2,47	2,30
13	2,77	2,71	2,67	2,60	2,53	2,46	2,38	2,21
14	2,70	2,65	2,60	2,53	2,46	2,39	2,31	2,13
15	2,64	2,59	2,54	2,48	2,40	2,33	2,25	2,07
16	2,59	2,54	2,49	2,42	2,35	2,28	2,19	2,01
17	2,55	2,49	2,45	2,38	2,31	2,23	2,15	1,96
18	2,51	2,46	2,41	2,34	2,27	2,19	2,11	1,92

5. pielikuma turpinājums

ν_2	ν_1							
	8	9	10	12	15	20	30	>30
19	2,48	2,42	2,38	2,31	2,23	2,16	2,07	1,88
20	2,45	2,39	2,35	2,28	2,20	2,12	2,04	1,84
21	2,42	2,37	2,32	2,25	2,18	2,10	2,01	1,82
22	2,40	2,34	2,30	2,23	2,15	2,07	1,98	1,78
23	2,37	2,32	2,27	2,20	2,13	2,05	1,96	1,76
24	2,36	2,30	2,25	2,18	2,11	2,03	1,94	1,73
25	2,34	2,28	2,24	2,16	2,09	2,01	1,92	1,71
26	2,32	2,27	2,22	2,15	2,07	1,99	1,90	1,69
27	2,31	2,25	2,20	2,13	2,06	1,97	1,88	1,67
28	2,29	2,24	2,19	2,12	2,04	1,96	1,87	1,65
29	2,28	2,22	2,18	2,10	2,03	1,94	1,85	1,64
30	2,27	2,21	2,16	2,09	2,01	1,93	1,84	1,65
40	2,18	2,12	2,08	2,00	1,92	1,84	1,74	1,51
60	2,10	2,04	1,99	1,92	1,84	1,75	1,65	1,39
120	2,02	1,96	1,91	1,83	1,75	1,66	1,55	1,25
	1,94	1,88	1,83	1,75	1,67	1,57	1,46	1,00

Nulles hipotēzi noraida, ja $F > F_{\alpha;\nu 1;\nu 2}$

6. Spirmena rangu korelācijas koeficienta kritiskās vērtības

r_{S_α} ($\alpha = 0{,}05$)

n	r_{S_α}	n	r_{S_α}	n	r_{S_α}
5	0,94	12	0,58	22	0,43
6	0,85	13	0,56	24	0,41
7	0,78	14	0,54	26	0,39
8	0,72	15	0,52	28	0,38
9	0,68	16	0,50	30	0,36
10	0,64	18	0,47	35	0,33
11	0,61	20	0,45	40	0,31

Korelācija ir ticama, ja $r_S > r_{S_\alpha}$

7. X^2-kritērija kritiskās vērtības

(α = 0,05)

ν	$X^2_{0,05;\nu}$	ν	$X^2_{0,05;\nu}$	ν	$X^2_{0,05;\nu}$	ν	$X^2_{0,05;\nu}$
4	9,49	13	22,36	22	33,92	40	55,76
5	11,07	14	23,68	23	35,17	50	67,50
6	12,59	15	25,00	24	36,42	60	79,08
7	14,07	16	26,30	25	37,65	70	90,53
8	15,51	17	27,59	26	38,89	80	101,88
9	16,92	18	28,87	27	40,11	90	113,14
10	18,31	19	30,14	28	41,34	100	124,34
11	19,68	20	31,41	29	42,56		
12	21,03	21	32,67	30	43,77		

Nulles hipotēzi noraida, ja $X^2 < X^2_{0,05;\nu}$

8. Koeficientu $a_{n;k}$ vērtības W-kritērija aprēķināšanai

n / *k*	8	9	10	11	12	13
1	0,60052	0,5888	0,5739	0,05601	0,5475	0,5359
2	0,3164	0,3244	0,3291	0,3315	0,3325	0,3325
3	0,1743	0,1976	0,2141	0,2260	0,2347	0,2412
4	0,0561	0,0947	0,1224	0,1429	0,1585	0,1707
5			0,0399	0,0695	0,0922	0,1099
6					0,0303	0,0539

n / *k*	14	15	16	17	18	19
1	0,5251	0,5150	0,5056	0,4968	0,4886	0,4808
2	0,3318	0,3306	0,3290	0,3273	0,3253	0,3232
3	0,2460	0,2495	0,2521	0,2540	0,2553	0,2561
4	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059
5	0,1240	0,1353	0,1447	0,1524	0,1587	0,1641
6	0,0727	0,0880	0,1005	0,1109	0,1197	0,1271
7	0,0240	0,0433	0,0593	0,0725	0,0837	0,0932
8			0,0196	0,0359	0,0496	0,0612
9					0,0163	0,0303

8. pielikuma turpinājums

n *k*	20	21	22	23	24	25
1	0,4734	0,4643	0,4590	0,4542	0,4493	0,4450
2	0,3211	0,3185	0,3156	0,3126	0,3098	0,3069
3	0,2565	0,2578	0,2571	0,2563	0,2554	0,2543
4	0,2085	0,2119	0,2131	0,2139	0,2145	0,2148
5	0,1686	0,1736	0,1764	0,1787	0,1807	0,1822
6	0,1334	0,1399	0,1443	0,1480	0,1512	0,1539
7	0,1013	0,1092	0,1150	0,1201	0,1245	0,1283
8	0,0711	0,0804	0,0878	0,0941	0,0997	0,1046
9	0,0422	0,0530	0,0618	0,0696	0,0764	0,0823
10	0,0140	0,0263	0,0368	0,0459	0,0539	0,0610
11			0,0122	0,0228	0,0321	0,0403
					0,0107	0,0200

n *k*	26	27	28	29	30	31
1	0,4407	0,4366	0,4328	0,4291	0,4254	0,4420
2	0,3043	0,3018	0,2992	0,2968	0,2944	0,2921
3	0,2533	0,2522	0,2510	0,2499	0,2487	0,2475
4	0,2151	0,2152	0,2151	0,2150	0,2148	0,2145
5	0,1836	0,1848	0,1857	0,1864	0,1870	0,1874
6	0,1563	0,1584	0,1601	0,1616	0,1630	0,1641
7	0,1316	0,1346	0,1372	0,1395	0,1415	0,1433
8	0,1089	0,1128	0,1162	0,1192	0,1219	0,1243
9	0,0876	0,0923	0,0965	0,1002	0,1036	0,1066
10	0,0672	0,0728	0,0778	0,0822	0,0862	0,0899
11	0,0476	0,0540	0,0598	0,0650	0,0697	0,0739
12	0,0284	0,0358	0,0424	0,0483	0,0537	0,0585
13	0,0094	0,0178	0,0253	0,0320	0,0381	0,0435
14			0,0084	0,0159	0,0227	0,0289
15					0,0076	0,0144

9. *W*-kritērija teorētiskās vērtības

(α = 0,05)

n	W_α	n	W_α	n	W_α
8	0,818	16	0,887	24	0,916
9	0,829	17	0,892	25	0,918
10	0,842	18	0,897	26	0,920
11	0,85	19	0,901	27	0,923
12	0,859	20	0,905	28	0.924
13	0,866	21	0,908	29	0,926
14	0,874	22	0,911	30	0,927
15	0,881	23	0,914	31	0,929

Nulles hipotēzi noraida, ja $W\alpha > W$

10. Diksona kritērija teorētiskās vērtības

(α = 0,05)

n	$\tau_{0,05}$	n	$\tau_{0,05}$	n	$\tau_{0,05}$
6	0,689	10	0,477	18	0,349
7	0,610	12	0,428	20	0,334
8	0,554	14	0,395	30	0,283
9	0,512	16	0,369		

Nulles hipotēzi noraida (variante jāatmet), ja $\tau > \tau_{0,05}$

11. Funkcijas Ψ vērtības

$\frac{g_i}{n_1+n_2+1}$	0	1	2	3	4	5	6	7	8	9
0.50	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
0.51	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05
0.52	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07
0.53	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.09	0.10	0.10
0.54	0.10	0.10	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.12
0.55	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.15	0.15
0.56	0.15	0.15	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17
0.57	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.20	0.20
0.58	0.20	0.20	0.21	0.21	0.21	0.21	0.22	0.22	0.22	0.23
0.59	0.23	0.23	0.23	0.24	0.24	0.24	0.24	0.25	0.25	0.25
0.60	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28
0.61	0.28	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30
0.62	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.33	0.33
0.63	0.33	0.33	0.34	0.34	0.34	0.35	0.35	0.35	0.35	0.36
0.64	0.36	0.36	0.36	0.37	0.37	0.37	0.37	0.38	0.38	0.38
0.65	0.39	0.39	0.39	0.39	0.40	0.40	0.40	0.40	0.41	0.41
0.66	0.41	0.42	0.42	0.42	0.42	0.43	0.43	0.43	0.43	0.44
0.67	0.44	0.44	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.46
0.68	0.47	0.47	0.47	0.48	0.48	0.48	0.48	0.49	0.49	0.49
0.69	0.50	0.50	0.50	0.50	0.51	0.51	0.51	0.52	0.52	0.52
0.70	0.52	0.53	0.53	0.53	0.54	0.54	0.54	0.54	0.55	0.55
0.71	0.55	0.56	0.56	0.56	0.57	0.57	0.57	0.57	0.58	0.58
0.72	0.58	0.59	0.59	0.59	0.59	0.60	0.60	0.60	0.61	0.61
0.73	0.61	0.62	0.62	0.62	0.63	0.63	0.63	0.63	0.64	0.64
0.74	0.64	0.65	0.65	0.65	0.66	0.66	0.66	0.67	0.67	0.67

11. pielikuma turpinājums

$\frac{g_i}{n_1+n_2+1}$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.75	0.67	0.68	0.68	0.68	0.69	0.69	0.69	0.70	0.70	0.70
0.76	0.71	0.71	0.71	0.72	0.72	0.72	0.73	0.73	0.73	0.74
0.77	0.74	0.74	0.75	0.75	0.75	0.76	0.76	0.76	0.77	0.77
0.78	0.77	0.78	0.78	0.78	0.79	0.79	0.79	0.80	0.80	0.80
0.79	0.81	0.81	0.81	0.82	0.82	0.82	0.83	0.83	0.83	0.84
0.80	0.84	0.85	0.85	0.85	0.86	0.86	0.86	0.87	0.87	0.87
0.81	0.88	0.88	0.89	0.89	0.89	0.90	0.90	0.90	0.91	0.91
0.82	0.92	0.92	0.92	0.93	0.93	0.93	0.94	0.94	0.95	0.95
0.83	0.95	0.96	0.96	0.97	0.97	0.97	0.98	0.98	0.99	0.99
0.84	0.99	1.00	1.00	1.01	1.01	1.02	1.02	1.02	1.03	1.03
0.85	1.04	1.04	1.05	1.05	1.05	1.06	1.06	1.07	1.07	1.08
0.86	1.08	1.09	1.09	1.09	1.10	1.10	1.11	1.11	1.12	1.12
0.87	1.13	1.13	1.14	1.14	1.15	1.15	1.16	1.16	1.17	1.17
0.88	1.18	1.18	1.19	1.19	1.20	1.20	1.21	1.21	1.22	1.22
0.89	1.23	1.23	1.24	1.24	1.25	1.25	1.26	1.26	1.27	1.28
0.90	1.28	1.29	1.29	1.30	1.30	1.31	1.32	1.32	1.33	1.33
0.91	1.34	1.35	1.35	1.36	1.37	1.37	1.38	1.39	1.39	1.40
0.92	1.41	1.41	1.42	1.43	1.43	1.44	1.45	1.45	1.46	1.47
0.93	1.48	1.48	1.49	1.50	1.51	1.51	1.52	1.53	1.54	1.55
0.94	1.55	1.56	1.57	1.58	1.59	1.60	1.61	1.62	1.63	1.64
0.95	1.64	1.65	1.66	1.67	1.68	1.70	1.71	1.72	1.73	1.74
0.96	1.75	1.76	1.77	1.79	1.80	1.81	1.83	1.84	1.85	1.87
0.97	1.88	1.90	1.91	1.93	1.94	1.96	1.98	2.00	2.01	2.03
0.98	2.05	2.07	2.10	2.12	2.14	2.17	2.20	2.23	2.26	2.29
0.99	2.33	2.37	2.41	2.46	2.51	2.58	2.65	2.75	2.88	3.09