

ივანე ჯავახიშვილის სახელობის
თბილისის სახელმწიფო უნივერსიტეტი

ქეთევან მანჯგალაძე, გრიგოლ სოხაძე
ომარ ფურთუხია

გამოყენებითი სტატისტიკის საფუძვლები



წინამდებარე სახელმძღვანელო წარმოადგენს ავტორების მიერ წლების განმავლობაში ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტში სტატისტიკაში წაკითხული ლექციების კურსის გადამუშავებულ ვარიანტს. ის მოიცავს აღწერითი (დესკრიფციული) სტატისტიკის, ალბათობის თეორიის და სტატისტიკური დასკვნების თეორიის მნიშვნელოვან ასპექტებს. მასში თვალსაჩინოადაა გადმოცემული ის ძირითადი საკითხები, რომლებიც ყველაზე უფრო ხშირად გვხვდება თანამედროვე გამოყენებით კვლევებში. წიგნი არაა გადატვირთული მათემატიკური დებულებების დამტკიცებებით. ძირითადი აქცენტი გადატანილია შინაარსობრივ მხარეზე და სამოტივაციო და საილუსტრაციო მაგალითებზე. წიგნით პირველადი სარგებლობისათვის სრულიად საკმარისია საშუალო სკოლის მათემატიკის ცოდნა. სახელმძღვანელო შედგენილია თბილისის სახელმწიფო უნივერსიტეტში ამჟამად მოქმედი სასწავლო პროგრამებისა და სილაბუსების მიხედვით ისეთ დისციპლინებში, როგორიცაა: „სტატისტიკა 1“, „სტატისტიკა 2“, „სტატისტიკა არქეოლოგიებისათვის“, „სამეცნიერო მონაცემთა ანალიზი“, „მათემატიკური სტატისტიკა ეკონომისტიებისათვის“, „გამოყენებითი სტატისტიკა“, „კვლევის სტატისტიკური მეთოდები“.

სახელმძღვანელო განკუთვნილია ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის სტატისტიკის (ფართო გაგებით) შემსწავლელი სამივე საფეხურის სწავლების სტუდენტებისათვის. ის დახმარებას გაუწევს სტატისტიკის გამოყენებითი ასპექტებით დაინტერესებულ ნებისმიერ მკითხველს.

რედაქტორები: საქართველოს მეცნიერებათა ეროვნული აკადემიის
წევრ-კორესპონდენტი, თსუ სრული პროფესორი ე. ნადარაია;
თსუ ასოცირებული პროფესორი, ფიზიკა-მათემატიკის
მეცნიერებათა დოქტორი ო. ლლონტი

რეცენზენტები: თსუ ასოცირებული პროფესორი, ფიზიკა-მათემატიკის
მეცნიერებათა დოქტორი ბ. დოჭვირი;
საქართველოს ტექნიკური უნივერსიტეტის სრული პროფესორი,
ტექნიკურ მეცნიერებათა დოქტორი რ. კაკუბავა

© ივანე ჯავახიშვილის სახელობის თბილისის
სახელმწიფო უნივერსიტეტის გამომცემლობა, 2015

ISBN 978-9941-13-397-8

შესავალი	7
სტატისტიკური მეთოდოლოგია. სტატისტიკური მეთოდების დანიშნულება. აღწერა და დასკვნა. პოპულაცია და შერჩევა. დასკვნითი სტატისტიკური მეთოდები. პარამეტრები და სტატისტიკა. პოპულაციის განსაზღვრა. კომპიუტერის როლი სტატისტიკაში. რატომ არის საჭირო სტატისტიკა. შეჯამება. კითხვები.	
თავი 1. შერჩევა და გაზომვა	13
ცვლადები და მათი გაზომვა. ცვლადი. თვისებრივი და რაოდენობრივი მონაცემები. ინტერვალების სკალა. რიგის სკალა. რიგის სკალის მონაცემთა რაოდენობრივი ბუნება. სტატისტიკური მეთოდები და გაზომვის ტიპი. დისკრეტული და უწყვეტი ცვლადები. შერჩევის სახეობები. შერჩევის შეცდომა. მარტივი შემთხვევითი შერჩევა. ალბათური და არალბათური შერჩევა. ცდომილება. საიმედოობა. შეჯამება. კითხვები.	
თავი 2. აღწერითი სტატისტიკა. ცხრილები და გრაფიკები	22
რაოდენობრივი მონაცემების სიხშირეთა განაწილება და მისი გრაფიკული წარმოდგენა. სიხშირეთა განაწილება. სიხშირეთა განაწილების გრაფიკულად გამოსახვა. პისტოგრამები. სიხშირეთა პოლიგონი. თვისებრივი მონაცემების სიხშირეთა განაწილება. სექტორული და წრიული დიაგრამები. ფოთლებიანი ღეროების მსგავსი დიაგრამები. პოპულაციის და შერჩევის განაწილება. შეჯამება. კითხვები.	
თავი 3. აღწერითი სტატისტიკები. ცენტრალური ტენდენციის საზომები	40
აღწერითი სტატისტიკები. პარამეტრები. სტატისტიკები. საშუალო. დამრგვალების ზოგადი წესი. საშუალოს დამრგვალების წესი. საშუალოს თვისებები. მედიანა. მედიანის შეფასება. მედიანის თვისებები. მოდა. მოდალური ინტერვალი. შეჯამება. კითხვები.	
თავი 4. მონაცემთა გაფანტულობის საზომები	59
გაბნევის დიაპაზონი. პროცენტილი. კვარტილები. კვარტილთშორისი გაბნევის დიაპაზონი. რანგები და პროცენტული რანგები. დისპერსია და სტანდარტული გადახრა. შერჩევითი დისპერსია. შესწორებული შერჩევითი დისპერსია. სტანდარტული გადახრა. შესწორებული სტანდარტული გადახრა. ვარიაციის კოეფიციენტი. ემპირიული წესი. ჩეიბშეის თეორემა. ბოქსპლოტი. შეჯამება. კითხვები.	
თავი 5. დანყილებული მონაცემები	83
დამოუკიდებელი სიდიდე. დამოკიდებული სიდიდე. გაბნევის დიაგრამა. მისადაგების წირი. რეგრესიული ანალიზი. უმცირეს კვადრატთა მეთოდი. კორელაციის კოეფიციენტი. კოვარიაციის კოეფიციენტი. სპირმენის რანგობრივი კორელაციის კოეფიციენტი. რეგრესია. რეგრესიის წრფე. ნიანსნარმეტყველება (პროგნოზირება). დეტერმინაციის კოეფიციენტი. შეჯამება. კითხვები.	
თავი 6. ალბათობა	106
რა არის ალბათობა? ალბათობის თვისებები. დენდროგრამა (ხისებრი დიაგრამა). შემთხვევითი სიდიდეები და მათი განაწილებები. განაწილების კანონი. ჯამის ალბათობის ფორმულა. პირობითი ალბათობა. ნამრავლის ალბათობის ფორმულა. სრული ალბათობის ფორმულით. ბაიესის ფორმულა. შეჯამება. კითხვები.	
თავი 7. პროპორციები და ბინომური განაწილება	118
ბინომური განაწილება. ბინომური განაწილების თვისებები. ბინომური ალბათობების გამოთვლა. ბერნულის ფორმულა. უალბათესი რიცხვი. ბინომური შემთხვევითი სიდიდის საშუალო და დისპერსია. ნორმალური მიახლოება. უაუსონის მიახლოება. ფარდობითი სიხშირები. შერჩევის ფარდობითი სიხშირის საშუალო და დისპერსია. შეჯამება. კითხვები.	
თავი 8. ნორმალური განაწილება	126
ნორმალური განაწილების თვისებები. სტანდარტული ნორმალური განაწილების ცხრილის გამოყენება. Z სიდიდე. მოცემული ალბათობით Z სიდიდის პოვნა. სტანდარტული	

ნორმალური განაწილების წირის ქვეშ მდებარე არის ფართობის მოსაძებნი პროცედურული ცხრილი. ცენტრალური ზღვარითი თეორემა (ცზთ). რამდენად დიდი უნდა იყოს შერჩევის მოცულობა n , რომ სრულდებოდეს ცზთ? შერჩევითი განაწილებები. ბინომური განაწილების ნორმალური აპროქსიმაცია. ცენტრალური ზღვარითი თეორემა და სიხშირეები. p -სიდიდეები. შეჯამება. კითხვები.

თავი 9. სტატისტიკური დასკვნების თეორია 143

ნდობის ინტერვალი პოპულაციის საშუალოსთვის. ჰიპოთეზათა შემოწმება საშუალოსთვის. ჰიპოთეზების ფორმულირება. კრიტერიუმის სტატისტიკის მნიშვნელობის გამოთვლა. უარყოფის არის დადგენა. კრიტერიუმის მნიშვნელოვნების დონე. გადაწყვეტილების მიღება. p -სიდიდეების გამოყენება ჰიპოთეზათა შემოწმებაში. პირველი და მეორე გვარის შეცდომები სტატისტიკური კრიტერიუმების გამოყენებისას. ჰიპოთეზათა შემოწმება ნორმალური პოპულაციის საშუალოსათვის ($n < 30$). ნდობის ინტერვალი დისპერსიისათვის. ჰიპოთეზათა შემოწმება (პოპულაცია ნორმალურია, $n < 30$). ჰიპოთეზათა შემოწმება პოპულაციის პროპორციისათვის ($n \geq 30$). ჰიპოთეზათა შემოწმება პოპულაციის პროპორციისათვის, როცა შერჩევის მოცულობა პატარაა. სასრული პოპულაციები. ცდომილება, სიზუსტე და შერჩევის მოცულობა. ჰიპოთეზათა შემოწმება დისპერსიისათვის. შეჯამება. კითხვები.

თავი 10. ორამოკრეფიანი ამოცანები 172

საშუალოთა შორის განსხვავების ჰიპოთეზის შემოწმება დიდი მოცულობის მქონე შერჩევებისას. ჰიპოთეზათა შემოწმება $\mu_1 - \mu_2$ -ის შესახებ. ორი დამოუკიდებელი პოპულაციის საშუალოთა შორის განსხვავების ჰიპოთეზათა შემოწმება (ორივე, ან ერთ-ერთი შერჩევის მოცულობა ნაკლებია 30-ზე). $\mu_1 - \mu_2$ -ის ნდობის ინტერვალი მცირე მოცულობის შერჩევებისათვის. ჰიპოთეზათა შემოწმება ორი დამოუკიდებელი ნორმალური პოპულაციის დისპერსიის შესახებ. ჰიპოთეზათა შემოწმება დანეცილებულ მონაცემთა საშუალოების სხვაობებისათვის. ორამოკრეფიანი ამოცანები ორი ალბათობის შედარებისას. ტესტი პროპორციათა სხვაობისთვის. შეჯამება. კითხვები.

თავი 11. დისპერსიული ანალიზი 188

ერთფაქტორიანი დისპერსიული ანალიზი (დისპირთა შორის სქემისათვის (ჰიპოთეზები). ცდისპირთა შიდა ორფაქტორიანი სქემა და დისპერსიული ანალიზი. ცდისპირთა შორის ორფაქტორიანი სქემა და დისპერსიული ანალიზი. შეჯამება. კითხვები.

თავი 12. არამარამეტრული სტატისტიკური კრიტერიუმები 211

ხი-კვადრატ (χ^2) თანხმობის კრიტერიუმი. დამოუკიდებლობის ხი-კვადრატ კრიტერიუმი. ჰიპოთეზები. ჰიპოთეზათა შემოწმება 2×2 შეუღლების ცხრილის შემთხვევაში. შანსების შეფარდება. შანსების შეფარდების თვისებები 2×2 შეუღლების ცხრილებისათვის. მან-უიტნის კრიტერიუმი. უილკოქსონის ნიშნის რანგების კრიტერიუმი. კრუსკალ-უოლისის კრიტერიუმი. შეჯამება. კითხვები.

თავი 13. რეგრესია და წინასწარმეტყველება (მარტივი რეგრესიის მოდელი) 239

დასკვნები კორელაციის კოეფიციენტის შესახებ. ჰიპოთეზის შემოწმება დახრილობის შესახებ. ჰიპოთეზის შემოწმება თანაკეთის შესახებ. დასკვნები კორელაციის კოეფიციენტის შესახებ. საპროგნოზო ინტერვალები. დეტერმინაციის კოეფიციენტი. მოდელის ვარისიანობის შემოწმება. R^2 -ის თვისებები. რეგრესიული ანალიზი მრუდწირული (არანრეფი) კავშირისთვის. მრავლობითი რეგრესია და კორელაცია. კორელაციური მატრიცა. ფიქტიური ცვლადები რეგრესიულ ანალიზში. შეჯამება. კითხვები.

თავი 14. ხშირად გამოყენებული ფორმულები. შეფასების და ჰიპოთეზათა

შემოწმების ამოცანათა მიმოხილვა 264

დანართი 1. EXCEL-ის სტატისტიკური ფუნქციების აღწერა 273

დანართი 2. სტატისტიკური ცხრილები 285

პასუხები 334

ლიტერატურა 339

საგნობრივი საძიებელი 340

შესავალი

სტატისტიკური მეთოდოლოგია

სტატისტიკა მნიშვნელოვანი იარაღია არა მარტო მათთვის, ვისაც ის ყოველდღიურად სჭირდება, არამედ მათთვისაც, ვისთვისაც ინფორმაციის უკეთ გააზრებაა აუცილებელი. სტუდენტი, რომელიც ეუფლება სოციალურ მეცნიერებებს, აუცილებლად უნდა გაეცნოს ბაზისურ სტატისტიკურ კურსს.

ინფორმაციის შეგროვება მეცნიერების ყველა დარგის საფუძველია. მეცნიერები იყენებენ ინფორმაციის შეგროვების სხვადასხვა ტექნიკას. გამოკითხვის, გაზომვის, დაკვირვების შედეგებს ზოგადად უწოდებენ მონაცემებს. მონაცემების გადამუშავებით ადგენენ პოპულაციის მახასიათებლებს. მაგალითად, თუ განვიხილავთ ერთი და იმავე ასაკის, სქესის და დანაშაულებრივი ისტორიის მქონე დამნაშავეების მონაცემებს, ვთქვათ, დამნაშავეების ციხიდან განთავისუფლებისა და მათ მიერ დანაშაულის ხელახლა ჩადენამდე დროს, სტატისტიკური მეთოდების მეშვეობით შეიძლება დადგინდეს, როგორ არის ეს დრო დამოკიდებული მისჯილი პატიმრობის ტიპზე.

სტატისტიკა გულისხმობს სხვადასხვა მეთოდის კრებულს მონაცემების შესაგროვებლად და გასაანალიზებლად. ის გვეხმარება სამყაროს ობიექტურ შეფასებაში.

სტატისტიკური მეთოდების დანიშნულება

სტატისტიკური მეთოდოლოგია

სტატისტიკა გვანდის მეთოდებს:

ა) **დაგეგმვისთვის** (დინამიკისათვის) – როგორ დაგეგმოთ და განვახორციელოთ კვლევები. რა საშუალებებით უნდა მივიღოთ კვლევისთვის სასურველი მონაცემები, როგორ წარვმართოთ კვლევა. ის შეიძლება გულისხმობდეს კითხვარების კონსტრუირებისა და კვლევაში მონაწილე ადამიანების შერჩევის პროცესსაც.

აღწერა და დასკვნა არის სტატისტიკური ანალიზის ორი ძირითადი ელემენტი. ჩვენ ყურადღებას გავამახვილებთ სტატისტიკურ ანალიზზე. ეს, რა თქმა უნდა, არ ნიშნავს იმას, რომ დინამიკი არ არის მნიშვნელოვანი, პირიქით, რაც არ უნდა ზუსტი იყოს სტატისტიკური ანალიზი, თუ კვლევა არასწორად არის დაგეგმილი ან მონაცემები არასათანადოდია შეგროვებული და ჩანერგილი, დასკვნებიც მცდარი იქნება.

ბ) **აღწერისთვის** – მონაცემების აღწერა და გამოკვლევა გულისხმობს მონაცემების შეჯამების სხვადასხვა გზას და იმ საზომების გამოყენებას, რომელიც უფრო მარტივი და ხელმისაწვდომია მკვლევრისთვის.

აღწერის მთავარი დანიშნულებაა გამოიყოს მნიშვნელოვანი მოპოვებულ ნედლ მონაცემებში და მარტივი და გასაგები ფორმით წარმოადგინოს ის.

ნედლი მონაცემები არის თითოეული მონაწილისგან მიღებული პასუხები ან გაზომვის შედეგები დამატებითი ანალიზის გარეშე.

სტატისტიკური ინფორმაციის მისაღებად ყველა ნედლი მონაცემი ჯამდება და მათი გამოსახვა ხდება რაოდენობრივად ცხრილებში ან გრაფიკული ფორმით. ასეთნაირად მიღებულ რიცხვებს *სტატისტიკები* ეწოდება.

მონაცემების შეგროვების, ორგანიზების, კლასიფიკაციის, სისტემატიზაციისა და პირველადი დამუშავების მეთოდებს *აღწერითი სტატისტიკის* უწოდებენ. მიღებულ სტატისტიკებს (შეფასებებს) უფრო მეტი მნიშვნელობა აქვს პოპულაციის პარამეტრების შესაფასებლად, ვიდრე მონაცემთა პირვანდელ სრულ სიას.

გ) დასკვნისთვის – დასკვნითი სტატისტიკა შეიცავს მეთოდებს, თუ როგორ უნდა გაეკეთოთ დასკვნა პოპულაციის მახასიათებლების შესახებ მიღებული აღწერითი სტატისტიკების საფუძველზე. მაგალითად, უკანასკნელი კვლევების მიხედვით, რომელიც ერთ-ერთი ორგანიზაციის მიერ ჩატარდა 150 თბილისელზე, გაირკვა რომ მათ 14%-ს სჯეროდა, რომ ფულით შესაძლებელია ბედნიერების ყიდვა. შეგვიძლია თუ არა ამ ინფორმაციის საფუძველზე გავაკეთოთ პროგნოზი: – თბილისის მთლიანი მოსახლეობის, ანუ 1.5 მილიონი თბილისელის რამდენ პროცენტს სჯერა ეს? არსებობს ამის გამოთვლის მეთოდი, რომლის საშუალებითაც ხდება მცირე ჯგუფიდან მიღებული ინფორმაციის განზოგადება უფრო დიდ ჯგუფზე.

აღწერა და დასკვნა

სტატისტიკური ანალიზის მეთოდების კლასიფიცირება დამოკიდებულია კვლევის მიზანზე. ამ განსხვავებების უფრო დანერგვით ასახსნელად განვსაზღვროთ რა არის პოპულაცია და შერჩევა.

პოპულაცია და შერჩევა

სამიზნე ჯგუფს, რომელზეც უნდა ჩატარდეს კვლევა, კვლევის სუბიექტი ეწოდება. ჩვეულებრივ, სუბიექტები ან ცდისპირები არიან ადამიანები, მაგრამ შეიძლება ასევე იყოს ოჯახები, სკოლები, ქალაქები, კომპანიები, ცხოველები, გამოშვებული პროდუქცია და ა.შ.

პოპულაცია. პოპულაცია არის ჩვენი სამიზნე ჯგუფის სრული შემადგენლობა.

შერჩევა. შერჩევა არის პოპულაციიდან გამოყოფილი ქვეჯგუფი, რომელიც გარკვეული წესების მიხედვითაა დალაგებული.

ნებისმიერი კვლევის მთავარი მიზანია პოპულაციის მახასიათებლების შესწავლა, ამის გაკეთება კი უფრო პრაქტიკული და ეკონომიურია ამ პოპულაციის შერჩევაზე დაყრდნობით.

აღწერითი სტატისტიკა. აღწერითი სტატისტიკა საშუალებას გვაძლევს დავადგინოთ შერჩევის ძირითადი მახასიათებლები. მაგალითად, ჩვენ შეგვიძლია აღწეროთ თბილისში მცხოვრები რაიმე შერჩევის ოჯახის წევრების ტიპური რაოდენობა, ტიპური შემოსავალი, სტუდენტების ჯგუფის მიერ სტატისტიკაში მიღებული ტიპური ქულა და ა.შ.

აღწერითი სტატისტიკის მთავარი დანიშნულებაა შერჩევის მონაცემების სრული სიის გამოკვლევა და ამ მონაცემთა ერთობლიობის თვალსაჩინოდ და კომპაქტურად წარმოდგენა. გავითვალისწინოთ, რომ ეს პროცედურა აუცილებლად უნდა ჩატარდეს ინფორმაციის დამახინჯებისა და დაკარგვის გარეშე.

დასკვნითი სტატისტიკური მეთოდები

დასკვნითი სტატისტიკური მეთოდები – განზოგადების მეთოდებია შერჩევიდან პოპულაციამდე.

განვიხილოთ მაგალითი: ერთ-ერთ მკვლევარს ფლორიდაში, კრიმინალური შემთხვევების მაღალი მაჩვენებლის გამო, სურდა გაეგო, თუ ფლორიდის მცხოვრებთა

რამდენი პროცენტი უჭერდა მხარს იარაღის გაყიდვაზე კონტროლის დაწესებას. პოპულაციის რაოდენობა ამ შემთხვევაში იყო 10 მილიონზე მეტი. ცხადია, თითოეული მათგანის გამოკითხვა შეუძლებელი იქნებოდა. ამიტომაც, მკვლევარმა გადაწყვიტა გამოეკითხა 800 რესპოდენტი და დასკვნა გამოეტანა ამ შერჩევიდან მიღებული შედეგების საფუძველზე. დადგინდა, რომ გამოკითხულთა 54% მხარს უჭერდა იარაღის გაყიდვაზე კონტროლის დაწესებას. ცხადია, მკვლევარს ანტიერესებდა არა მხოლოდ 800 ფლორიდელის, არამედ მთლიანად ფლორიდის მცხოვრებთა აზრი. დასკვნითი სტატისტიკა კი სწორედ იმის შესაძლებლობას გვაძლევს, რომ შერჩევის მონაცემების დამუშავების შედეგად გაკეთდეს დასკვნა მთლიან პოპულაციაზე. სათანადო სტატისტიკური პროცედურების ჩატარების შემდეგ (რომელსაც შემდეგ ლექციებში გავივლით) ვნახავთ, რომ პოპულაციის 51-57% დაუჭერდა მხარს იარაღის გაყიდვაზე კონტროლის დაწესებას.

დასკვნითი სტატისტიკური მეთოდების საშუალებით, თუ შერჩევა პოპულაციიდან სათანადო წესების დაცვითაა მიღებული, შეიძლება ამ შერჩევაზე დაყრდნობით განისაზღვროს მთლიანი პოპულაციის მახასიათებლები.

პარამეტრები და სტატისტიკა

პარამეტრი. პარამეტრი არის რიცხობრივად გამოხატული პოპულაციის მახასიათებელი.

სტატისტიკა. სტატისტიკა (სტატისტიკური სიდიდე) არის რიცხობრივად გამოხატული შერჩევის მახასიათებელი.

კვლევების უპირველესი მიზანია პოპულაციის პარამეტრების განსაზღვრა. შერჩევა და სტატისტიკური აღწერა მხოლოდ იმდენადაა მნიშვნელოვანი, რამდენადაც ის გვაძლევს ინფორმაციას პოპულაციის უცნობი მახასიათებლების შესახებ. იმ შემთხვევაში, როდესაც არსებობს პოპულაციის სრული შემადგენლობის მონაცემები, არ არის საჭირო სტატისტიკური დასკვნითი მეთოდების გამოყენება. მაგალითად, როდესაც გვაქვს მცირე ზომის პოპულაცია ან როდესაც არჩევნების ხმები უნდა დავითვალოთ. მაგრამ, ძალიან ბევრი სოციალური კვლევების მიზნებიდან გამომდინარე, საკმაოდ არაპრაქტიკულია და ხანდახან შეუძლებელია მთლიანი პოპულაციის მონაცემების შეგროვება როგორც მატერიალური, ადამიანური, ასევე დროის შეზღუდული რესურსების გამო.

პოპულაციის პარამეტრების შესახებ დასკვნის გაკეთება, სტატისტიკური მეთოდების გამოყენებით, შესაძლებელია შედარებით მცირე ჯგუფიდანაც.

პოპულაციის განსაზღვრა

დასკვნითი სტატისტიკა ითხოვს ზუსტად დადგინდეს საკვლევი პოპულაცია და ის, თუ რა სახის დასკვნები გვაინტერესებს.

ხანდახან პოპულაციაში შემავალი სუბიექტების რაოდენობა განსაზღვრულია. მაგალითად, საქართველოში მცხოვრები მოზარდების სრული რაოდენობა ცნობილია. ხშირად განზოგადება ეხება კონცეპტუალურ პოპულაციას ანუ პოპულაციას, რომელიც სინამდვილეში არ არსებობს, მაგრამ რომლის პიპოთეტურად წარმოდგენა შესაძლებელია. მაგალითად, განვიხილოთ მკვლევართა ჯგუფი, რომელიც ცდის ახალ აბებს დეპრესიის სანინაჟმდგოდ. მათ, დეპრესიაში მყოფი პაციენტების შერჩევიდან მიღებული მონაცემების სტატისტიკური ანალიზის საფუძველზე, უნდა გააკეთონ დასკვნა ყველა იმ ინდივიდზე, რომელიც ახლა იმყოფება ან მომავალში შეიძლება აღმოჩ-

ნდეს დეპრესიაში. ამ შემთხვევაში დასკვნა კეთდება საყარაუდო, შესაძლო პოპულაციაზე, რომელიც რეალურად არც ვიცით რა რაოდენობას მოიცავს.

მკვლევრები ხშირად ცდილობენ შერჩევის შედეგების უფრო ფართო პოპულაციაზე განზოგადებას, რამაც შეიძლება არასწორ დასკვნებამდე მიგვიყვანოს, რადგანაც შერჩევა შეიძლება პოპულაციისგან განსხვავებულად იყოს ფუნდამენტური მახასიათებლებით, როგორიცაა რასა, სოციალური ან ეკონომიკური სტატუსი და სხვა.

კომპიუტერის როლი სტატისტიკაში

კომპიუტერის როლი მნიშვნელოვანია არა მარტო სტატისტიკაში, არამედ თითოეული ჩვენთვის ცხოვრებაში. ის უფრო და უფრო მისაღები ხდება იმ ადამიანებისთვისაც, რომელთაც არა აქვთ სათანადო ტექნიკური ცოდნა. გამოთვლითი ტექნიკის ასე ფართოდ გამოყენების მიზეზია ის, რომ კომპიუტერული პროგრამები უმაღლეს დონეზეა სპეციალიზებული და მათი გამოყენებისათვის სრულიად არ არის საჭირო ტექნიკურ საკითხებში ღრმად გარკვევა.

სტატისტიკური კომპიუტერული პროგრამებიდან ფართოდ გამოყენებადია SPSS და SAS. რა თქმა უნდა, უფრო მარტივია ამ პროგრამების საშუალებით მონაცემების დამუშავება, ვიდრე მათი ხელით დათვლა. ამით თავიდან იცილებენ ძალიან ბევრ შეცდომას და დროსაც მნიშვნელოვნად ზოგავენ განსაკუთრებით მაშინ, როდესაც დიდი რაოდენობის მონაცემებთან აქვთ საქმე.

კომპიუტერული პროგრამების გამოყენებას ბევრი უპირატესობა აქვს, მაგრამ, ამავე დროს, შეიძლება ნაკლოვანებაც ახლდეს. იმ შემთხვევაში, თუ ჩვენ არ ვართ სტატისტიკაში სათანადოდ გაცნობიერებულნი, შეიძლება არასწორი სტატისტიკური მეთოდები გამოვიყენოთ და არასათანადო დასკვნები გავაკეთოთ. არსებითია, რომ სანამ გამოვიყენებდეთ კომპიუტერულ პროგრამებს, ჯერ კარგად უნდა გავიგოთ სტატისტიკური მეთოდების არსი.

სტატისტიკური პროგრამების გამოყენება არ არის იმის გარანტია, რომ ანალიზის შედეგი ზუსტი იქნება. უნდა ვიცოდეთ, რა მიზნით, რა სახის მონაცემებისთვის რომელი მეთოდები შევარჩიოთ და შეგვეძლოს მიღებული შედეგების სწორი ინტერპრეტირება.

რატომ არის საჭირო სტატისტიკა

სამყაროს ცვალებადობის არარსებობის პირობებში სრულიად მოხსნილი იქნებოდა სტატისტიკური ანალიზის საჭიროება. ჩვენი ცვალებადობით სავსე ცხოვრების აღწერა კი საჭიროებს სტატისტიკური მეთოდების გამოყენებას მოვლენათა კანონზომიერების დასადგენად და შემდგომი დასკვნებისა და პროგნოზის გასაკეთებლად.

შეჯამება

სტატისტიკა ეფუძნება მეთოდებს, რომლებიც გამოიყენება მონაცემთა შესაგროვებლად, აღსაწერად, ანალიზისა და დასკვნების გამოსატანად.

სტატისტიკა აღწერს შერჩევას, პარამეტრი კი პოპულაციას.

სტატისტიკის აღწერითი მეთოდები გამოიყენება შერჩევის შესაჯამებლად, დასკვნითი მეთოდები კი – დასკვნების და პროგნოზის გასაკეთებლად.

ძირითადი ტერმინები

სტატისტიკა (მეცნიერება), აღწერითი სტატისტიკა, დასკვნითი სტატისტიკა, პოპულაცია, შერჩევა, სტატისტიკა (სტატისტიკური სიდიდე), პარამეტრი.

კითხვები

1. 12-13 წლის 50 000 თბილისელი ბავშვიდან შემთხვევით შერჩეული იყო 100 ბავშვი. მათ შეავსებინეს კითხვარი, რომლის ერთ-ერთი კითხვა იყო – „გუშინ რამდენი საათი გაატარეთ კომპიუტერთან?“
 - ა) ამ კვლევაში რას წარმოადგენს პოპულაცია?
 - ბ) ეთქვათ, გამოკითხული 100 ბავშვიდან 5%-მა მიუთითა 3 საათი. აღნიშნული დრო არის სტატისტიკა თუ პარამეტრი?
2. სოციალური კვლევის მიზანი იყო გასული საუკუნის 50-60-იან წლებში მოცემულ ქალაქში დაედგინათ ქალიშვილების გათხოვების საშუალო ასაკი. ქალაქის ერთ-ერთი რაიონის საარქივო მასალებზე დაყრდნობით გათხოვების საშუალო ასაკი გამოვიდა 24.5 წელი. სტატისტიკური მეთოდების გამოყენებით დადგინდა, რომ ამ ქალაქში, 50-60-იან წლებში, გათხოვების საშუალო ასაკი 23.9 წელსა და 25.1 წელს შორის იყო.
 - ა) კვლევის რა ნაწილი იყო აღწერითი?
 - ბ) კვლევის რა ნაწილი იყო დასკვნითი?
 - გ) რა პოპულაციაზე იყო გაკეთებული დასკვნა?

თავი 1

შერჩევა და გაზომვა

სტატისტიკური კვლევის მთავარი მიზანია ინფორმაციის მიღება საკვლევი მოვლენის შესახებ, ამ ინფორმაციის გაგება, ახსნა და დასკვნების გამოტანა. ამისათვის კი საჭიროა გვექონდეს მონაცემები.

როგორ მიიღება მონაცემები ისეთ აბსტრაქტულ კონცეფციებზე, როგორიცაა: სიყვარული, დაზიანება, შფოთვა, აგრესიულობა, გაღიზიანება? ამისათვის საჭიროა ჩავატაროთ გარკვეული გაზომვები და, შესაბამისად, უნდა განვიხილოთ თვით გაზომვის სტატისტიკური ასპექტები. ჯერ შემოვიღოთ ცვლადის ცნება.

ცვლადები და მათი გაზომვა

სტატისტიკური მეთოდები გვიჩვენებს გზებს, როგორ გავუმკლავდეთ ვარიანტობას (ცვალებადობას), რომელიც ყოველ ნაბიჯზე გვხვდება. მაგალითად, პიროვნებათშორისი განსხვავებულობა შემოსავლების, ოჯახური მდგომარეობის, მუსიკალური ნიჭის, IQ-ს, პოლიტიკური კუთვნილების და ა.შ. მიხედვით. ცვალებადობის ბუნება მნიშვნელოვან გავლენას ახდენს როგორც აღწერითი, ასევე დასკვნითი სტატისტიკის მეთოდებზე.

ცვლადი

თითოეულ მახასიათებელს, რომელიც იცვლება შერჩევაში ან პოპულაციაში, *ცვლადი* ეწოდება. ცვლადების მაგალითებია: სქესი, ასაკი, ოჯახში ბავშვების რაოდენობა, აგრესიულობა, ტემპერატურა და ა.შ.

თითოეულ ცდის, პირს (საგანს) აქვს თავისი ცალკეული მნიშვნელობა ცვლადისათვის. ცვლადის შესაძლო მნიშვნელობა იზომება სათანადო სკალაზე. მაგალითად, სქესი იზომება სქესის სკალაზე, რომელსაც ორი დონე აქვს: მდედრობითი და მამრობითი. ოჯახში ბავშვების რაოდენობა იზომება სკალაზე, რომლის შესაძლო სიდიდეებია: 0, 1, 2, 3 და ა.შ.

სტატისტიკური მეთოდების გამოყენება, ცვლადების ანალიზისათვის, დამოკიდებულია გაზომვის სკალაზე. განსხვავებულად ვეპყრობით რაოდენობრივი მნიშვნელობის მქონე ცვლადებს (ყოველთვიური შემოსავალი ლარში, ასაკი, დრო), ვიდრე ცვლადებს დონეების შემცველ სკალაზე (პოლიტიკური კუთვნილების სკალაზე, შფოთვის სკალაზე). ასეთ სკალებს მეცნიერები სპეციალურად ქმნიან.

თვისებრივი და რაოდენობრივი მონაცემები

მონაცემს ეწოდება *თვისებრივი*, როცა საზომი სკალა მოიცავს არაშედარებად, არათანმიმდევრულ კატეგორიებს. მაგალითად, ოჯახური მდგომარეობის

აღწერისას გვაქვს ასეთი კატეგორიები: მარტოხელა, დაქორწინებული, დაშორებული, ქერივი. თვისებრივს განეკუთვნება: სქესი, ფერი, საცხოვრებელი ადგილი, პოლიტიკურ პარტიაზე კუთვნილება. თითოეული ცვლადისათვის კატეგორიები არათანმიმდევრულია და სკალას არა აქვს „მაღალი“, „დაბალი“, „საწყისი“, „ბოლო“. თვისებრივი ცვლადებისათვის კატეგორიები თვისებრივადაა განსხვავებული და არა რაოდენობრივად ანუ სიდიდით. ამასთან, სხვადასხვა კატეგორიას *სკალის დონე* ეწოდება და არცერთი დონე არ არის უფრო მაღალი, ან დაბალი სხვა დონეებთან მიმართებაში. ამ დონეებს განასხვავებს მხოლოდ სახელები.

როცა ცვლადის შესაძლო მნიშვნელობა სიდიდით განსხვავებულია, ცვლადს *რაოდენობრივი* ეწოდება. რაოდენობრივი ცვლადის თითოეული შესაძლო მნიშვნელობა ტოლია, უფრო დიდია, ან უფრო მცირეა ნებისმიერ სხვა შესაძლო მნიშვნელობაზე. რაოდენობრივი ცვლადების მაგალითებია ყოველთვიური შემოსავალი, სრული განათლების წლების რაოდენობა, დედამამიშვილების რაოდენობა, ტემპერატურა, მანძილი, დროის ხანგრძლივობა და ა.შ.

სკალას რომელზეც განლაგდება თვისებრივი ცვლადები, *სახელდების სკალა* ეწოდება.

ინტერვალების სკალა

სკალას, რომელზეც იზომება რაოდენობრივი ცვლადები – ინტერვალების სკალა ეწოდება. ინტერვალების სკალას აქვს სპეციფიკური, რაოდენობრივი დისტანცია, ანუ ინტერვალის ნებისმიერ ორ დონეს შორის. ყოველთვიური შემოსავალი, ჩვეულებრივ, იზომება ინტერვალების სკალაზე. ინტერვალის 125 ლარსა და 126 ლარს შორის უდრის 1 ლარს ან 100 თეთრს; ინტერვალის 175 სმ-სა და 178 სმ-ს შორის უდრის სამ სანტიმეტრს.

რიგის სკალა

მესამე სკალა სახელდების და ინტერვალების სკალას შორის მდებარეობს. მას აქვს ბუნებრივი რიგის მნიშვნელობა, თუმცა მნიშვნელობებს შორის არ არის განსაზღვრული მანძილი, ინტერვალის. მაგალითად: სოციალური კლასები (დაბალი, საშუალო, მაღალი), მთავრობის დანახარჯი გარემოზე (ძალიან ცოტა, საკმარისი, მაღალი, ძალიან მაღალი). მიუხედავად იმისა, რომ კატეგორიებს აქვს ამკარა რიგი, დისტანცია მათ შორის უცნობია.

სახელდების და რიგის სკალები იძლევა მონაცემების კატეგორიებში გაერთიანების საშუალებას. მაგრამ, განსხვავებით რიგის სკალისგან, სახელდების სკალა არ იძლევა საშუალებას განვასხვავოთ უფრო მაღალი ან დაბალი, ვიდრე სხვა მონაცემი.

რიგის სკალის მონაცემთა რაოდენობრივი ბუნება

როგორც ითქვა, სახელდების სკალის მონაცემები თვისებრივია, ინტერვალის სკალის კი – რაოდენობრივი. რიგის სკალის პოზიცია თვისებრივ-რაოდენობრივ კლასიფიკაციაზე გაურკვეველია. თუ სკალის შემცველი მონაცემები თვისებრივია,

გამოიყენება სახელდების სკალის ანალიზის მეთოდები. ზოგჯერ რიგის სკალის მონაცემებს ამუშავებენ, როგორც ინტერვალურს, რათა უზრუნველყონ მეთოდები რაოდენობრივი მონაცემების ხელმისაწვდომად. მაგალითად, აშშ-ში დისციპლინის (საგნის) ცოდნის დონის შეფასება (A, B, C, D, E კატეგორიებში) რიგის სკალაზე ხდება, მაგრამ ამ მონაცემებს ზოგჯერ ამუშავებენ, როგორც ინტერვალურს, როცა ამ კატეგორიებს აფასებენ რიცხვებით (4, 3, 2, 1, 0), რათა გამოითვალოს, მაგალითად, საშუალო ქულა.

სტატისტიკური მეთოდები და გაზომვის ტიპი

შეუძლებელია თვისებრივი მონაცემების გაანალიზება რაოდენობრივი ცვლადების მეშვეობით. თუ ცვლადი განეკუთვნება სახელდების სკალას, შეუძლებელია გამოიყენოთ ინტერვალების სკალის მეთოდები, რადგან სკალის დონეებს არა აქვთ რიცხობრივი მნიშვნელობა. ინტერვალების სკალაზე შესაძლებელია დავითვალოთ, მაგალითად, ცვლადის საშუალო (თვიური საშუალო შემოსავალი), მაგრამ საშუალოს გამოთვლა შეუძლებელია სახელდების სკალაზე. მაგალითად, რელიგიური კუთვნილების სკალაზე.

დისკრეტული და უწყვეტი ცვლადები

ცვლადი შეიძლება იყოს დისკრეტული (ოჯახში ბავშვების რაოდენობა), ან უწყვეტი (წონა, დროის ხანგრძლივობა, სიმაღლე, ასაკი). დისკრეტული ცვლადი იღებს მხოლოდ იზოლირებულ მნიშვნელობებს. მაგალითად, 0, 1, 2, 3, ... და, როგორც წესი, მიიღება თვლის შედეგად, ხოლო უწყვეტ ცვლადს შეუძლია მიიღოს ნებისმიერი რიცხვითი მნიშვნელობა რაიმე საზღვრებში. მაგალითად, დრო შეიძლება გავზომოთ საათების, წუთების, წამების და ა.შ. სიზუსტით.

ითვლება, რომ ყველა თვისებრივი ცვლადი დისკრეტულია (სახელდების და რიგის სკალაზე), ხოლო რაოდენობრივი ცვლადები შეიძლება იყოს დისკრეტულიც და უწყვეტიც.

განსხვავება უწყვეტ და დისკრეტულ ცვლადებს შორის პრაქტიკაში ხშირად ბუნდოვანია. მაგალითად, ვამბობთ, რომ პიროვნება 20 წლისაა (დისკრეტული), მაშინ, როცა მისი ასაკი მერყეობს 20-დან 21 წლამდე, ანუ შეიძლება გამოთვლილი იყოს თვეების, კვირების, საათების და ა.შ. სიზუსტით (უწყვეტი ცვლადი).

როცა ცვლადი სახელდების სკალაზეა, მაშინ ის თვისებრივია. ინტერვალების სკალაზე ცვლადი რაოდენობრივია, ხოლო რიგის სკალაზე ხან რაოდენობრივია, ხან თვისებრივი. ცვლადის მნიშვნელობები სახელდების და რიგის სკალაზე თავსდება კატეგორიებში და, შესაბამისად, კატეგორიულია. კატეგორიული ცვლადები დისკრეტულია, ხოლო ინტერვალების სკალაზე ცვლადები შეიძლება იყოს დისკრეტულიც და უწყვეტიც.

შერჩევის სახეობები

დასკვნითი სტატისტიკა იყენებს შერჩევაზე დაყრდნობით გამოთვლილ აღწერით სტატისტიკას პოპულაციაზე დასკვნის გასაკეთებლად.

დასკვნების ხარისხი დამოკიდებულია იმაზე, რამდენად არის შერჩევა წარმომადგენლობითი ანუ რეპრეზენტაციული მოცემული პოპულაციისათვის. მიღებული შერჩევის რეპრეზენტაციულობას ხელს უშლის შერჩევის შეცდომა.

შერჩევის შეცდომა განიხილებულია განსხვავებით შერჩევასა და პოპულაციას შორის, რადგან შერჩევა ხდება მხოლოდ მონაწილე კონკრეტული სუბიექტებიდან.

შერჩევის შეცდომა არის შეცდომა, რომელსაც ადგილი აქვს მაშინ, როცა სტატისტიკა ემყარება შერჩევის შეფასებას და წინასწარმეტყველებს პოპულაციის პარამეტრის მნიშვნელობას.

ასეთი შეცდომა ადგილი წარმოსადგენია. მაგალითად, ვთქვათ ყუთში 10 სხვადასხვა ფერის 100 ბირთვია, თითოეული 10-ის ოდენობით. 10-ელემენტიან შერჩევაში შეიძლება ყველა მწვანე აღმოჩნდეს და აქედან გაკეთებული დასკვნა, რომ ყუთში ბირთვების უმეტესობა მწვანეა, სწორი არ იქნება. შემდგომ ჩვენ ვნახავთ, როგორ უნდა შევაფასოთ შერჩევის მოსალოდნელი შეცდომა.

შერჩევის შეცდომის შესაფასებლად, თვით შერჩევა უნდა იყოს ჩატარებული გარკვეული წესების დაცვით. არსებობს შერჩევის მიღების ბევრი მეთოდი. განვიხილოთ რამდენიმე მათგანი.

მარტივი შემთხვევითი შერჩევა – შერჩევა, რომელშიც პოპულაციის ნებისმიერ სუბიექტს აქვს მოხვედრის ერთი და იგივე შანსი და ერთი წევრის არჩევა ხდება ნებისმიერი სხვა წევრებისაგან დამოუკიდებლად. სუბიექტები შეიძლება იყვნენ როგორც ინდივიდები, ასევე ოჯახები, სკოლები, ქალაქები და ა.შ.

აღვნიშნოთ n -ით შერჩევის სუბიექტთა რაოდენობა. n -ს ეწოდება შერჩევის მოცულობა. N -ით აღნიშნავენ პოპულაციის მოცულობას. პოპულაციიდან აღებული, ნებისმიერი n მოცულობის შერჩევას, უნდა ჰქონდეს ამორჩევის ერთი და იგივე შანსი. ამის მიღწევა ხშირად ხდება ე.წ. შემთხვევითი რიცხვების მეშვეობით.

მაგალითი 1.1. უნივერსიტეტის დამამთავრებელი კურსების $N = 500$ სტუდენტისგან საჭიროა $n = 20$ მოცულობის შერჩევის გამოყოფა, რათა მოეწიოს გამოკითხვა. გადავნიშნოთ ეს სტუდენტები (პირველი ნომერი – 001, ბოლო ნომერი 500). ვისარგებლოთ თანაბრად განაწილებული შემთხვევითი რიცხვების ცხრილით (იხილეთ დანართი). ამ ცხრილის ფრაგმენტი შეიცავს შემთხვევითი ციფრების 30 სტრიქონსა და 30 სვეტს.

ცხრილში შემთხვევით უნდა ავირჩიოთ ერთი რიცხვი. ვთქვათ, ჩვენ ავირჩიეთ ცხრიანი მე-15 სტრიქონისა და მე-2 სვეტის გადაკვეთაზე, გავყვეთ მარჯვნივ ამ სტრიქონს და ავირჩიოთ ციფრების 20 სამეული ისე, რომ გამოვტოვოთ 500-ზე მეტი რიცხვები. მივიღებთ:

137 468 130 250 384 397 374 287 245 384 421 444 212 492 372 321 137 485 104 153.

ყველა ეს რიცხვი 500-ზე ნაკლებია. ამრიგად, გამოსაყოფი შერჩევა იქნება შედგენილი პირებისაგან ამ ნომრებით.

ცხადია, შეგვეძლო სტრიქონის გასწვრივ გავყოლოდით მარცხნივ, სვეტის გასწვრივ ზემოთ, ან ქვემოთ, ოღონდ აუცილებლად წინასწარ განსაზღვრული გეგმით. შემთხვევითი რიცხვების არჩევა კომპიუტერული პროგრამების მეშვეობითაც შეიძლება.

ალბათური და არაალბათური შერჩევა

- **ალბათური (შემთხვევითი) შერჩევა**
პოპულაციის ყოველ ელემენტს აქვს შერჩევაში მოხვედრის განსაზღვრული (არანულოვანი) ალბათობა.
- **არაალბათური (მიზანმიმართული) შერჩევა**
შესწავლისათვის ვირჩევთ პოპულაციის გარკვეული ნიშან-თვისებების მქონე ელემენტებს.

არაალბათური შერჩევის სახეები

- ✓ მისანედომი შერჩევა;
- ✓ ქვოტური შერჩევა;
- ✓ „თოვლის გუნდის“ პრინციპი;
- ✓ მიზანმიმართული შერჩევა:
 - პილოტური გამოკვლევა;
 - ფოკუს-ჯგუფები.

როდის ვიყენებთ შემთხვევით და როდის მიზანმიმართულ შერჩევებს?

- ✓ შემთხვევით შერჩევით კვლევებს ვიყენებთ ყოველთვის, როდესაც გვინდა ვიცოდეთ გამოკვლევით მიღებული დასკვნების სიზუსტე;
- ✓ მიზანმიმართულ შერჩევით კვლევებს ვიყენებთ გამოკვლევის პირველ ეტაპზე, პრობლემების, შერჩევის ბაზისა და კითხვარის დასაზუსტებლად (პილოტური გამოკვლევა, ფოკუს-ჯგუფები);
ან
- ✓ როდესაც არ არის მნიშვნელოვანი გამოკვლევის რაოდენობრივი დასკვნების სიზუსტე (ზოგიერთ მარკეტინგულ კვლევაში).

შემთხვევითი შერჩევის სახეები

- ✓ შერჩევა დაბრუნების გარეშე – შერჩევაში მოხვედრილ ერთეულს მეორედ შერჩევაში მოხვედრის შანსი არა აქვს;
- ✓ შერჩევა დაბრუნებით – შერჩევაში მოხვედრილ ელემენტს შანსი აქვს განმეორებით მოხვდეს შერჩევაში.

როგორი უნდა იყოს კარგი შერჩევა?

- ✓ **შემთხვევითი** – ყოველ ელემენტს უნდა გააჩნდეს შერჩევაში მოხვედრის ნინასწარ განსაზღვრული არანულოვანი ალბათობა.
- ✓ **წარმომადგენლობითი** – შერჩევა განფენილი უნდა იყოს მთელ პოპულაციაზე.
- ✓ **მოცულობითი** – შერჩევის მოცულობა საშუალებას უნდა იძლეოდეს, რომ კვლევით მიღებული რიცხვები იყოს საიმედო.

ყველაზე ხშირად გამოყენებული შერჩევის სახეები:

- მარტივი შემთხვევითი (SRS)
- სისტემური
- ზომის პროპორციული ალბათობით (PPS)
- განშრეებული (სტრატეფიცირებული)
- კლასტერული

1. **მარტივი შემთხვევითი შერჩევა (SRS)** ალბათური შერჩევის ერთ-ერთი სახეობა. ამ შემთხვევაში ყველა შერჩევით ერთეულს აქვს შერჩევაში მოხვედრის ერთი და იგივე შანსი.

განხორციელების გზები:

- ✓ ლატარია.
 - ✓ შემთხვევითი რიცხვების ცხრილი.
 - ✓ ფსევდოშემთხვევითი რიცხვების გენერატორი.
- დადებითი მხარეები:** მარტივი განხორციელება, მარტივი გამოთვლები.
უარყოფითი მხარე: ცუდ შედეგებს იძლევა, როცა ერთობლიობა არაერთგვაროვანია.

2. სისტემური შერჩევა

ნრფივი სისტემური შერჩევა: როდესაც N (პოპულაციის მოცულობა) არის n -ის (შერჩევის მოცულობის) ჯერადი.

შერჩევის პერიოდი: $N/n = k$

შევაჩიოთ ნატურალური r რიცხვი 1-დან k -მდე შემთხვევით (შემთხვევითი სტარტი)

შერჩევა იქნება – ' r ', ' $r+k$ ', ' $r+2k$ ', ..., ' $r+(n-1)k$ '.

ნრიული სისტემური შერჩევა: როდესაც N არ არის n -ის ჯერადი.

' k ' – შემთხვევითი სტარტი 1-დან N -მდე.

შერჩევა არის – k , $2k$, $3k$, ..., $(n-1)k$.

3. **ზომის პროპორციული ალბათობით შერჩევა (PPS).** ამ შემთხვევაში ყოველი შერჩევის ერთეულის შერჩევაში მოხვედრის ალბათობა პროპორციულია ერთეულის ზომისა.

4. სტრატეფიცირებული შემთხვევითი შერჩევა

თუ პოპულაციის ნევრების გაერთიანება შესაძლებელია ჰომოგენურ ქვეჯგუფებად, სტრატებად, ირჩევენ სტრატეფიცირებული შემთხვევითი შერჩევის მეთოდს. **შრე (სტრატა)** – პოპულაციის, რაიმე თვალსაზრისით ერთგვაროვანი, შერჩევითი ერთეულების ერთობლიობა. გენერალური ერთობლიობის შრეებად დაყოფის პროცესს განშრევებას (სტრატეფიკაციას) უწოდებენ.

სტრატა გულისხმობს მსგავსი ობიექტების გაერთიანებას, მაგალითად, ისეთი მახასიათებლის მიხედვით, როგორიცაა: სქესი, ასაკი, განათლების დონე. ასეთი პოპულაციის სტრატებიდან შერჩევის წევრებს ირჩევენ შემთხვევითად იმ პროპორციით, რა პროპორციითაც გვხვდება პოპულაციაში.

5. კლასტერული შერჩევა

პოპულაციის წევრები ჯგუფდებიან კლასტერებში. როგორც წესი, წევრების დაჯგუფება არის ბუნებრივი.

კლასტერი	სიითი ერთეული	ელემენტარული ერთეული	გამოსავალი ინფორმაცია
რეგიონი	ქალაქი	მალაზია	ვაჭრობაში დასაქმებულთა საერთო რაოდენობა
ქალაქი	სკოლა	მოსწავლე	გამოცდაზე ჩაჭრილ მოსწავლეთა რაოდენობა
კვირა	დღე	დღე	შემოსულ და გასულ ტურისტთა რაოდენობა

ყველაზე ცნობილი არააღმართური შერჩევა არის *მოხალისეების შერჩევა*, სადაც შერჩევის სუბიექტები თავს შერჩევის მოხალისეებად მოიხსენიებენ და ხელმისაწვდომნი არიან (მკვლევრისათვის) შერჩევისათვის. არააღმართური შერჩევის მეთოდების ნაკლი ისაა, რომ ასეთმა შერჩევამ შეიძლება საკმარისად არ წარმოადგინოს პოპულაცია და მისგან შესაძლოა მივიღოთ არასწორი დასკვნები. შემდეგ თავებში ვნახავთ, რომ არააღმართურ შერჩევებს იყენებენ უფრო ხშირად მაშინ, როცა კვლევის მიზანია დავადგინოთ განსხვავდება თუ არა ორი ან მეტი ჯგუფი ერთმანეთისაგან რაიმე საკვლევი მახასიათებლით.

შემთხვევითი შერჩევითი კვლევით მიღებული დასკვნები ალბათური ხასიათისაა: გამოკვლევის შედეგად მიღებული რიცხვები (შერჩევითი მახასიათებლები) გვაძლევს პოპულაციის შესაბამისი მახასიათებლების მიახლოებით მნიშვნელობებს.

გამოკვლევის შედეგად მიღებულ რიცხვს აუცილებლად უნდა ახლდეს ორი პარამეტრი:

ცდომილება – მაქსიმუმ რამდენი პროცენტია განსხვავება მიღებულ რიცხვსა და პოპულაციის შესაბამის მახასიათებელს შორის;

საიმედოობა – დასკვნის საიმედოობის რამდენი პროცენტია გაკეთებული.

სოციოლოგიურ და მარკეტინგულ კვლევებში, როგორც წესი, ცდომილების საზღვრად იღებენ 5%-ს, ხოლო საიმედოობის დონედ 95%-ს.

შეჯამება

სტატისტიკა შეისწავლის სხვადასხვა ცვლადს – საკვლევი სუბიექტების მახასიათებლებს:

- თვისებრივი და რაოდენობრივი ცვლადების ანალიზისათვის გამოიყენება განსხვავებული მეთოდები;

- თვისებრივი ცვლადები იზომება სახელდების სკალებზე, რაოდენობრივი კი – ინტერვალების სკალებზე;
- ცვლადი შეიძლება იყოს დისკრეტული ან უწყვეტი;
- აღწერითი სტატისტიკა იძლევა მეთოდებს მონაცემთა აღსაწერად და შესაფერხებლად, ხოლო დასკვნითი სტატისტიკა იყენებს მიღებულ ინფორმაციას პოპულაციაზე დასკვნის გასაკეთებლად;
- საიმედო დასკვნების გასაკეთებლად აუცილებელია, რომ შერჩევა იყოს პოპულაციის რეპრეზენტაციული, ანუ კარგად წარმოადგენდეს პოპულაციას;
- კვლევის შედეგები დამოკიდებულია იმაზე, თუ სუბიექტების როგორი შერჩევა იქნება გამოყენებული;
- მარტივი შემთხვევითი შერჩევა და სტრატეგიული შემთხვევითი შერჩევა საშუალებას გვაძლევს ვაკონტროლოთ შერჩევის შეცდომა.

ძირითადი ტერმინები

ცვლადი, თვისებრივი ცვლადი, რაოდენობრივი ცვლადი, უწყვეტი ცვლადი, დისკრეტული ცვლადი, სახელდების სკალა, ინტერვალების სკალა, რიგის სკალა, შერჩევის შეცდომა, მარტივი შემთხვევითი შერჩევა, სტრატეგიული შერჩევა, ალბათური შერჩევა, არაალბათური შერჩევა.

კითხვები

1. ახსენით განსხვავება:
 - ა. თვისებრივ და რაოდენობრივ ცვლადებს შორის;
 - ბ. დისკრეტულ და უწყვეტ ცვლადებს შორის.

ახდენს, თუ არა ეს განსხვავება გავლენას სტატისტიკურ ანალიზზე?
2. განსაზღვრეთ, ქვემოთ ჩამოთვლილი ცვლადები დისკრეტულია, თუ უწყვეტი:
 - ა. მანძილი სახლიდან უახლოეს მარკეტამდე;
 - ბ. თქვენი ჯგუფელების რაოდენობა სტატისტიკის ლექციაზე;
 - გ. თქვენი ახლობლების წონები;
 - დ. სქესი.
3. საბავშვო ბაღში ბავშვების ურთიერთქმედება გააერთიანეს კატეგორიებში: ანტაგონისტური, ნეიტრალური, ალტრუისტული. გაზომვის რომელ სკალას მიეკუთვნება ეს ცვლადი?
4. დრო, რომელსაც უთმობენ აბიტურიენტები ყოველდღიურად საშინაო დავალების შესრულებას. გაზომვის რომელ სკალას მიეკუთვნება ეს ცვლადი?
5. თანამშრომელთა ჯგუფში სამსახურებრივი სტრესი გაზომეს 10-დონიან სკალაზე. გაზომვის რომელი სკალა გამოიყენეს?
6. გაზომვის რომელ სკალაზე გაზომავთ ქვემოთ ჩამოთვლილ ცვლადებს?
 - ა. სპეციალობა (მშენებელი, ექიმი, მასწავლებელი);
 - ბ. სოციალური სტატუსი (დაბალი, საშუალო, მაღალი);
 - გ. დანაშაულობათა ინდექსი რაიონების მიხედვით (1000 კაცზე დანაშაულობათა რაოდენობა);
 - დ. თვიური შემოსავალი;

- ე. საცხოვრებელი ადგილი (სოფელი, დაბა, პატარა ქალაქი, დედაქალაქი);
 ვ. ოჯახში წევრთა რაოდენობა (1, 2, 3, 4, ...).
7. მოიყვანეთ მაგალითები:
- ა. თვისებრივი ცვლადის;
 - ბ. რაოდენობრივი ცვლადის;
 - გ. სახელდების სკალაზე გაზომილი ცვლადის;
 - დ. რიგის სკალაზე გაზომილი ცვლადის;
 - ე. ინტერვალების სკალაზე გაზომილი ცვლადის;
 - ვ. დისკრეტული ცვლადის;
 - ზ. უწყვეტი ცვლადის.
8. ქვემოთ მოყვანილი ცვლადებიდან რომელი შეიძლება ჩაითვალოს უწყვეტად?
- ა. ასაკი;
 - ბ. მისჯილი პატიმრობის ვადა;
 - გ. მანძილი სახლიდან ავტობუსის გაჩერებამდე;
 - დ. სოფლის მოსახლეობის რაოდენობა.
9. ქვემოთმოყვანილი ცვლადებიდან რომელია უწყვეტი?
- ა. მკურნალობის მეთოდი;
 - ბ. ამოცანის ამოხსნის დრო;
 - გ. გასხვისება;
 - დ. ავტორიტარიზმი;
 - ე. შფოთვა;
 - ვ. საცხოვრებელი ადგილი.
10. სტატისტიკის კურსს გადის 800 სტუდენტი. თქვენ დაგავალეს 40-კაციანი შერჩევის შექმნა. ხელთ გაქვთ 800 სტუდენტის პირადი ნომერი. როგორ შეასრულებთ ამ დავალებას?

თავი 2

აღწერიითი სტატისტიკა. ცხრილები და გრაფიკები

აღწერიითი სტატისტიკის მეთოდების ერთ-ერთი დანიშნულებაა მონაცემთა აღწერა, მათი წარმოდგენა და შეჯამება. განვიხილოთ რაოდენობრივ მონაცემთა აღწერის მეთოდები.

რაოდენობრივი მონაცემების სიხშირეთა განაწილება და მისი გრაფიკული წარმოდგენა

მაგალითი 2.1. 2.1 ცხრილში წარმოდგენილია მკვლელობის დონე, გათვლილი 100 000 კაცზე (აშშ, 50 შტატის მონაცემები, 1993 წელი). როგორ გამოითვლება მკვლელობის დონე?

ეთქვათ, რომელიმე შტატში 1993 წელს მოხდა 120 მკვლელობა და ამ შტატის მოსახლეობა შეადგენს 2 300 000 ადამიანს. მკვლელობის დონის გამოთვლა შეიძლება 100 000 კაცზე, 10 000 კაცზე, 1 000 კაცზე და ა.შ. მკვლელობის დონე, გამოთვლილი 100 000 კაცზე, უდრის:

$$\frac{120}{2300000} \cdot 100000 \approx 5.2.$$

მკვლელობის დონის ამგვარი გამოთვლა საშუალებას გვაძლევს შევადაროთ მკვლელობის დონე არა მხოლოდ ქვეყნის შიგნით, არამედ ნებისმიერ ქვეყნებს (ან შტატებს) შორის.

ცხრილი 2.1. მკვლელობის დონე 100 000 კაცზე (აშშ, 50 შტატი, 1993 წ.)

შტატი	მკვლელობის დონე	შტატი	მკვლელობის დონე
აიდაჰო	3.5	მერილენდი	12.7
ალაბამა	11.6	მიჩიგანი	9.8
ალასკა	9.0	მონტანა	3.0
არიზონა	8.6	ნებრასკა	3.9
არკანზასი	10.2	ნევადა	10.4
დასავლეთ ვირჯინია	6.9	ნიუ იორკი	13.3
დელავერი	5.0	ნიუმექსიკო	8.0
ვაიომინგი	3.4	ნიუ ჯერსი	5.3
ვაშინგტონი	5.2	ნიუჰემპშირი	2.0
ვერმონტი	3.6	ოკლაჰომა	8.4
ვირჯინია	8.3	ორეგონი	4.6

შტატი	მკვლელობის დონე	შტატი	მკვლელობის დონე
ილინოისი	11.4	ოჰაიო	6.0
ინდიანა	7.5	პენსილვანია	6.8
იუტა	2.3	როდ აილენდი	3.9
კალიფორნია	13.1	სამხრეთი დაკოტა	3.4
კანზასი	6.4	სამხრეთი კაროლინა	10.3
ენტუკი	6.3	ტენესი	10.2
კოლორადო	5.8	ტეხასი	11.9
კონექტიკუტი	6.6	უისკონსინი	4.4
ლუიზიანა	20.3	ფლორიდა	8.9
მასაჩუსეტსი	3.9	ჩრდილოეთ დაკოტა	1.7
მაიამი	1.6	ჩრდილოეთ კაროლინა	11.3
მინესოტა	3.4	ჯორჯია	11.4
მისისიპი	13.5	ჰავაი	3.8
მისური	11.3	იუტა	3.1

სიხშირეთა განაწილება

2.1 ცხრილში განხილული მონაცემების ანალიზის და მათგან ინფორმაციის მიღების პირველი ნაბიჯია მონაცემების წარმოდგენა სიხშირეთა განაწილების სახით. ამისათვის ხდება მონაცემების დაჯგუფება ინტერვალებში. ინტერვალები ლაგდება მინიმალურიდან მაქსიმალურამდე. შემდეგ ითვლება თითოეულ ინტერვალში მოხვედრილი მონაცემების რაოდენობა (სიხშირეები) და ხდება მათი გამოსახვა ცხრილის სახით.

ინტერვალების რაოდენობა დამოკიდებულია მკვლევრის გადანყვეტილებაზე, მონაცემთა დიაპაზონსა და დაკვირვებების რაოდენობაზე. გასათვალისწინებელია ის ფაქტი, რომ მონაცემების დამუშავება რთულდება, თუ ინტერვალები ძალიან მცირეა. დიდი ინტერვალების შემთხვევაში კი შეიძლება დაიკარგოს მნიშვნელოვანი ინფორმაცია.

მონაცემთა დიაპაზონის მისაღებად 2.1 ცხრილში ვიპოვოთ მკვლელობის დონის მაქსიმალური და მინიმალური მნიშვნელობა. მაქსიმალური დონე დაფიქსირდა ლუიზიანაში – 20.3, მინიმალური კი მაიამიში – 1.6.

დიაპაზონი აღინიშნება Δ ასოთი და გამოითვლება ფორმულით:

$$\Delta = x_{\max} - x_{\min}.$$

სტატისტიკაში სიმბოლო X ხშირად გამოიყენება თავად მონაცემის აღსანიშნად. მკვლელობის დონის დიაპაზონი უდრის: $\Delta = 20.3 - 1.6 = 18.7$.

თუ ჩვენ ვისურვებთ, რომ ინტერვალების რაოდენობა იყოს 7 და ინტერვალის სიგრძეს l ასოთი აღვნიშნავთ, მაშინ:

$$l = 18.7 / 7 = 2.67.$$

სასურველია, რომ ინტერვალის სიგრძე მთელი რიცხვით გამოისახებოდეს. თუ მიღებულ ათწილადს დავამრგვალებთ მთელამდე, ინტერვალის სიგრძე 3-ის ტოლი იქნება. დიაპაზონი დაიყოს შვიდ ინტერვალად, თითოეული 3-ის ტოლი სიდიდით. ისეთი კომპიუტერული პროგრამები, როგორიც არის SAS ან SPSS, თვითონ ირჩევენ ინტერვალების რაოდენობას.

მკვლელობის დონის ინტერვალები იქნება (0.0–2.9, 3.0–5.9, 6.0–8.9, 9.0–11.9, 12.0–14.9, 15.0–17.9, 18.0–20.9).

ინტერვალების შექმნის შემდეგ უნდა დავთვალოთ შტატების რაოდენობა, რომლის დონე „ჩავარდება“ თითოეულ ინტერვალში. მიღებულ სიხშირეთა განაწილება მოცემულია 2.2 ცხრილში.

ცხრილი 2.2. მკვლელობის დონის სიხშირეთა განაწილება

მკვლელობის დონე	სიხშირე (შტატების რაოდენობა)	ფარდობითი სიხშირე	ფარდობითი სიხშირე %-ში
0.0–2.9	5	0.1	10
3.0–5.9	16	0.32	32
6.0–8.9	12	0.24	24
9.0–11.9	12	0.24	24
12.0–14.9	4	0.08	8.0
15.0–17.9	0	0.0	0.0
18.0–20.9	1	0.02	2.0
სულ	50	1	100

რა ინფორმაცია მივიღეთ? ამ ცხრილიდან ჩანს, რომ ხუთ შტატში მკვლელობის დონე სამზე ნაკლებია, 16 შტატში მკვლელობის დონე მერყეობს 3-დან 5.9-მდე. მხოლოდ ერთ შტატშია მკვლელობის დონე ყველაზე მაღალი, რადგან ხვდება ინტერვალში (18.0 – 20.9). 24 შტატის დონე მერყეობს 6.0-დან 11.9-მდე. პირველი ინტერვალის სიხშირე $f = 5$, მეორე ინტერვალის სიხშირეა $f = 16$ და ა.შ. (სიხშირეს ხშირად აღნიშნავენ f ასოთი – პირველი ასო სიტყვაში frequency).

ფარდობითი სიხშირის გამოსათვლელად სიხშირე უნდა გავყოთ მონაცემთა სრულ რაოდენობაზე, ჩვენს შემთხვევაში 50-ზე (50 შტატი). მაგალითად, პირველი ინტერვალის ფარდობითი სიხშირის მისაღებად 5 უნდა გავყოთ 50-ზე.

$$rf = 5 / 50 = 0.1$$

(rf – relative frequency). ყველა ფარდობითი სიხშირის ჯამი უდრის 1-ს.

პროცენტებში გამოხატული ფარდობითი სიხშირის მისაღებად ფარდობითი სიხშირე უნდა გავამრავლოთ 100-ზე:

$$rf\% = \frac{5}{50} \cdot 100\% = 10\%.$$

პროცენტებში გამოთვლილ ფარდობით სიხშირეთა ჯამი უდრის 100%-ს.

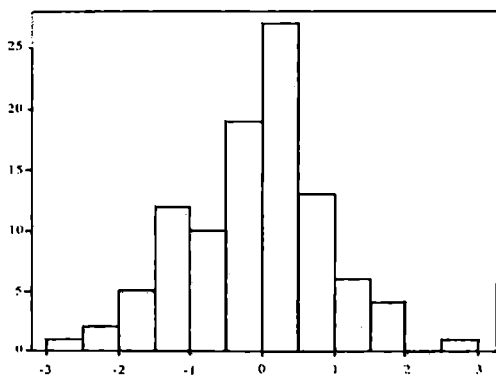
სიხშირეთა განაწილების გრაფიკულად გამოსახვა

ჰისტოგრამები

მონაცემების უკეთ აღსაქმელად, სიხშირეთა განაწილებით მოცემული ინფორმაცია ხშირად გამოისახება გრაფიკული სახით.

რაოდენობრივი მონაცემების სიხშირეთა განაწილების გრაფიკული გამოსახვის ერთ-ერთი ფორმაა – *ჰისტოგრამა*.

შესაძლო მონაცემების ინტერვალები გადაზომილია ჰორიზონტალურ ox ანუ აბსცისთა ღერძზე. სიხშირის საზომი, მაგალითად, სიხშირე (ან ფარდობითი სიხშირე) შეფარდებული ინტერვალის სიგრძესთან, გადაზომილია oy ღერძზე ანუ ორდინატთა ღერძზე. გრაფიკის სიმაღლე h დამოკიდებულია მონაცემთა სიხშირეზე კონკრეტული ინტერვალისთვის: $h_i = f_i / \Delta$, თითოეული მართკუთხედის ფართობი $h_i \cdot \Delta = f_i$ -ის ტოლია, ამიტომ ყველა ფართობის ჯამი უდრის შერჩევის მოცულობას (ან, შესაბამისად, $h_i = f_i / (n\Delta)$, თითოეული მართკუთხედის ფართობი $h_i \cdot \Delta = f_i / n$ -ის ტოლია, ამიტომ ყველა ფართობის ჯამი უდრის ერთს). სტატისტიკური პროგრამები თვითონ ირჩევენ ჰისტოგრამის ოპტიმალურ ზომას.



ნახ. 2.1. მკვლელობათა დონის სიხშირეთა განაწილების ჰისტოგრამა

მაგალითი 2.2. ქვემოთ მოყვანილ ცხრილში მოცემულია დიდი ზომის ავტომანქანების სანავაგის დანახარჯის დასადგენად ჩატარებული 100 დაკვირვების მონაცემები. ცხრილი გვიჩვენებს 1 გალონი (≈ 3.78 ლიტრი) სანავაგით ავტომანქანის გავლილ მანძილს მილებში (1 მილი = 1.609 კმ):

მილი/გალონზე

19.0	20.8	22.0	22.7	20.0	18.9	16.6	16.8	20.8	14.7
15.1	21.8	21.1	21.5	21.1	15.5	19.3	15.1	20.6	16.8
18.2	20.5	15.3	16.2	16.3	22.8	22.7	21.9	22.5	17.1
19.1	21.6	19.0	18.3	18.6	22.1	17.5	22.9	21.7	18.7
21.9	20.2	14.5	14.1	22.9	20.2	17.3	22.6	19.3	21.7
21.5	22.6	18.7	19.2	22.8	21.6	21.7	20.5	22.7	20.4
18.8	15.1	16.5	20.5	19.1	17.4	19.7	19.2	16.4	21.9
14.3	19.2	19.7	17.1	21.4	21.9	21.7	19.2	23.9	19.6
20.9	18.5	20.2	18.2	20.2	22.4	20.4	21.6	21.3	22.4
20.5	18.1	20.7	21.3	16.9	20.3	23.9	18.8	21.1	21.9

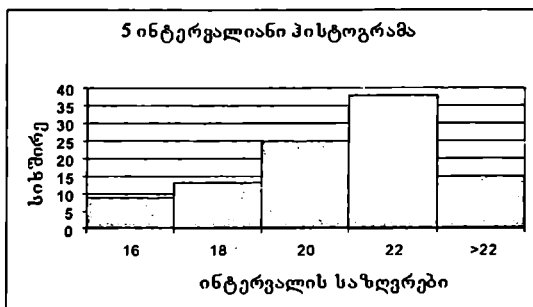
ამ მონაცემების დამუშავება მდგომარეობს შემდეგში:

I. დაკვირვებულ მონაცემებს შორის უნდა მოიძებნოს მინიმალური და მაქსიმალური მნიშვნელობები. წარმოდგენილ მონაცემებში უმცირესი რიცხვითი მნიშვნელობაა $x_{\min}=14.1$, ხოლო უდიდესი $-x_{\max}=23.9$. ეს ორი რიცხვი განსაზღვრავს მონაცემთა გაფანტულობის უმარტივეს მახასიათებელს – გაბნევის დიაპაზონს, რომელიც არის სხვაობა მონაცემთა მაქსიმალურ და მინიმალურ რიცხვით მნიშვნელობებს შორის. შესაბამისად, მონაცემთა გაბნევის დიაპაზონი ტოლია: $23.9-14.1=9.8$.

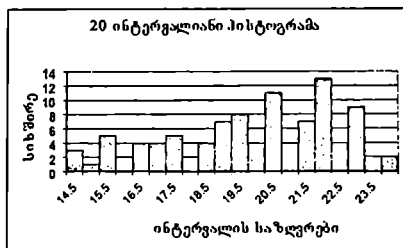
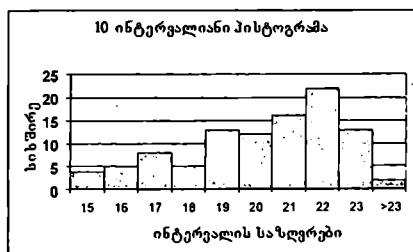
II. შემდეგი ნაბიჯია – მონაცემთა დაჯგუფების ინტერვალების გამოყოფა. ამისათვის უმარტივეს შემთხვევაში გაბნევის დიაპაზონს ყოფენ ერთი და იმავე სიგრძის დაახლოებით 5 – 20 ქვეინტერვალად (ხშირად გამოიყენება ქვეინტერვალთა რაოდენობის სტრადფესის წესი: $[\log_2 n] + 1$ ან ბრუკსისა და კარუზერის წესი: $5 \log_{10} n$). მოცემულ შემთხვევაში ავირჩიოთ 5 ტოლი სიგრძის ინტერვალი. სიმარტივისათვის მინიმალური და მაქსიმალური მნიშვნელობები დავამრგვალოთ უახლოეს მთელ რიცხვებამდე: კერძოდ, მინიმუმად ავიღოთ 14, ხოლო მაქსიმუმად – 24. მაშინ თითოეული ქვეინტერვალის სიგრძედ უნდა ავიღოთ: $\Delta=(24-14)/5=2$. შესაბამისად, დაჯგუფების ინტერვალები იქნება: $(-\infty, 16]$, $(16, 18]$, $(18, 20]$, $(20, 22]$, $(22, +\infty)$. გამოვთვალოთ ამ ქვეინტერვალებში მოხვედრილი მონაცემების სიხშირეები და ჩავწეროთ ისინი ცხრილის სახით:

ინტერვალი	სიხშირეთა გამოთვლა	სიხშირე
$(-\infty, 16]$	//// //	9
$(16, 18]$	//// //// //	13
$(18, 20]$	//// //// //// //// //	25
$(20, 22]$	//// //// //// //// //// //// //	38
$(22, +\infty)$	//// //// //// //	15
	სულ	100

ამ ცხრილის მიხედვით აგებულ ჰისტოგრამას ექნება შემდეგი სახე (აქ, თვალსაჩინოების მიზნით, სიხშირე არაა გაყოფილი ინტერვალის სიგრძეზე):



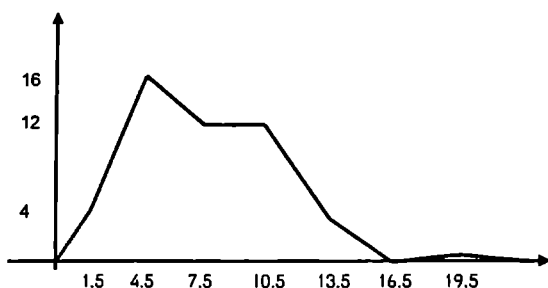
თუ იმავე მონაცემებს დავაჯგუფებთ 10 ან 20 ქვეინტერვალში, მაშინ ჰისტოგრამებს ექნებათ, შესაბამისად, შემდეგი სახეები:



როგორც ვხედავთ, მონაცემთა დაჯგუფების ქვეინტერვალის რაოდენობის ზრდასთან ერთად, ჰისტოგრამა სულ უფრო და უფრო დაკბილული ხდება. ამასთანავე, ამ შემთხვევაში დაჯგუფების 5 და 10 ინტერვალის იძლევა უფრო მკაფიოდ გამოხატულ კანონზომიერებას, მაშინ, როდესაც 20 ინტერვალის შემთხვევაში გაჩნდა ბევრი, მოვლენის შინაარსიდან გამომდინარე, აზრს მოკლებული "რხევა", რაც გამოწვეულია მონაცემთა მოცემული რაოდენობისათვის დაჯგუფების ინტერვალთა დიდი რაოდენობით.

სიხშირეთა პოლიგონი

სიხშირეთა პოლიგონის ასაგებად საჭიროა, აბსცისთა ღერძზე განლაგდეს ინტერვალის შუანერტილები, ინტერვალის სიხშირე კი აღინიშნება შუანერტილების თავზე წერტილებით. ეს წერტილები შემდეგ ერთდება სწორი ხაზებით. იმისათვის, რომ სიხშირეთა განაწილების პოლიგონი განაწილების ორივე ბოლოში ნულამდე იყოს მიყვანილი, ამატებენ თითო-თითო ინტერვალს განაწილების ბოლოებში, რომლებიც არ შეიცავენ არცერთ მონაცემს.



ნახ. 2.2. მკვლელობათა დონის სიხშირეთა განაწილების პოლიგონი

ჰისტოგრამისა და სიხშირეთა პოლიგონის ასაგებად ხშირად იყენებენ ფარდობით სიხშირეებს. ფარდობითი სიხშირეების გამოყენების უპირატესობა ისაა, რომ ფარდობითი სიხშირეები გვიჩვენებენ სხვადასხვა ინტერვალში მოხვედრილ მონაცემთა წილს და ამიტომ ორი განსხვავებული განაწილებისათვის მონაცემთა სხვადასხვა რაოდენობის შედარებაა შესაძლებელი.

თვისებრივი მონაცემების სიხშირეთა განაწილება.

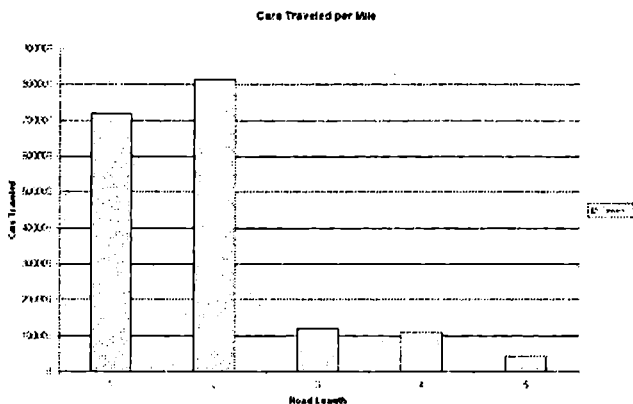
სვეტოვანი და წრიული დიაგრამები

თვისებრივი მონაცემებისთვის, მაგალითად, ზრდასრული მოსახლეობის ოჯახური მდგომარეობის განაწილება კატეგორიებში: მარტოხელა, დაქორწინებული, დაშორებული, ქერივი, აგებენ სიხშირეთა განაწილებას კატეგორიების მიხედვით, ანუ ითვლება თითოეულ კატეგორიაში მოხვედრილი მონაცემების რაოდენობა. ამის შემდეგ აგებენ სვეტოვან დიაგრამებს. სვეტების რაოდენობა დამოკიდებულია თვისებრივი მონაცემების კატეგორიების რაოდენობაზე, სვეტების სიმაღლე კი – შესაბამის კატეგორიაში მოხვედრილ მონაცემთა რაოდენობაზე (სიხშირეზე). როგორც ნახ. 2.3-დან ჩანს, სვეტები ერთმანეთისგან დაშორებულია იმის საჩვენებლად, რომ მონაცემები თვისებრივია. 2.3 ცხრილი და 2.3 სურათი თვისებრივი მონაცემების სიხშირეთა განაწილების მაგალითია.

ცხრილი 2.3. ოჯახების სტრუქტურა, 1994 წელი, აშშ

#	ოჯახის ტიპი	რაოდენობა (მილიონი)	პროცენტი
1	ბავშვებით	25.1	36.6
2	უბავშვო	28.1	41.0
3	მარტოხელა დედა	7.6	11.1
4	მარტოხელა მამა	1.3	1.9
5	სხვა	6.4	9.3
	სულ	68.5	99.9

ნახ. 2.3-ზე გამოსახულია თვისებრივი მონაცემების სიხშირეთა განაწილების გრაფიკი – სვეტოვანი დიაგრამა.

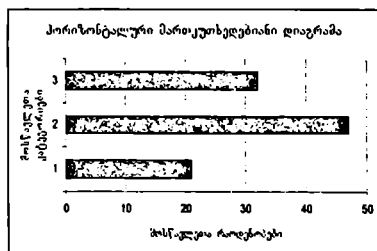
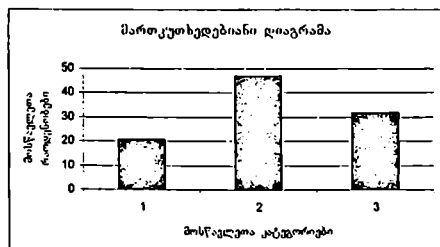


ნახ. 2.3 ოჯახების განაწილება ოჯახის ტიპის მიხედვით

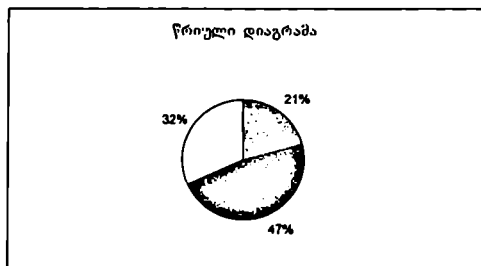
თვისებრივ მონაცემთა განაწილების გრაფიკული წარმოდგენისათვის ორი ძირითადი საშუალება არსებობს: მართკუთხედებიანი (ანუ სვეტოვანი) დიაგრამები და წრიული (ანუ სექტორებიანი) დიაგრამები.

მაგალითი 2.3. ქვემოთ მოყვანილია სკოლის დამამთავრებელი კლასების მოსწავლეთა აკადემიური მოსწრება ნიშნების მიხედვით. I კატეგორია: „9-10-ბალიანი“ – 21 მოსწავლე, II კატეგორია: „7-8-ბალიანი“ – 47 მოსწავლე, III კატეგორია: „5-6-ბალიანი“ – 32 მოსწავლე.

ავილთ აბსცისთა ღერძზე მდებარე ტოლი ფუძეების მქონე სამი მართკუთხედი, რომელთა სიმაღლეები პროპორციულია მოსწავლეთა რაოდენობების და ისინი ერთმანეთისაგან ტოლი დაშორებით განვალაგოთ. ორდინატთა ღერძზე გადაიზომება სიხშირეები. ასე მიღებულ ნახატს *მართკუთხედებიანი დიაგრამა* ეწოდება. ცხადია, რომ ეერტიკალურ ღერძზე შეიძლება გადაგვეზომა ფარდობითი სიხშირე.



წრიული დიაგრამა განკუთვნილია დროის მოცემულ მომენტში ან მოცემულ შუალედში ობიექტთა მოცემული რაოდენობის კომპონენტებად დაყოფის ანუ ამ ობიექტთა კლასიფიკაციის აღსანიშნავად. ის უფრო მოხერხებულია პროცენტებში გამოსახული ნაწილების ან ფარდობითი სიხშირეების გამოსახატად. ჩვენს მაგალითში ვაგებთ სექტორებს, რომლებიც შეადგენენ წრიული კუთხის – 360° -ის შესაბამისად 21/100-ს ანუ 21%-ს, 47/100-ს ანუ 47%-ს, 32/100-ს ანუ 32%-ს.



სტატისტიკაში ინტენსიურად გამოიყენება წრიული დიაგრამა. წრიული დიაგრამის გამოყენების მიზანია ვაჩვენოთ მთელის ნაწილებს შორის ურთიერთმიმართება სექტორების ზომების ვიზუალური შედარების გზით; გამოიყენება როგორც პროცენტები, ისე პროპორციები. სიდიდეები შეიძლება იყოს როგორც რიცხობრივი, ისე კატეგორიული ბუნების.

წრიული დიაგრამა არის სექტორებად დაყოფილი წრე, სადაც თითოეული სექტორი შეესაბამება ნებისმიერი კატეგორიის განაწილებაში მონაწილე სიხშირეების პროცენტულ გამოსახულებებს ან ფარდობით სიხშირეებს.

მაგალითი 2.4. ქვემოთ მოყვანილია 1998 წლის განმავლობაში მსუბუქ საუზმეზე დახარჯული ფუნტების რაოდენობის განაწილება საკვების სახეობების მიხედვით. ავაგოთ წრიული დიაგრამა.

მსუბუქი საუზმე	ფუნტები (სიხშირე)
კარტოფილის „ჩიფსი“	11.2 მილიონი
სიმინდის „ჩიფსი“	8.2 მილიონი
ფუნტუშა	4.3 მილიონი
სიმინდის ბურბუშელა	3.8 მილიონი
კექსი	2.5 მილიონი
ჯამი $n = 30.0$ მილიონი	

ამოხსნა:

ნაბიჯი 1. ვინაიდან წრე შედგება 360 გრადუსისაგან, თითოეული კლასის სიხშირე უნდა გადავიყვანოთ წრის პროპორციულ ნაწილში. ეს გადაყვანა უნდა მოხდეს შემდეგი ფორმულის გამოყენებით:

$$\text{გრადუსი} = \frac{f_i}{n} \cdot 360^\circ,$$

სადაც f_i – არის i -ური კლასის შესაბამისი სიხშირე, ხოლო n – სიხშირეთა ჯამი. ამიტომ გვექნება შემდეგი თანაფარდობები (ცხადია, რომ ყველა გრადუსის ჯამიუნდა იყოს 360°):

$$\text{კარტოფილის „ჩიფსი“} - \frac{11.2}{30} \cdot 360^\circ = 134^\circ;$$

$$\text{სიმინდის „ჩიფსი“} - \frac{8.2}{30} \cdot 360^\circ = 98^\circ;$$

$$\text{ფუნთუშა} - \frac{4.3}{30} \cdot 360^\circ = 52^\circ;$$

$$\text{სიმინდის ბურბუშელა} - \frac{3.8}{30} \cdot 360^\circ = 46^\circ;$$

$$\text{კექსი} - \frac{2.5}{30} \cdot 360^\circ = 30^\circ;$$

$$\text{ჯამი} - 360^\circ.$$

ნაბიჯი 2. თითოეული სიხშირე უნდა გადავიყვანოთ აგრეთვე პროცენტებში, რისთვისაც უნდა ვისარგებლოთ ფორმულით:

$$\text{პროცენტი} - \% = \frac{f_i}{n} \cdot 100\%.$$

შესაბამისად, ჩვენ მივიღებთ შემდეგ პროცენტებს (ცხადია მათი ჯამი უნდა იყოს 100%):

$$\text{კარტოფილის „ჩიფსი“} - \frac{11.2}{30} \cdot 100\% = 37.3\%;$$

$$\text{სიმინდის „ჩიფსი“} - \frac{8.2}{30} \cdot 100\% = 27.3\%;$$

$$\text{ფუნთუშა} - \frac{4.3}{30} \cdot 100\% = 14.3\%;$$

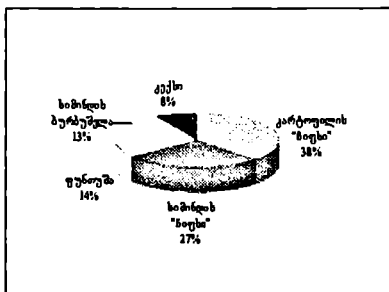
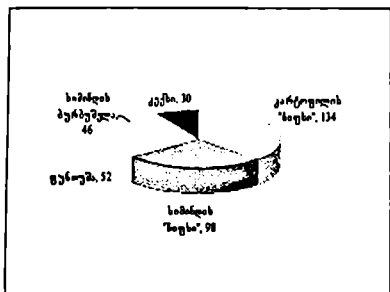
$$\text{სიმინდის ბურბუშელა} - \frac{3.8}{30} \cdot 100\% = 12.6\%;$$

$$\text{კექსი} - \frac{2.5}{30} \cdot 100\% = 8.3\%;$$

$$\text{ჯამი} - 100\%.$$

შენიშვნა: დამრგვალების გამო, გრადუსების ჯამი შეიძლება ყოველთვის არ იყოს 360° , ისევე როგორც პროცენტების ჯამი – 100%.

ნაბიჯი 3. ვისარგებლოთ ტრანსპორტირით, დავხაზოთ წრიული დიაგრამა და მასზე მივუთითოთ თითოეული სექტორის დასახელება და შესაბამისი ერთეული, როგორც ეს ნახაზზეა ნაჩვენები.



მაგალითი 2.5. ავგოთ წრიული დიაგრამა არმიის ახალწვეულთა სისხლის ჯგუფების სიხშირეთა განაწილების ქვემოთ მოყვანილი ცხრილის მიხედვით:

კლასი	სიხშირე
A	5
BB	7
O	9
AB	4
	ჯამი 25

ამოხსნა.

ნაბიჯი 1. ვიპოვოთ თითოეული კლასის შესაბამისი გრადუსული ზომა, შემდეგი ფორმულის გამოყენებით:

$$\text{გრადუსი} = \frac{f_i}{n} \cdot 360^\circ.$$

ადვილად დავრწმუნდებით, რომ მივიღებთ გრადუსების შემდეგ რიცხვებს:

$$A - 72^\circ, \quad BB - 100.8^\circ, \quad O - 129.6^\circ, \quad AB - 57.6^\circ.$$

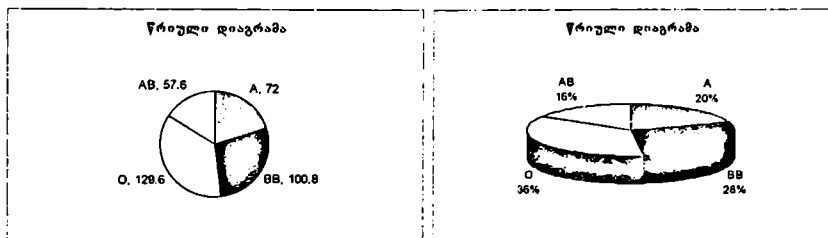
ნაბიჯი 2. ვიპოვოთ თითოეული კლასის შესაბამისი პროცენტი, შემდეგი ფორმულის მიხედვით:

$$\text{პროცენტი} - \% = \frac{f_i}{n} \cdot 100\%.$$

მივიღებთ, შესაბამისად, პროცენტების შემდეგ რიცხვებს:

$$A - 20\%, \quad BB - 28\%, \quad O - 36\%, \quad AB - 16\%.$$

ნაბიჯი 3. ვისარგებლოთ ტრანსპორტირით, აეგოთ წრიული დიაგრამა და მასზე მოენიშნოთ თითოეული სექტორის სახელი. გვექნება:



ამ დიაგრამიდან ჩანს, რომ ყველაზე მეტად გავრცელებული სისხლის ჯგუფია – O (36%). იმისათვის, რომ გავანალიზოთ წრიულ დიაგრამაზე წარმოდგენილი მონაცემები, უნდა შევადაროთ სექტორები. მაგალითად, არის თუ არა რომელიმე სექტორი გაცილებით დიდი, ყველა დანარჩენთან შედარებით? ეს დიაგრამა გვიჩვენებს, რომ სისხლის ტიპებს შორის ყველაზე მეტად გავრცელებულია O ტიპის სისხლი. ადამიანები, რომელთაც აქვთ AB ტიპის სისხლი, არიან უმცირესობაში. ადამიანების ორჯერ უფრო მეტ რაოდენობას აქვს O ტიპის სისხლი, AB ტიპის სისხლთან შედარებით.

ფოთლებიანი ღეროების მსგავსი დიაგრამები

ნახაზზე 2.4 ნაჩვენებია მკვლელობათა დონის ალტერნატიული დიაგრამა, რომელსაც ფოთლებიანი ღეროების მსგავს დიაგრამას უწოდებენ. ამ დიაგრამით ხდება მონაცემების რანჟირება ზრდის მიხედვით. ჩვეულებრივ, ჰისტოგრამისგან განსხვავებით, ის ასახავს იმ ინფორმაციას, რომელიც ინტერვალურებში იქარგება.

ღერო	ფოთლო
1	67
2	039
3	0144468999
4	46
5	0238
6	034689
7	5
8	03469
9	08
10	2234
11	334469
12	7
13	135
14	
15	
16	
17	
18	
19	
20	3

ნახ. 2.4. მკვლელობათა დონის ფოთლებიანი დიაგრამა

ამ დიაგრამაზე მკვლელობათა დონის ღერო არის რიცხვის მთელი ნაწილი, ფოთოლი კი წილადი ნაწილია. მაგალითად, მაიამის მკვლელობის დონე უდრის 1.6-ს (ცხრილი 2.1). ამ რიცხვის ღერო არის 1, ფოთოლი 6. ჩრდილოეთ დაკოტას დონე უდრის 1.7-ს. აქ ღეროა – 1, ფოთოლი – 7. ყველაზე დიდი დონე აქვს ლუიზიანას – 20.3. ამ რიცხვის ღეროა – 20, ფოთოლი კი – 3.

თუ ასაგებია ფოთლებიანი ღეროების მსგავსი დიაგრამა ისეთი რიცხვები-სათვის, როგორიცაა: 5, 12, 7, 15, 24, ღერო იქნება რიცხვის ათეული, ფოთოლი კი – ერთეული. რიცხვი 5-ის ღეროა – 0, ფოთოლი – 5; რიცხვი 24-ის ღეროა – 2, ფოთოლი – 4.

მაგალითი 2.6. აშშ-ის და კანადის მკვლელობათა დონეების შედარება. ხშირად ფოთლებიანი ღეროების მსგავს დიაგრამას იყენებენ ორი ჯგუფის (ორი ქვეყნის) მონაცემთა შესადარებლად. ამ დროს ფოთლები ლაგდება ერთი ღეროს სხვადასხვა მხარეს.

კანადა და აშშ

ფოთოლი (კანადა)	ღერო	ფოთოლი (აშშ)
7	0	
3 2 1	1	6 7
9 7 6 3 2 0	2	0 3 9
	3	0 1 4 4 4 6 8 9 9 9
	4	4 6
	5	0 2 3 8
	6	0 3 4 6 8 9
	7	5
	8	0 3 4 6 9
	9	0 8
	10	2 2 3 4
	11	3 3 4 4 6 9
	12	7
	13	1 3 5
	14	
	15	
	16	
	17	
	18	
	19	
	20	3

ნახ. 2.5. მკვლელობათა დონეების ფოთლებიანი ღეროების მსგავსი დიაგრამა.

2.5 ნახაზიდან ჩანს, რომ კანადაში მკვლელობის ყველა დონე ნაკლებია 3-ზე. მაქსიმალური დონე 2.9-ის ტოლია, მინიმალური 0.7-ის ტოლია.

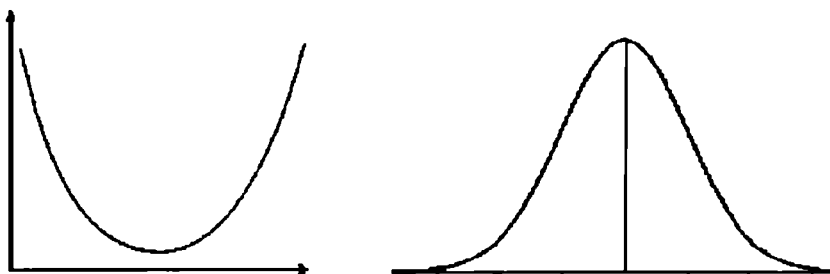
პოპულაციის და შერჩევის განაწილება

სიხშირეთა განაწილებას აგებენ როგორც შერჩევისათვის, ასევე პოპულაციისათვის, ანუ განიხილავენ შერჩევისა და პოპულაციის განაწილებას. არსებობს გამოთქმა, რომ შერჩევის განაწილება პოპულაციის განაწილების ბუნდოვან-

ნი სურათია. შერჩევის მოცულობის (n -ის) გაზრდასთან ერთად, შერჩევის პროპორციები ინტერვალებში უახლოვდებიან პოპულაციის პროპორციებს. ამგვარად, სურათი უფრო ნათელი ხდება და შერჩევის განაწილება უფრო და უფრო ემსგავსება პოპულაციის განაწილებას.

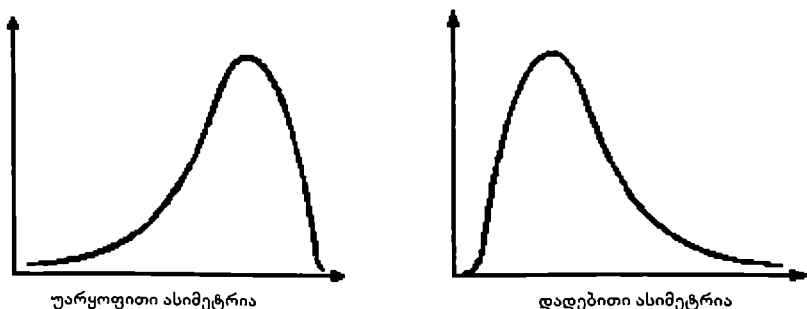
პოპულაციის განაწილებამ შეიძლება მიიღოს ნებისმიერი ფორმა არსებული მთელი რიგი სხვადასხვაგვარი ფორმებიდან.

ნახ. 2.6-ზე გამოსახულია ორი განაწილება – U -სებური და ზარისებური. U -სებურ განაწილებაზე მონაცემთა უმრავლესობა თავმოყრილია მონაცემთა დიაპაზონის ბოლოებში. ზარისებურ განაწილებაში მონაცემთა უმრავლესობა თავმოყრილია დიაპაზონის შუაში.



ნახ. 2.6. U -სებური და ზარისებური განაწილება

2.6 ნახაზზე გამოსახულ სიხშირეთა განაწილებებს *სიმეტრიული* ეწოდება. მონაცემების განაწილებები ყოველთვის არ არის სიმეტრიული. ხშირად, სიხშირეთა განაწილების კუდი გადახრილია რომელიმე მხარეს. ასეთ განაწილებას *ასიმეტრიულს* უწოდებენ. უნდა გავარჩიოთ მარცხნივ ასიმეტრიული და მარჯვნივ ასიმეტრიული განაწილებები.



ნახ. 2.7. ასიმეტრიული განაწილებები

შეჯამება

სიხშირეთა განაწილება არის ცხრილი, რომელიც გვიჩვენებს მონაცემთა სიხშირებს, ფარდობით სიხშირებს, პროცენტებში გამოსახულ სიხშირებს.

პისტოგრამა – რაოდენობრივი მონაცემების სიხშირეთა განაწილების გრაფიკული გამოსახვაა.

სვეტოვანი დიაგრამა – თვისებრივი მონაცემების სიხშირეთა განაწილების გრაფიკული გამოსახვაა.

ფოთლებიანი ლეროების მსგავსი დიაგრამა – რაოდენობრივი მონაცემების სიხშირეთა განაწილების გრაფიკული გამოსახვის ალტერნატიული საშუალებაა.

სიხშირეთა განაწილების მსგავსად, განაწილების გრაფიკები მონაცემებზე ერთსა და იმავე ინფორმაციას გვაწვდის. გრაფიკების უპირატესობა ისაა, რომ ხშირად გაცილებით იოლია გრაფიკულად წარმოდგენილი ინფორმაციის აღქმა. სიხშირეთა განაწილებას აგებენ როგორც შერჩევისათვის, ასევე პოპულაციისათვის.

ძირითადი ტერმინები

სიხშირეთა განაწილება, პისტოგრამა, პოლიგონი, სვეტოვანი დიაგრამა, წრიული დიაგრამა, ფოთლებიანი ლეროების მსგავსი დიაგრამა.

კითხვები

1. მოცემული მონაცემები იძლევა ორი წლის ბავშვის სიტყვათა მარაგის მოცულობას: 205 377 292 300 179 240 300 190 680 250 280 170 211 266 303 350 375 288 360 225.
 - ა. როგორია ეს მონაცემები, – დისკრეტული თუ უწყვეტი?
 - ბ. ააგეთ სიხშირეთა განაწილება. აიღეთ 5 ან 6 ინტერვალი.
 - გ. როგორია ამ განაწილების ფორმა?
2. ქვემოთ მოყვანილ ცხრილში მოცემულია სხვადასხვა რეგიონში დაბადებული მოქალაქეების რაოდენობა (მილიონობით) 1990 წლის მონაცემებით:

დაბადების ადგილი	რაოდენობა
ეროპა	4.0
საქართველო	0.03
აზია	5.0
კანადა	0.7
მექსიკა	4.3
კარიბის კუნძულები	1.9
ცენტრალური ამერიკა	1.1
სამხრეთი ამერიკა	1.0
აფრიკა	0.4
კენია	0.1
სულ	18.53

- ა. დაბადების ადგილის მიხედვით, თვისებრივია თუ რაოდენობრივი ეს მონაცემები?
 - ბ. ააგეთ ფარდობით სიხშირეთა განაწილება;
 - გ. ააგეთ სვეტოვანი დიაგრამა;
 - დ. ცვლადი „დაბადების ადგილი“ თვისებრივია, თუ რაოდენობრივი?
3. ადამიანის განვითარების ინდექსი შედგება სამი კომპონენტისგან: დაბადებისას სიცოცხლის მოსალოდნელი ხანგრძლივობა, განათლების მიღების შესაძლებლობა, ოჯახის შემოსავალი. ინდექსი იღებს მნიშვნელობებს 0-დან 1-მდე. უფრო დიდი ინდექსი შეესაბამება უფრო მაღალ განვითარებას. 1992 წლის მონაცემებით, ცენტრალური ამერიკის რვა ქვეყნის ინდექსები იყო: ბელიზი – 0.884, კოსტა რიკა – 0.884, სალვადორი – 0.579, გვატემალა – 0.591, პონდურასი – 0.578, მექსიკა – 0.842, ნიკარაგუა – 0.611, პანამა – 0.856.
- ა. ააგეთ ფოთლებიანი ღეროების დიაგრამა. რიცხვის ნილადი ნაწილის პირველი ციფრი ჩათვალეთ ღეროდ, ორი დანარჩენი ფოთლად;
 - ბ. როგორია ამ განაწილების ფორმა?
4. ცხრილში მოცემულია რესპონდენტების პასუხები კითხვაზე – „ბოლო 12 თვის მანძილზე თქვენი რამდენი ნაცნობი გახდა ძალადობის მსხვერპლი?“ (1990 წლის მონაცემები).

მსხვერპლი	სიხშირე	პროცენტი
0	1244	90.8
1	81	5.9
2	27	2.0
3	11	0.8
4	4	0.3
5	2	0.1
6	1	0.1

- ა. რამდენმა კაცმა შემოხაზა პასუხებში „0“, რამდენმა შემოხაზა „6“?
 - ბ. აღწერეთ სიხშირეთა განაწილების ფორმა.
5. ფოთლებიანი ღეროების მსგავს დიაგრამაზე გამოსახულია სტატისტიკის გამოცდაში მიღებული ქულები (მაქსიმალური შეფასება - 30).

ღერო	ფოთოლი
0	6 9 9
1	0 1 2 5 8 8 8 9 9 9 9
2	1 2 2 3 4 5 5 6 7 7 8 8 9 9 9 9 9
3	0 0 0 0 0 0

- ა. რომელია ყველაზე მაღალი და დაბალი ქულა?
- ბ. რამდენმა სტუდენტმა მიიღო უმაღლესი შეფასება?
- გ. რამდენმა სტუდენტმა მიიღო 18-ზე ნაკლები ქულა?
- დ. რამდენი სტუდენტი ამ ჯგუფში?

6. გამოკითხეს 50 სტუდენტი, ყოველდღიურად კომპიუტერთან გატარებული საათების რაოდენობის შესახებ. დასახელებული დრო (წუთებში) სიხშირეთა შემდეგი განაწილების სახითაა მოცემული:

ინტერვალი	სიხშირე
0-29	1
30-59	7
60-89	5
90-119	9
120-149	2
150-179	7
180-209	3
210-239	8
240-269	9

- ა. რა სიდიდის ინტერვალია გამოყენებული ამ ცხრილში?
 ბ. გამოთვალეთ ფარდობითი სიხშირეები;
 გ. აღწერეთ განაწილების ფორმა.
7. ცხრილში მოცემულია სტატისტიკაში, შუალედურ გამოცდაში სტუდენტების მიერ მიღებული ქულების განაწილება:

ქულები	სიხშირე
0-4	5
5-9	10
10-14	15
15-20	20

- ა. რა სიდიდის ინტერვალია გამოყენებული?
 ბ. რამდენი სტუდენტი იყო შეფასებული?
 გ. რას ნიშნავს მნიშვნელობა „5“ ინტერვალისთვის 0 - 4?
8. ბიოლოგმა გაზომა 1 000 ერთი და იმავე ჯიშის დაჭერილი თევზის სიგრძე (სმ) და ააგო ფარდობითი სიხშირეთა შემდეგი განაწილება:

სიგრძე	ფარდობითი სიხშირე
8-10	0.22
11-13	0.15
14-16	0.25
17-19	0.24
20-22	0.06
23-25	0.05
26-28	0.03

- ა. ააგეთ სიხშირეთა განაწილება და შესაბამისი ჰისტოგრამა;
 ბ. შესაძლებელია თუ არა, ამ მონაცემებზე დაყრდნობით, ფოთლებიანი ღეროების მსგავსი დიაგრამის აგება?

9. გაყიდვების მენეჯერმა შეაგროვა 15 მაღაზიის მონაცემები კვირაში სამუშაო საათების რაოდენობისა და საღაროში შემოსული თანხის შესახებ:

მაღაზია	თანხა	საათი	მაღაზია	თანხა	საათი
1	23 300	120	9	27 886	140
2	25 600	135	10	54 156	300
3	19 200	96	11	34 080	254
4	10 211	102	12	25 900	180
5	19 330	240	13	36 400	270
6	35 789	190	14	25 760	175
7	12 540	108	15	31 500	256
8	43 150	234			

- ა. ააგეთ შემოსავლების და სამუშაო საათების ფოთლებიანი ღეროების მსგავსი ორი დიაგრამა;
- ბ. რა კომენტარის გაკეთება შეგიძლიათ?

თავი 3

აღწერითი სტატისტიკაში.

ცენტრალური ტენდენციის საზომები

§1. საშუალო

ახლა ჩვენ ვინყებთ იმ სტატისტიკური მეთოდების შესწავლას, რომლებიც მონაცემების დაჯამებისა და შესაბამისი დასკვნების გაკეთების საშუალებას მოგვცემენ. ამ მეთოდებს შორის ყველაზე ცნობილია საშუალოს მოძებნის მეთოდი. პოპულარულ ენაზე საშუალო ნიშნავს განაწილების ცენტრს ან, უფრო, ტიპურ შემთხვევას. საშუალოს საზომებს აგრეთვე უწოდებენ ცენტრალური ტენდენციის საზომებს და მათ მიეკუთვნება: საშუალო, მედიანა და მოდა.

მონაცემთა სიმრავლის საშუალოს ცოდნა არ არის საკმარისი მონაცემთა სიმრავლის სრულად აღსაწერად. დამატებით, საშუალოს ცოდნასთან ერთად, საჭიროა ვიცოდეთ, თუ როგორაა მონაცემთა მნიშვნელობები გაბნეული. უფრო ზუსტად, არის თუ არა მონაცემთა მნიშვნელობები დაჯგუფებული (თავმოყრილი) საშუალოს ირგვლივ, თუ ისინი უფრო თანაბრად არიან გაბნეული მთელი განაწილების გასწვრივ? საზომებს, რომლებიც განსაზღვრავენ მონაცემთა მნიშვნელობების გაფანტულობას (გაბნევას), უწოდებენ გადახრის ან გაფანტულობის (გაბნევის) საზომებს. ამ ტიპის საზომებია: გაბნევის დიაპაზონი, სტანდარტული გადახრა და საშუალო კვადრატული გადახრა. გარდა ამისა, გამოიყენება ე.წ. მონაცემთა პოზიციის მაჩასიათებლები. ისინი გვიჩვენებენ სპეციფიკური მნიშვნელობების მდებარეობას მონაცემთა სიმრავლეში ან მათ ფარდობით მდებარეობას მონაცემთა სხვა მნიშვნელობებთან მიმართებაში. ყველაზე გავრცელებული პოზიციის მაჩასიათებლებია: პროცენტილები და რანგები. ეს მაჩასიათებლები ინტენსიურად გამოიყენება ფსიქოლოგიაში და განათლებაში.

საზომებს (მაჩასიათებლებს), რომელთა საპოვნელად გამოიყენება პოპულაციის ყველა მონაცემის მნიშვნელობები, ეწოდება პარამეტრები, ხოლო საზომებს (მაჩასიათებლებს), რომელთა მისაღებად გამოიყენება შერჩევის მონაცემების მნიშვნელობები, სტატისტიკები ეწოდება.

დამრგვალების ზოგადი წესი: სტატისტიკაში დამრგვალების ძირითადი წესი ასეთია: იმაში, სანამ გამოთვლები გრძელდება დამრგვალება არ უნდა მოხდეს, ვიდრე არ იქნება მიღებული საბოლოო პასუხი. როდესაც დამრგვალება ხდება შუალედურ ეტაპზე, მაშინ იზრდება განსხვავება პასუხსა და ზუსტ შედეგს შორის.

საშუალო, რომელიც აგრეთვე ცნობილია როგორც არითმეტიკული საშუალო, მიიღება, თუ შევკრებთ მონაცემების მნიშვნელობებს და გავყოფთ მნიშვნელობების საერთო რაოდენობაზე. მაგალითად: 3, 2, 6, 5 და 4-ის საშუალო მიიღება შეკრებით $3+2+6+5+4 = 20$ და 5-ზე გაყოფით. შესაბამისად, ამ მონაცემების საშუალოა $20/5 = 4$. მონაცემების მნიშვნელობები გამოისახება x -ებით. მონაცემთა ამ

სიმრავლეში $x_1 = 3$, $x_2 = 2$, $x_3 = 6$, $x_4 = 5$ და $x_5 = 4$. x -ის ყველა მნიშვნელობა-თა ჯამის აღსანიშნავად გამოიყენება $\sum$ სიმბოლო (პერძნული დიდი სიგმა ასო) და $\sum x$ - აღნიშნავს მონაცემთა სიმრავლის ყველა x მნიშვნელობათა ჯამს.

საშუალო (mean) არის მნიშვნელობათა ჯამი გაყოფილი მნიშვნელობათა საერთო რაოდენობაზე. შერჩევის საშუალოს აღნიშნავენ $\bar{x}$ სიმბოლოთი:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i,$$

სადაც n გვიჩვენებს მნიშვნელობათა საერთო რაოდენობას შერჩევაში.

თუ შერჩევაში x_i მონაცემის სიხშირეა f_i , x_2 მონაცემის სიხშირეა f_2 და ა. შ. x_k მონაცემის სიხშირეა f_k , მაშინ საშუალო გამოითვლება შემდეგნაირად:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = \left(\sum_{i=1}^k f_i x_i \right) / \left(\sum_{i=1}^k f_i \right).$$

მაგალითი 3.1.1. ფირმის წლიური საპროცენტო შემოსავალი ბოლო ათი წლის განმავლობაში იყო: 8.1, -6.2, 20.9, -2.7, 33.6, 42.9, 24.4, 5.2, 3.1, 30.5. როგორია ფირმის საშუალო წლიური შემოსავალი ამ ათი წლის განმავლობაში?

ცხადია, რომ

$$\bar{x} = (8.1 + (-6.2) + 20.9 + (-2.7) + 33.6 + 42.9 + 24.4 + 5.2 + 3.1 + 30.5) / 10 = 16.$$

ე. ი. საშუალო წლიური შემოსავალი არის 16%.

არითმეტიკულ საშუალოს შეიძლება მიეცეს შემდეგი მარტივი ინტერპრეტაცია: არითმეტიკული საშუალო არის შერჩევის ერთ ინდივიდზე ან ობიექტზე მოსული ჯამურ დაკვირვებათა წილი (საშუალო რაოდენობა).

მაგალითი 3.1.2. 10 ოჯახში ბავშვთა რაოდენობა:

$$3, 2, 1, 1, 2, 3, 3, 1, 2, 2.$$

მაშინ ერთ ოჯახზე მოსული ბავშვების საშუალო რაოდენობა იქნება:

$$(3 + 2 + 1 + 1 + 2 + 3 + 3 + 1 + 2 + 2) / 10 = 2.$$

ამრიგად, თუ 10 ოჯახს ჯამში ყავს 20 ბავშვი, მაშინ არითმეტიკული საშუალო უდრის ბავშვთა იმ რაოდენობას, რომელიც მოვიდოდა ერთ ოჯახზე იმ შემთხვევაში, ბავშვები თანაბრად რომ გაენაწილებინათ.

საშუალოს დამრგვალების წესი: საშუალო უნდა დამრგვალდეს ერთი ათობითი ნიშნით მეტზე, ვიდრე ნედლი მონაცემებია წარმოდგენილი. ამ წესის არსი ისაა, რომ განასხვავოს საშუალო ნედლი მონაცემებისაგან.

დაჯგუფებული მონაცემების შემთხვევაში საშუალოს პოვნის პროცედურა იყენებს კლასების შუაწერტილებს. ამ პროცედურას ჩვენ ქვემოთ შევხებით.

მაგალითი 3.1.3. მოცემულია კეირის განმავლობაში შემთხვევით შერჩეული 20 მორბენალის მიერ გარბენილი მანძილების სიხშირული განაწილების ცხრილი

კლასის საზღვრები	სიხშირე
5.5-10.5	1
10.5-15.5	2
15.5-20.5	3
20.5-25.5	5
25.5-30.5	4
30.5-35.5	3
35.5-40.5	2
	$n = 20$

ვიპოვოთ საშუალო.

ამოხსნა:

ნაბიჯი 1. ცხრილს დავამატოთ შუანერტილებისა და შუანერტილების სიხშირეებზე ნამრავლის სვეტები.

ნაბიჯი 2. ვიპოვოთ თითოეული კლასის შუანერტილი და შევიტანოთ C სვეტში:

$$x_m = \frac{5.5+10.5}{2} = 8, \quad \frac{10.5+15.5}{2} = 13, \text{ და ა. შ.}$$

ნაბიჯი 3. ყოველი კლასის შესაბამისი სიხშირე გავამრავლოთ კლასის შუანერტილზე, როგორც ქვემოთაა ნაჩვენები და შედეგები შევიტანოთ D სვეტში: $1 \cdot 8 = 8$, $2 \cdot 13 = 26$ და ა. შ. დასრულებულ ცხრილს ექნება შემდეგი სახე:

A კლასები	B სიხშირე (f)	C შუანერტილი (x_m)	D $f \cdot x_m$
5.5—10.5	1	8	8
10.5—15.5	2	13	26
15.5—20.5	3	18	54
20.5—25.5	5	23	115
25.5—30.5	4	28	112
30.5—35.5	3	33	99
35.5—40.5	2	38	76
	$n = 20$		$\sum f \cdot x_m = 490$

ნაბიჯი 4. ვიპოვოთ D სვეტის მნიშვნელობათა ჯამი.

ნაბიჯი 5. საშუალოს მისაღებად მიღებული ჯამი გავყოთ n -ზე:

$$\bar{x} = \frac{\sum f \cdot x_m}{n} = \frac{490}{20} = 24.5 \text{ მილი.}$$

მაგალითი 3.1.4. მეათე კლასელი 204 მოსწავლის ძმებისა და დების რიცხვის ქვემოთ მოყვანილი სიხშირული ცხრილის მიხედვით გამოთვალეთ საშუალო.

ძმებისა და დების რაოდენობა, x_i	სიხშირე, f_i
0	36
1	94
2	48
3	15
4	7
5	3
6	1
	სულ: 204

ამოხსნა: დავამატოთ სიხშირული განაწილების ცხრილს მესამე სვეტი, რომელშიც მოხდება მონაცემებისა და მათი სიხშირეების გამრავლება $f_i x_i$:

ძმებისა და დების რაოდენობა, x_i	სიხშირე, f_i	$f_i x_i$
0	36	0
1	94	94
2	48	96
3	15	45
4	7	28
5	3	15
6	1	6
ჯამი	$\sum_{i=1}^k f_i = 204$	$\sum_{i=1}^k f_i x_i = 284$

$$\text{შესაბამისად, საშუალო იქნება } \bar{x} = \left(\sum_{i=1}^7 f_i x_i \right) / \left(\sum_{i=1}^7 f_i \right) = 284 / 204 = 1.39.$$

შენიშვნა. ზოგიერთ შემთხვევაში ჩვენ შეიძლება დაგვჭირდეს საშუალოს შეფასება დაგროვილ სიხშირეთა დიაგრამის მიხედვით. ამ შემთხვევაში დაგროვილ სიხშირეთა დიაგრამიდან წინასწარ შევადგინოთ სიხშირეთა ცხრილი, ხოლო შემდგომ ვიმოქმედოთ ჩვეულებრივი გზით. სიხშირული ცხრილის მისაღებად საჭიროა: ა) აბსცისთა ლერძზე მოვნიშნოთ მონაცემთა უმცირესი და უდიდესი მნიშვნელობები - $x_{\min}$ და $x_{\max}$; ბ) მათ შორის მოთავსებული ინტერვალი დავეყოთ თანაბარი სიგრძის (მოხერხებული რაოდენობა) ინტერვალებად და მოვნიშნოთ მათი საზღვრები - $x_1, x_2, \dots, x_k$; გ) თითოეული მონიშნული წერტილიდან გავავლოთ ვერტიკალური ხაზები გრაფიკის გადაკვეთამდე და მოვძებნოთ შესაბამისი

გამოყენებითი სტატისტიკის საფუძვლები

დაგროვილი სიხშირეები (გრაფიკთან გადაკვეთის წერტილებიდან გავავლოთ პორიზონტალური ხაზები ორდინატა ღერძის გადაკვეთამდე) - $y_{\min} = 0$, y_1 , $y_2, \dots, y_k, y_{\max}$; დ) შევადგინოთ ცხრილი:

მონაცემები, რომლებიც	დაგროვილი სიხშირე
$< x_{\min}$	0
$< x_1$	y_1
$< x_2$	y_2
...	...
$< x_k$	y_k
$< x_{\max}$	$y_{\max}$

ე) ავაგოთ სიხშირული განაწილების ცხრილი:

კლასის საზღვრები	შუა ინტერვალური მნიშვნელობა	სიხშირე, f_i
$x_{\min} \leq x < x_1$	$(x_{\min} + x_1)/2$	y_1
$x_1 \leq x < x_2$	$(x_1 + x_2)/2$	$y_2 - y_1$
$x_2 \leq x < x_3$	$(x_2 + x_3)/2$	$y_3 - y_2$
...	...	...
$x_{k-1} \leq x < x_k$	$(x_{k-1} + x_k)/2$	$y_k - y_{k-1}$
$x_k \leq x < x_{\max}$	$(x_k + x_{\max})/2$	$y_{\max} - y_k$

მოვიყვანოთ საშუალოს თვისებები:

1. მონაცემთა დაკვირვებული მნიშვნელობების საშუალოდან გადახრების ჯამი

ნულის ტოლია:
$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

2. ნებისმიერი a რიცხვისათვის:
$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - a)^2.$$

3. თუ არსებული $x_1, \dots, x_n$ მონაცემებიდან გადავალთ ახალ $y_1 = ax_1 + b, \dots, y_n = ax_n + b$ მონაცემებზე, სადაც a და b ნამდვილ რიცხვთა ნებისმიერი

წყვილია, მაშინ
$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = a\bar{x} + b.$$

4. თუ $x_1, \dots, x_n$ და $y_1, \dots, y_n$ ტოლი მოცულობის ორი შერჩევაა და გადავალთ ახალ შერჩევაზე $z_1 = ax_1 \pm by_1, \dots, z_n = ax_n \pm by_n$, მაშინ $\bar{z} = a\bar{x} \pm b\bar{y}$.
5. თუ $x_1, \dots, x_n$ და $y_1, \dots, y_m$ ორი შერჩევაა, მაშინ $x_1, \dots, x_n, y_1, \dots, y_m$ გაერთიანებული შერჩევის საშუალო მოიცემა ფორმულით:

$$\bar{z} = \frac{n}{n+m} \bar{x} + \frac{m}{n+m} \bar{y}.$$

მიუხედავად იმისა, რომ საშუალო ადვილი გამოსათვლელია, გააჩნია მარტივი ინტერპრეტაცია – მონაცემთა ახალი მასივის გაჩენისას ინარჩუნებს არითმეტიკულ ოპერაციებს და წარმოადგენს მომაცემთა ცენტრის ერთ-ერთ ყველაზე გავრცელებულ რიცხვით მახასიათებელს. აღსანიშნავია, რომ საშუალოს გააჩნია ერთი სერიოზული ნაკლი – ის ზედმეტად რეაგირებს ექსტრემალური დაკვირვებების უმნიშვნელო რაოდენობაზეც კი. ასეთი დაკვირვებები შეიძლება იყოს ორი ტიპის. პირველ ტიპს მიეკუთვნება ე. წ. „ამოვარდნილი“ დაკვირვებები ანუ ის დაკვირვებები (მონაცემები), რომლებიც მკვეთრად განსხვავებული მონაცემთა ძირითადი მასივისაგან. ამ შემთხვევაში საშუალო მკვეთრად იქნება წანაცვლებული „ამოვარდნილი“ დაკვირვებების მიმართულებით. ექსტრემალურ დაკვირვებებთან საქმე გვაქვს იმ შემთხვევაშიც, როცა მონაცემებით აგებული (გაგლეჯებული) ჰისტოგრამა ძლიერ ასიმეტრიულია რომელიმე მიმართულებით, ანუ მას გააჩნია რომელიმე მიმართულებით განელილი გრძელი „კუდი“, ხოლო სანინალმდეგო მიმართულებით კი – „მოკლე“ კუდი. ამ შემთხვევაშიც საშუალო წანაცვლებულია გრძელი „კუდის“ მიმართულებით და მონაცემთა ძირითადი მასა მის ერთ მხარეს აღმოჩნდება. ასეთი მონაცემებისათვის საშუალო არ წარმოადგენს ცენტრალური ტენდენციის საიმედო საზომს.

მაგალითი 3.1.5. ვთქვათ, მუსიკალურ ნაწარმოებთა კატალოგიდან ამოვარჩიეთ ხუთი ნაწარმოები, რომელთა ხანგრძლივობებია (წუთებში): 37, 46, 40, 57 და 50. მაშინ $\bar{x} = 46$. თუ მეხუთე ნაწარმოებს, რომლის ხანგრძლივობა იყო 50 წუთი, შევცვლით სხვა ნაწარმოებით, რომლის ხანგრძლივობაა 200 წუთი, მაშინ საშუალო იქნება $\bar{x} = 76$. ანუ საშუალომ გადაინაცვლა ექსტრემალური მნიშვნელობისაკენ. ყველა დანარჩენი მონაცემი საგრძნობლად ნაკლებია მასზე. ამ შემთხვევაში $\bar{x}$ არ არის განლაგების (ლოკალიზაციის) მდგრადი მახასიათებელი. თუ მეხუთე მონაცემს შევცვლით 400-ით, მაშინ $\bar{x} = 116$.

მკვეთრად ასიმეტრიული (მარჯვნივ) ფორმა აქვს მოსახლეობის შემოსავლების სიხშირულ განაწილებებს, ამიტომ $\bar{x}$ მკვეთრად წანაცვლებული აღმოჩნდება მარჯვნივ (ექსტრემალურად დიდი შემოსავლების მიმართულებით) და გასაშუალოებული შემოსავლების გამოყენება ცხოვრების დონის აღსაწერად, აზრს მოკლებულია.

მაგალითი 3.1.6. ქვევით მოყვანილია ავსტრალიის ერთ-ერთი უნივერსიტეტის თანამშრომელთა წლიური შემოსავლები 1000 დოლარებში: 28, 109, 26, 32, 30, 26, 29. მაშინ საშუალო იქნება $\bar{x} = 40$. როგორც ვხედავთ, ყველა მონაცემი, გარდა ერთისა (109), 40-ზე ნაკლებია და, ამდენად, საშუალო არ აღწერს ტიპობრივ შემოსავალს.

§2. მედიანა

მედიანა (median) განეკუთვნება მონაცემთა მდებარეობის ცენტრალური ტენდენციის მდგრადი საზომების ჯგუფს (განსხვავებით საშუალოსაგან). მონაცემთა რაიმე რიცხვითი მახასიათებლის **მდგრადობა** ნიშნავს, რომ დაკვირვებათა მცირე რაოდენობის ცვლილებას შეზღუდული გავლენა აქვს აღნიშნულ მახასიათებელზე, მიუხედავად იმისა, თუ როგორია ამ ცვლილების სიდიდე. სწორედ ამ თვისების გამო, მედიანა, საშუალოს შემდეგ, წარმოადგენს ცენტრალური ტენდენციის ყველაზე გავრცელებულ საზომს.

როგორც ცნობილია, გეომეტრიაში მედიანა არის სამკუთხედის წვეროს მოპირდაპირე გვერდის შუანერტილთან შემაერთებელი მონაკვეთი, ხოლო ფართო გაგებით, მედიანა არის "შუა"-ს სინონიმი. სტატისტიკაშიც მედიანა განიმარტება, როგორც მონაცემთა ვარიაციული მწკრივის შუაში მდგომი მნიშვნელობა იმ აზრით, რომ მასზე ნაკლები რიცხვითი მნიშვნელობის მონაცემთა რაოდენობა მასზე დიდი რიცხვითი მნიშვნელობის მონაცემთა რაოდენობის ტოლია. გასაგებია, რომ თუ დაკვირვებულ მონაცემთა რაოდენობა კენტია, მაშინ მედიანა მართლაც არის ვარიაციული მწკრივის შუა ელემენტი. თუ დაკვირვებათა რაოდენობა ლუნია $n = 2k$, მაშინ მედიანის განმარტებას დააკმაყოფილებს ნებისმიერი რიცხვი (x_k, x_{k+1}) რიცხვითი შუალედიდან, ე. ი. ამ შემთხვევაში მედიანა ცალსახად არ განიმარტება. შესაბამისად, ამ გაუგებრობის თავიდან აცილების მიზნით, მიღებულია შემდეგი შეთანხმება: როცა $n = 2k$, მაშინ მედიანას უწოდებენ ვარიაციული მწკრივის შუაში მდგომი ორი ელემენტის საშუალო არითმეტიკულს.

დავუშვათ, რომ $x_1, \dots, x_n$ შერჩევის ელემენტებია, ხოლო $x_{(1)}, \dots, x_{(n)}$ კი — შესაბამისი ვარიაციული მწკრივია. მედიანა აღინიშნება $\tilde{x}$ (ან m_e) სიმბოლოთი და ის აიგება შემდეგი წესით:

$$\tilde{x} = x_{((n+1)/2)}, \text{ როცა } n \text{ კენტია, და } \tilde{x} = \frac{x_{(n/2)} + x_{((n/2)+1)}}{2}, \text{ როცა } n \text{ ლუნია.}$$

ამრიგად, მედიანის მოსაძებნად საჭიროა შემდეგი ნაბიჯები: I. დავალაგოთ ნედლი მონაცემები ზრდადობის მიხედვით (ავაგოთ ვარიაციული მწკრივი) და II. შევარჩიოთ (ან ავაგოთ) შუანერტილი. როგორც ვხედავთ, მედიანის მოსაძებნად საჭიროა ვარიაციული მწკრივის აგება (განსხვავებით საშუალოს მოძებნის შემთხვევისაგან), რაც დაკვირვებათა დიდი რაოდენობის დროს გარკვეულ სიძნელეებთანაა დაკავშირებული.

მაგალითი 3.2.1. ქვემოთ მოყვანილია შვიდი სამხედრო ახალწვეულის წონა (ფუნტებში, 1 ფუნტი = 453,6 გრამი):

180, 201, 220, 191, 219, 209, 186.

იპოვეთ მედიანა.

ამოხსნა:

ნაბიჯი 1. დავალაგოთ მონაცემები ზრდადობის მიხედვით:

180, 186, 191, 201, 209, 219, 220.

ნაბიჯი 2. შევარჩიოთ შუანერტილი. ვინაიდან $n = 7$, ამიტომ მედიანა იქნება

$$\bar{x} = x_{((7+1)/2)} = x_{(4)} = 201.$$

მაგალითი 3.2.2. აშშ-ში რვა წლის განმავლობაში მომხდარი ტორნადოების რაოდენობებია:

684, 764, 656, 702, 856, 1133, 1132, 1303.

იპოვეთ მედიანა.

ამოხსნა. ვარიაციული მწკრივი იქნება:

656, 684, 702, 764, 856, 1132, 1133, 1303.

ვინაიდან შუანერტილი მდებარეობს 764-სა და 856-ს შორის, ამიტომ მედიანას ეპოულობთ ამ ორი მნიშვნელობის შეკრებითა და 2-ზე გაყოფით:

$$\bar{x} = \frac{x_{(8/2)} + x_{((8/2)+1)}}{2} = \frac{x_{(4)} + x_{(5)}}{2} = \frac{764 + 856}{2} = 810,$$

ე. ი. ტორნადოების რაოდენობის მედიანაა 810.

მოვიყვანოთ მედიანის თვისებები:

I. შერჩევის მედიანა წარმოადგენს შემდეგი განტოლების ამოხსნას:

$$\sum_{i=1}^n \text{sign}(x_i - \bar{x}) = 0,$$

სადაც $\text{sign}x = 1$, თუ $x > 0$; $\text{sign}0 = 0$ და $\text{sign}x = -1$, თუ $x < 0$.

II. ნებისმიერი a რიცხვისათვის ადგილი აქვს უტოლობას:

$$\sum_{i=1}^n |x_i - \bar{x}| \leq \sum_{i=1}^n |x_i - a|.$$

III. თუ არსებული $x_1, \dots, x_n$ მონაცემებიდან გადავალთ ახალ $z_1 = ax_1 + b, \dots, z_n = ax_n + b$ მონაცემებზე (სადაც a და b ნამდვილ რიცხვთა ნებისმიერი წყვილია), მაშინ: $\bar{z} = a\bar{x} + b$.

მაგალითი 3.2.3. დაეუბრუნდეთ §1-ის მე-5 მაგალითის მონაცემებს და წარმოვადგინოთ ვარიაციული მწკრივის სახით:

$$x_{(1)} = 26, x_{(2)} = 26, x_{(3)} = 28, x_{(4)} = 29, x_{(5)} = 30, x_{(6)} = 32, x_{(7)} = 109.$$

მაშინ, ვინაიდან $n = 8$ კენტია, მედიანა იქნება $\bar{x} = x_{((7+1)/2)} = x_{(4)} = 29$. გავიხსენოთ, რომ საშუალო იყო $\bar{x} = 40$, საიდანაც ჩანს, რომ მედიანა ამ მაგალითში უკეთ აღწერს ტიპობრივ შემოსავალს, ვიდრე საშუალო.

ახლა მოვახდინოთ დაკვირვებათა ცვლილება და შევხედოთ როგორ იქცევა საშუალო და მედიანა. ვთქვათ, დაკვირვება $x_{(7)} = 109$ შეეცვალეთ 200-ით, მაშინ შეცვლილი მონაცემებისათვის მედიანა დარჩება იგივე $\bar{x} = 29$, ხოლო საშუალო წაინაცვლებს დიდი დაკვირვების (200-ის) მიმართულებით $\bar{x} = 53$. ამდენად, მედიანა მდებარეობის უფრო მდგრადი მახასიათებელია, ვიდრე საშუალო.

ვთქვათ, ახლა დაკვირვება $x_{(2)} = 26$ შეეცვალეთ 300-ით. მაშინ ახალი მონაცემების ვარიაციული მწკრივი იქნება:

$$x_{(1)} = 26, x_{(2)} = 28, x_{(3)} = 29, x_{(4)} = 30, x_{(5)} = 32, x_{(6)} = 109, x_{(7)} = 300.$$

ამ შემთხვევაში მედიანა იცვლება მხოლოდ ერთი ერთეულით $\bar{x} = x_{(4)} = 30$, მაშინ, როდესაც საშუალო კვლავ მკვეთრად იზრდება $\bar{x} = 78.85$, რაც ისევ მეტყველებს მედიანის, როგორც ცენტრის მახასიათებლის, მდგრადობაზე.

მედიანას გააჩნია კიდევ ერთი სასარგებლო თვისება: მის გამოსათვლელად საჭიროა ვარიაციული მწკრივის მხოლოდ შუაში მოქცეული ინდივიდის (თუ n კენტია) ან ორი შუა ინდივიდის (თუ n ლუნია) გაზომვა.

მედიანა განსაკუთრებით მნიშვნელოვან ინფორმაციას იძლევა მაშინ, როცა მოცემულ შერჩევაში დაკვირვებების შედარებით მცირე რაოდენობა მნიშვნელოვნად განსხვავდება დანარჩენი სიმრავლისაგან. მაგალითად, მოსახლეობის გარკვეული ჯგუფის შემოსავლების განაწილების მონაცემებში მედიანა მოსახლეობის ამ ჯგუფის ცხოვრების დონის უფრო აზრიან სურათს იძლევა, ვიდრე საშუალო (რომელზედაც დიდ გავლენას ახდენს ერთი ინდივიდის შემოსავალი).

თუმცა ეს არ უნდა გავიგოთ ისე, რომ მედიანას ყოველთვის უნდა მიენიჭოს უპირატესობა საშუალოსთან შედარებით. საშუალო გამოითვლება ყველა მონაცემის მეშვეობით (ყველა მონაცემის სიდიდის გათვალისწინებით) და შეიცავს მეტ ინფორმაციას, ვიდრე მედიანა. მისი მგრძნობიარობა კი, ხშირ შემთხვევაში, მას მოცემული შერჩევის კარგ მახასიათებლად აქცევს. საშუალო შეიძლება გამოყენებულ იქნეს როგორც საზომი, რომელიც ასახავს, თუ საშუალოდ რა სიდიდისაა შერჩევის ელემენტები. ორი სხვადასხვა ჯგუფის ელემენტების შედარების მიზნით, ჩვეულებრივ, პირველ რიგში, ადარებენ მათ არითმეტიკულ საშუალოებს.

მაგალითი 3.2.4. მოცემულია ორი – A და B ფირმის შემოსავლიანობა (პროცენტებში) ბოლო 10 წლის განმავლობაში:

ფირმა: 8.3, -6.2, 20.9, -2.7, 33.6, 42.9, 24.4, 5.2, 3.1, 30.5;

ფირმა: 12.1, 6.4, 12.2, 27.8, 25.7, 18.2, 10.7, -1.3, 11.4, -2.8.

თუ გამოვთვლით საშუალოებს A და B ფირმების შემოსავლიანობებისათვის, მივიღებთ:

$$\bar{x}_A = 16\% \quad \text{და} \quad \bar{x}_B = 12\%.$$

ამ ორი საშუალოს შედარება გვაძლევს საფუძველს დავასკვნათ, რომ A ფირმის შემოსავლიანობა საშუალოდ უფრო მაღალია, ვიდრე B ფირმისა (შემდგომი ანალიზი კი წარმოებს სხვა მახასიათებლების საშუალებით).

ხშირად მონაცემთა სიმრავლე საკმაოდ დიდია და ის მოცემულია სიხშირეთა ცხრილის საშუალებით.

მაგალითი 3.2.5. გამოვთვალოთ მედიანა 204 მეათეკლასელი მოსწავლის ძმებისა და დების რიცხვის ზემოთ მოყვანილი სიხშირული ცხრილის მიხედვით.

ამოხსნა. მედიანის მოძებნის ტრადიციული მეთოდი გულისხმობს ვარიაციული მწკრივის შედგენას და მისი შუა მონაცემის პოვნას. ამ შემთხვევაში უნდა გავამწკრივოთ ერთ სტრიქონში 36 ცალი 0, შემდგომ 94 ცალი 1 და ა. შ. ბოლოს 1 ცალი 6; მოვძებნოთ ბოლოებიდან თანაბრად დაშორებული $\frac{1}{2} \times 204 = 102$ -ე და 103-ე მნიშვნელობები და გამოვთვალოთ მათი საშუალო არითმეტიკული. გაცილებით მარტივი მეთოდია, სიხშირული განაწილების ცხრილს დავამატოთ დაგროვილი სიხშირეების სვეტი, როგორც ეს ქვემოთაა გაკეთებული:

ძმებისა და დების რაოდენობა	სიხშირე	დაგროვილი სიხშირე
0	36	36
1	94	130
2	48	178
3	15	193
4	7	200
5	3	203
6	1	204
	სულ: 204	

ამ ცხრილიდან ჩანს, რომ, როცა ჩავალთ 0-ების ბოლოში, ჯერ კიდევ არა ვართ მიღწეული 102-ე მნიშვნელობასთან, მაგრამ 1-იანების ბოლოში უკვე მივალ-ნიეთ 130-ე მნიშვნელობასთან, რაც იმას ნიშნავს, რომ როგორც 102-ე, ისე 103-ე მნიშვნელობა 1-იანია და, შესაბამისად, მედიანა ისევე 1-იანია.

უნყვეტი ტიპის მონაცემთა დიდი მასივების შემთხვევაში, როგორც წესი, ხშირად მოცემული გვაქვს არა კონკრეტული მონაცემები, არამედ ერთმანეთთან ახლოს მყოფი დაჯგუფებული მონაცემები (შესაბამისად, კონკრეტული მნიშვნელობები დაკარგულია). ამ შემთხვევაში არ შეგვიძლია ზუსტად დავადგინოთ მედიანა, მაგრამ შეგვიძლია შევაფასოთ ის.

მაგალითი 3.2.6. განვიხილოთ CD დისკების საკმაოდ დიდი რაოდენობის ჩანაწერის ხანგრძლივობის (წუთებში) სიხშირული განაწილების ზემოთ მოყვანილი ცხრილი და შევაფასოთ ჩანაწერის ხანგრძლივობის მედიანა.

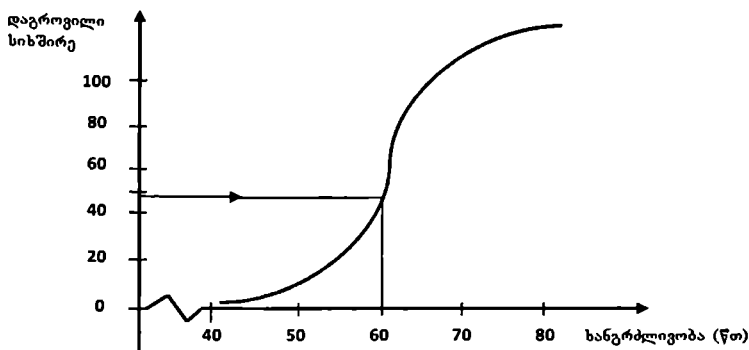
ამოხსნა. სიხშირული განაწილების ცხრილს დავამატოთ დაგროვილი სიხშირეების სვეტი, როგორც ეს ქვემოთაა გაკეთებული:

ჩანაწერის ხანგრძლივობა, x (წთ)	ინტერვალის საზღვრები	სიხშირე	დაგროვილი სიხშირე
40-44	$39.5 \leq x < 44.5$	1	1
45-49	$44.5 \leq x < 49.5$	7	8
50-54	$49.5 \leq x < 54.5$	12	20
55-59	$54.5 \leq x < 59.5$	24	44
60-64	$59.5 \leq x < 64.5$	29	73
65-69	$64.5 \leq x < 69.5$	14	87
70-74	$69.5 \leq x < 74.5$	5	92
75-79	$74.5 \leq x < 79.5$	3	95
		სულ: 95	

მედიანის შეფასება შესაძლებელია დაგროვილ სიხშირეთა დიაგრამის (ოგივას) საშუალებით: – აბსცისთა ლერძზე გადავზომოთ ჩანაწერის ხანგრძლივობა, ორდინატთა ლერძზე კი დაგროვილი სიხშირეები (დაგროვილი სიხშირის მნიშვნელობა გადაიზომება ინტერვალის ზედა საზღვრის გასწვრივ). ავავსოთ ოგივა და მოვძებნოთ დაგროვილ სიხშირეთა ლერძზე ისეთი წერტილი, რომელიც შუაზე

ყოფს ერთობლივ სიხშირეს. ამ შემთხვევაში ეს არის $\frac{1}{2} \times 95 = 47.5$. ამ წერტილის

შესაბამისი აბსცისა კი იქნება მედიანის შეფასება (ორდინატთა ლერძის 47.5-ის შესაბამისი წერტილიდან ვავლებთ პორიზონტალურ ხაზს გრაფიკის გადაკვეთამდე და ვპოულობთ ამ გადაკვეთის წერტილის შესაბამის აბსცისას – ვუშვებთ მართობს აბსცისთა ლერძზე), ვინაიდან ჩანაწერის ხანგრძლივობის დაახლოებით ნახევარი იქნება მასზე ნაკლები და დაახლოებით ნახევარი – მასზე მეტი. ჩვენ შემთხვევაში მედიანის შეფასება გამოვა 60.



ამრიგად, იმისათვის, რომ ვიპოვოთ დაჯგუფებული მონაცემების მედიანა, დაგროვილი სიხშირეების გრაფიკის საშუალებით, უნდა მოვძებნოთ დაკვირვებულ მონაცემთა დიაპაზონში ისეთი მნიშვნელობა, რომელიც შეესაბამება იმ დაგროვილ სიხშირეს, რომელიც ერთობლივი სიხშირის ნახევრის ტოლია.

§3. მოდა

ცენტრალური ტენდენციის მესამე საზომს წარმოადგენს **მოდა (mode)**.

მოდა წარმოადგენს იმ მნიშვნელობას შერჩევიდან, რომელიც ყველაზე ხშირად გვხვდება. ზოგჯერ ამბობენ, რომ მოდა არის ყველაზე ტიპური შემთხვევა. მონაცემთა სიმრავლეს შეიძლება ჰქონდეს ერთი მოდა, ერთზე მეტი მოდა ან საერთოდ არ ჰქონდეს მოდა. გავეცნოთ მაგალითებს.

მაგალითი 3.3.1. ქვემოთ მოყვანილია აშშ-ის მრავალჯერადი გამოყენების კოსმოსური ხომალდის ფრენის ხანგრძლივობა (დღეებში) 1992–94 წლებში:

8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11.

ვიპოვოთ მოდა.

ამოხსნა: მოდის მოსაძებნად სასარგებლოა (მოხერხებულია) მონაცემების დალაგება ზრდის მიხედვით, თუმცა ეს არ არის აუცილებელი:

6, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10, 11, 11, 14, 14, 14.

ვინაიდან 8-დღიანი ფრენა მოხდა 5-ჯერ (8-ის სიხშირე უფრო დიდია, ვიდრე ნებისმიერი სხვა რიცხვის), ამიტომ ამ მონაცემების მოდა არის 8.

მაგალითი 3.3.2. იპოვეთ სამხრეთ-დასავლეთ პენსილვანიის 10 შერჩეულ საგრაფოში ქვანახშირის მოპოვებაში დასაქმებულ მუშათა რიცხვის მოდა. ეს მონაცემებია:

110, 731, 1031, 84, 20, 118, 1162, 1977, 103, 752.

ამოხსნა: ვინაიდან ყველა მნიშვნელობა გვხვდება მხოლოდ ერთხელ, აქ მოდა არ არსებობს.

შენიშვნა: არ უნდა ვთქვათ, რომ მოდა არის ნული. ეს იქნება არაკორექტული, ვინაიდან მონაცემთა ზოგიერთი სიმრავლისათვის (მაგალითად, როგორიცაა ტემპერატურა) ნული შეიძლება აღმოჩნდეს ფაქტობრივი (რეალური) მნიშვნელობა.

მაგალითი 3.3.3. შემომდა 11 სხვადასხვა ავტომობილის სამუხრუჭე მანძილის სიდიდე, როცა ისინი მოძრაობდნენ 15 მილი/საათში სიჩქარით. ვიპოვოთ მოდა, თუ მიღებულ იქნა შემდეგი მონაცემები:

15, 18, 18, 18, 20, 22, 24, 24, 24, 24, 26, 26.

ამოხსნა: ვინაიდან 18 და 24, ორივე გვხვდება მაქსიმალური რაოდენობით – 3-ჯერ, ამიტომ მოდა არის 18 და 24. ამ შემთხვევაში ამბობენ, რომ მონაცემთა სიმრავლე არის **ბიმოდალური**.

შეიძლება ისე მოხდეს, რომ შერჩევაში ისეთი მნიშვნელობა, რომელიც ყველაზე ხშირად გვხვდება, არ აღმოჩნდეს. ამ შემთხვევაში შეიძლება ვილაპარაკოთ ე. წ. **მოდალური ინტერვალზე**, რომელიც ასე განიმარტება: თუ აგებული გვაქვს ასე თუ ისე მისაღები ჰისტოგრამა (ანუ ჰისტოგრამა, რომელიც ადეკვატურად აღწერს მონაცემთა სიხშირულ განაწილებას), მაშინ **მოდალური ინტერვალის** ენოდეება იმ ინტერვალს, რომელსაც შეესაბამება ჰისტოგრამის მაქსიმალური მნიშვნელობა, ანუ მოდალური ინტერვალის არის ის ინტერვალის, რომელშიც მოხვდა მონაცემთა ყველაზე დიდი რაოდენობა. გარკვეულობისათვის, მოდას უწოდებენ მოდალური ინტერვალის შუანერტილს. თუ შერჩევის მოცულობა დიდია, მაშინ სწორედ მონაცემთა დაჯგუფებას და მოდალური ინტერვალის მოძებნას ამჯობინებენ.

გასაგებია, რომ მოდის (მოდალური ინტერვალის) მოძებნა საკმაოდ რთულია. იმ ინტერვალის მითითება, სადაც მოხვდა მონაცემთა მაქსიმალური რაოდენობა, დამოკიდებულია იმაზე, თუ როგორ დავაჯგუფებთ მონაცემებს, კერძოდ, იმაზე, თუ რა სიგრძის ინტერვალებად დავყოფთ ინტერვალს (x_{min}, x_{max}). შესაბამისად, მოდალური ინტერვალის დადგენას თან ახლავს ყველა ის სიძნელე, რაც დაკავშირებულია ჰისტოგრამის აგებასთან.

მაგალითი 3.3.4. ვიპოვოთ მოდალური კლასი შემთხვევით შერჩეული 20 მორბენალის მიერ კვირის განმავლობაში გარბენილი მანძილების სიხშირული განაწილების ცხრილის მიხედვით:

კლასის საზღვრები	სიხშირე
5.5—10.5	1
10.5—15.5	2
15.5—20.5	3
20.5—25.5	5
25.5—30.5	4
30.5—35.5	3
35.5—40.5	2
	$n = 20$

ამოხსნა: მოდალური კლასი იქნება კლასი 20.5–25.5, ვინაიდან მას აქვს ყველაზე დიდი სიხშირე. როგორც ზემოთ იყო აღნიშნული, ზოგჯერ მოდას უწოდებენ მოდალური კლასის შუანერტილს. შესაბამისად, მოდად შეიძლება მიჩნეულ იქნეს $(20.5+25.5)/2=23$ მილის გარბენა კვირაში.

მოდა არის ერთადერთი ცენტრალური ტენდენციის საზომებიდან, რომლის გამოყენებაც შეიძლება სახელდებით ან კატეგორიული მონაცემებისათვის, ვინაიდან, ცხადია, რომ კატეგორიზებული მონაცემებისათვის საშუალოსა და მედიანის გამოყენება უაზრობაა, ხოლო მოდა კი იძლევა სასარგებლო ინფორმაციას ყველაზე ტიპური შემთხვევის გამოსავლენად.

მაგალითი 3.3.5. სტუდენტებში ჩატარებულ იქნა გამოკითხვა იმის შესახებ, თუ რა სფეროში აპირებდნენ ისინი მოღვაწეობას სწავლის დასრულების შემდეგ. მიღებულ იქნა სიხშირეთა განაწილების შემდეგი ცხრილი:

მოღვაწეობის სფერო	სიხშირე
ბიზნესი	1425
ჰუმანიტარული მეცნიერებები	878
კომპიუტერული მეცნიერებები	632
განათლება	471
სახელმწიფო სტრუქტურები	95

ამოხსნა: ვინაიდან ყველაზე მაღალი სიხშირის მქონე კატეგორიაა ბიზნესი, ამიტომ ყველაზე ტიპური დასაქმების სფეროა ბიზნესში დასაქმება.

მონაცემთა განლაგების ცენტრის სამივე მახასიათებელი: საშუალო, მედიანა და მოდა წარმოადგენს პოპულაციის შესაბამისი მახასიათებლების: საშუალოს, მედიანისა და მოდის შერჩევით ანალოგებს. ალბათობის თეორიაში მტკიცდება, რომ შერჩევის მოცულობის ზრდისას შერჩევის რიცხვითი მახასიათებლები იკრიბება (უახლოვდება) პოპულაციის ანალოგიურ მახასიათებლებს. აქედან გამომდინარე, შერჩევითი მახასიათებლები შეიძლება ჩაითვალოს პოპულაციის მახასიათებლების შეფასებად და გამოყენებულ იქნეს ამ უკანასკნელის შესახებ სტატისტიკური დასკვნების მისაღებად.

ცხადია, რომ საშუალო, მედიანა და მოდა ხშირად სხვადასხვა ინფორმაციას იძლევა. როცა მოცემულია ცენტრის მახასიათებელი, მნიშვნელოვანია ვიცოდეთ ის საშუალოა, მედიანა თუ მოდა. როცა ვითვლით ცენტრის მახასიათებელს, მნიშვნელოვანია შევარჩიოთ ის მახასიათებელი, რომელიც ყველაზე უფრო ტიპურია მოცემულ შემთხვევაში.

მაგალითი 3.3.6. ქალაქგარეთ მცხოვრებ ადამიანს სამსახურში წასასვლელად შეუძლია ისარგებლოს ორი მარშრუტით: მარშრუტი X და მარშრუტი Y . მან გადანყვიტა შეადაროს მგზავრობაზე დახარჯული დროები თითოეული მარშრუტისათვის. მან დააღაგა 10 მომდევნო სამუშაო დღისათვის მგზავრობაზე დახარჯული დროები (წუთებში) თითოეული მარშრუტისათვის და მიიღო შემდეგი სურათი:

მარშრუტი X: 53 52 48 51 49 47 42 48 57 53

მარშრუტი Y: 43 41 39 108 52 42 38 45 39 51

გამოთვალეთ საშუალო და მედიანა თითოეული მარშრუტისათვის. რომელი მახასიათებელია უფრო მოხერხებული ამ მარშრუტებზე მგზავრობის დროების შესადარებლად?

ამოხსნა: X და Y მარშრუტების საშუალო იქნება, შესაბამისად:

$$\bar{x} = (\sum_{i=1}^{10} x_i) / 10 = 500 / 10 = 50 \quad \text{და} \quad \bar{y} = (\sum_{i=1}^{10} y_i) / 10 = 498 / 10 = 49.8.$$

მედიანების მოსაძებნად დავალაგოთ მონაცემები ზრდადობის მიხედვით:

მარშრუტი X: 42 47 48 48 49 51 52 53 53 57

მარშრუტი Y: 38 39 39 41 42 43 45 51 52 108

ვინაიდან მონაცემთა რაოდენობა 10-ის ტოლია, მედიანა იქნება მე-5 და მე-6 მონაცემების საშუალო არითმეტიკული. ამიტომ X და Y მარშრუტების მედიანა იქნება, შესაბამისად:

$$(49 + 51) / 2 = 50 \quad \text{და} \quad (42 + 43) / 2 = 42.5.$$

როგორც ვხედავთ, X მარშრუტის მედიანა და საშუალო ერთი და იგივეა, ხოლო Y მარშრუტის შემთხვევაში საშუალო მეტია მედიანაზე, რაც გამოწვეულია ერთი უჩვეულოდ დიდი მონაცემით (კერძოდ, 108 წუთის ტოლი მგზავრობის დროით). ეს არაჩვეულებრივი (ამოვარდნილი) მგზავრობის დრო სავარაუდოდ განპირობებულია ცუდი ამინდით ან ავტოსაგზაო შემთხვევით და არაა ტიპური. ამიტომ მგზავრობის დროის შეფასებისას უკეთესია ვისარგებლოთ მედიანით, რამდენადაც ის არ განიცდის ამ ტიპის ამოვარდნების ზემოქმედებას. შესაბამისად, Y მარშრუტი 42.5 წუთის ტოლი მედიანით უფრო სწრაფია, ვიდრე X მარშრუტი 50 წუთის ტოლი მედიანით.

კითხვები

1. 7 სტუდენტის ტესტირების ქულებია:

18 2 5 0 17 15 16

იპოვეთ საშუალო ქულა.

2. ა) იპოვეთ $\bar{x}$, თუ $\sum_{i=1}^{20} x_i = 226$.

ბ) იპოვეთ $\bar{y}$, თუ ცნობილია, რომ $\sum_{i=1}^{12} (y_i - 100) = 66$.

3. 182-გვერდიანი წიგნის თითოეულ გვერდზე შეცდომების რაოდენობა მოცემულია შემდეგი ცხრილის სახით:

შეცდომების					
რაოდენობა	0	1	2	3	4
გვერდების					
რაოდენობა	144	24	10	2	2

იპოვეთ გვერდზე შეცდომების რაოდენობის საშუალო.

4. ქვემოთ მოყვანილია სპორტულ შეჯიბრებაზე პაუზების ხანგრძლივობის (წამებში) სიხშირული განაწილების ცხრილი:

პაუზის								
ხანგრძლივობა	1	2	3	4	5	6	7	8
სიხშირე	2	20	15	12	10	5	3	1

გამოთვალეთ საშუალო.

5. ქვემოთ მოყვანილი ცხრილი გვიჩვენებს თსუ 50 სტუდენტის მიერ წლის განმავლობაში ლექციაზე არასაპატიო მიზეზით დაგვიანებების რაოდენობას:

დაგვიანებების									
რაოდენობა	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
სიხშირე	1	1	1	6	17	16	4	2	2

გამოთვალეთ დაგვიანებათა საშუალო რიცხვი.

6. ავტოსტრადაზე რადარის მიერ დაფიქსირებული 200 სატვირთო მანქანის სიჩქარეების (მილი/სთ-ში) დაჯგუფებული სიხშირეების ქვემოთ მოყვანილი ცხრილის მიხედვით შეაფასეთ სიჩქარეების საშუალო.

სიჩქარის						
ინტერვალი	30-39	40-49	50-59	60-69	70-79	80-89
სიხშირე	12	32	56	72	20	8

7. გამყიდველი ქალის მიერ განხორციელებული 30 სატელეფონო საუბრის ხანგრძლივობა (წუთებში) თავმოყრილია ქვემოთ მოყვანილ ცხრილში:

საუბრის					
ხანგრძლივობა	0-2	3-5	6-8	9-11	12-14
საუბრების					
რაოდენობა	17	6	4	2	1

ა) ამოწერეთ კლასების საზღვრები.

ბ) შეაფასეთ სატელეფონო საუბრების ხანგრძლივობის საშუალო.

8. 8 სპორტსმენის საშუალო წონა შეადგენს 100კგ-ს. მათ შეუერთდა მეცხრე სპორტსმენი, რომელიც იწონის 55კგ-ს. გამოთვალეთ 9 სპორტსმენის საშუალო წონა.
9. როცა 6 მეზღვაურისაგან შემდგარ ეკიპაჟს შეუერთდა მე-7 მეზღვაურ, მათი საშუალო წონა გაიზარდა 110 კგ-დან 111 კგ-მდე. იპოვეთ მე-7 მეზღვაურის წონა.
10. CD დისკების ჩანაწერის ხანგრძლივობის (წუთებში) სიხშირული განაწილების ქვემოთ მოყვანილი ცხრილის მიხედვით შეაფასეთ ჩანაწერის ხანგრძლივობის საშუალო:

ჩანაწერის ხანგრძლივობა, x (წთ)	კლასის საზღვრები	სიხშირე
40-44	$39.5 \leq x < 44.5$	1
45-49	$44.5 \leq x < 49.5$	7
50-54	$49.5 \leq x < 54.5$	12
55-59	$54.5 \leq x < 59.5$	24
60-64	$59.5 \leq x < 64.5$	29
65-69	$64.5 \leq x < 69.5$	14
70-74	$69.5 \leq x < 74.5$	5
75-79	$74.5 \leq x < 79.5$	3
		სულ: 95

11. სხვადასხვა მაღაზიაში შეძენილი 60 CD დისკის ფასების ჯამი აღმოჩნდა

$\sum_{i=1}^{60} x_i = 773.4$. გამოთვალეთ CD დისკების საშუალო ფასი. გარდა ამისა, გამოთვალეს 40 სხვა CD დისკის საშუალო ფასი და ის აღმოჩნდა 11.64. იპოვეთ 100-ივე CD დისკის საშუალო ფასი.

12. იპოვეთ მედიანა ქვემოთ მოყვანილი 20 სტუდენტის მასისათვის (ფუნტებში):

118 147 146 138 175 118 155 146 135 127
136 122 114 140 106 159 127 143 153 139.

13. იპოვეთ შემდეგი მასების მედიანა:

6.6კგ 3.2კგ 4.8 კგ 7.6კგ 5.4კგ 7.1კგ 2.0კგ 6.3კგ 4.3კგ

მასების ამ სიმრავლეს დაუმატეს 6.0კგ. რა იქნება 10 მასის მედიანა?

14. გამოთვალეთ მონაცემთა შემდეგი სიმრავლეების მედიანები.

ა) ქალაქის ქუჩაზე გაზომილი 20 ავტომობილის სიჩქარე კმ/სთ-ებში:

41 15 4 27 21 32 43 37 18 25

29 34 28 30 25 52 12 36 6 25

ბ) ტექნიკური მომსახურების სადგურში 17 ავტომობილის შეკეთებაზე დახარჯული დრო სთ-ებში

0.9 1.0 2.1 4.2 0.7 1.1 0.9 1.8

0.9 1.2 2.3 1.6 2.1 0.3 0.8 2.7 0.4

15. იპოვეთ კლინიკის 45 პაციენტის სისხლში პემოგლობინის დონის მედიანა, თუ გაზომვის შედეგებია:

9.1 10.1 10.7 10.7 10.9 11.3 11.3 11.4 11.4 11.4 11.6 11.8 12.0 12.1 12.3
12.4 12.7 12.9 13.1 13.2 13.4 13.5 13.5 13.6 13.7 13.8 13.8 14.0 14.2 14.2
14.2 14.6 14.6 14.8 14.8 15.0 15.0 15.0 15.1 15.4 15.6 15.7 16.2 16.3 16.9

16. 2006 წლის 360 დღის განმავლობაში მომუშავე სუპერმარკეტის დღიური ნაგაჭრის ჩანაწერები (გაზომილი 10 000 ლარებში) ასე გამოიყურება:
- | | | | | |
|--------------|----------------|----------------|----------------|----------------|
| ნაგაჭრი, x | $x < 2$ | $2 \leq x < 3$ | $3 \leq x < 4$ | $4 \leq x < 5$ |
| დღეთა რიცხვი | 15 | 27 | 64 | 72 |
| ნაგაჭრი, x | $5 \leq x < 6$ | $6 \leq x < 7$ | $7 \leq x < 8$ | $8 \leq x < 9$ |
| დღეთა რიცხვი | 86 | 70 | 16 | 10 |
- შეაფასეთ დღიური ნაგაჭრის მედიანა.
17. ქვემოთ მოყვანილია 80 აალებადი ნივთიერების წამებში გაზომილი აალების დრო (დალაგებული):
- | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| 1.2 | 1.4 | 1.4 | 1.5 | 1.5 | 1.6 | 1.7 | 1.8 | 1.8 | 1.9 | 2.1 | 2.2 | 2.3 | 2.5 | 2.5 | 2.5 |
| 2.5 | 2.6 | 2.7 | 2.8 | 3.1 | 3.2 | 3.5 | 3.6 | 3.7 | 3.8 | 3.8 | 3.9 | 3.9 | 4.0 | 4.1 | 4.2 |
| 4.3 | 4.5 | 4.5 | 4.6 | 4.7 | 4.7 | 4.8 | 4.9 | 5.1 | 5.1 | 5.1 | 5.2 | 5.2 | 5.3 | 5.4 | 5.5 |
| 5.6 | 5.8 | 5.9 | 5.9 | 6.0 | 6.3 | 6.4 | 6.4 | 6.4 | 6.4 | 6.7 | 6.8 | 6.8 | 6.9 | 7.3 | 7.4 |
| 7.4 | 7.6 | 7.9 | 8.0 | 8.6 | 8.8 | 8.8 | 9.2 | 9.4 | 9.6 | 9.7 | 9.8 | 10.6 | 11.2 | 11.8 | 12.4 |
- ა) დააჯგუფეთ მონაცემები 8 თანაბარი სიგრძის კლასად (დაიწყეთ 1.0–2.4-დან) და შეადგინეთ დაჯგუფებული სიხშირების ცხრილი. ააგეთ სიხშირეთა პისტოგრამა. რას გვიჩვენებს პისტოგრამა?
- ბ) დაჯგუფებული სიხშირების ცხრილის მიხედვით ააგეთ დაგროვილი სიხშირების დიაგრამა და შეაფასეთ მედიანა.
- გ) პირდაპირი გზით მოძებნეთ მონაცემების ზუსტი მედიანა და იპოვეთ განსხვავება მასა და მედიანის შეფასებას შორის.
18. ქვემოთ მოყვანილია CD დისკების მწარმოებელ კომპანიაში 100 დღის განმავლობაში დაკვირვებული დეფექტური დისკების დაჯგუფებული სიხშირული განაწილების ცხრილი:
- | | | | | | | |
|------------------------------|-----|-------|-------|-------|-------|-------|
| დეფექტური დისკების რაოდენობა | 0–9 | 10–19 | 20–29 | 30–39 | 40–49 | 50–59 |
| დღეების რიცხვი | 5 | 8 | 19 | 37 | 22 | 9 |
- შეაფასეთ დეფექტური დისკების მედიანა.
19. მიუთითეთ რომელ განაწილებას აქვს მოდა ან მოდალური კლასი:
- ა) x_i 0 1 2 3 4 ბ) x_i 70 75 80 85 90
- | | | | | | |
|-------|---|---|---|---|---|
| f_i | 7 | 4 | 2 | 5 | 1 |
| f_i | 7 | 7 | 7 | 7 | 7 |
- გ) x_i 2–3 4–5 6–7 8–9 10–11
- | | | | | | |
|-------|---|---|---|---|---|
| f_i | 7 | 4 | 2 | 5 | 1 |
|-------|---|---|---|---|---|
- დ) თვალის ფერი ცისფერი თაფლისფერი მწვანე
- | | | | |
|-------|----|----|---|
| f_i | 23 | 39 | 3 |
|-------|----|----|---|
20. ცენტრალური ტენდენციის რომელი მაჩვენებელით სარგებლობა იქნება უფრო მიზანშეწონილი ქვემოთ მოყვანილ სიტუაციებში:
- ა) ფეხსაცმლის მაღაზიის მენეჯერს სურს შეუკვეთოს სხვადასხვა ზომის ფეხსაცმელი.

- ბ) ქალაქის საკრებულოს სურს დაგეგმოს ახალი სკოლის მშენებლობა. ამიტომ მოსწავლეთა რაოდენობის შეფასებისთვის იკვლევს ერთნაირი მოცულობისა და მდგომარეობის ოჯახებს.
- გ) ივანე რეგულარულად მგზავრობს თბილისიდან გორში და აღრიცხავს მგზავრობის დროს. მას სურს შეაფასოს მომავალი მგზავრობის დრო.
21. უძრავი ქონების აგენტმა გააკეთა შემდეგი განცხადება: მიმდინარე თვეში სახლების დაახლოებით 60% გაიყიდა უფრო ძვირად, ვიდრე საშუალო გასაყიდი ფასია. განიხილეთ ამ მტკიცებულების შესაძლო მართებულობა და რა იგულისხმება „საშუალოში“.
22. მიუთითეთ ქვემოთ ჩამოთვლილთაგან რომელ სიდიდეს ექნება ასიმეტრიული ან, დაახლოებით, სიმეტრიული განაწილება:
- მდედრობითი სქესის უნივერსიტეტის სტუდენტების სიმაღლე.
 - მარათონულ შეჯიბრებაში მონაწილეთა სირბილის დრო.
 - ადვილ გამოცდაზე სტუდენტების მიერ მიღებული ქულები.
 - ბიბლიოთეკის წიგნების გვერდების რაოდენობა.
23. ქვემოთ მოცემულია 32 მოსწავლის მიერ 8-კითხვიან მათემატიკურ ტესტში მიღებული ქულები:
- | | | | | | | | | | |
|-----------------|---|---|---|---|---|---|---|---|---|
| სწორი პასუხების | | | | | | | | | |
| რაოდენობა | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| მოსწავლეთა | | | | | | | | | |
| რაოდენობა | 1 | 2 | 1 | 4 | 4 | 6 | 7 | 4 | 3 |
- იპოვეთ სწორი პასუხების რაოდენობის საშუალო, მედიანა და მოდა. გააკეთეთ მედიანისა და მოდის ინტერპრეტაცია მათემატიკური ტესტის კონტექსტში.
 - დაახასიათეთ განაწილების მოხაზულობა.

თავი 4

მონაცემთა გაფანტულობის საზომები

§1. გაბნევის დიაპაზონი. პროცენტული

სტატისტიკაში, იმისათვის, რომ ზუსტად (ადეკვატურად) აღვწეროთ მონაცემთა სიმრავლე, სტატისტიკოსმა უნდა იცოდეს უფრო მეტი, ვიდრე ცენტრალური ტენდენციის საზომებია. ჩვენ შევისწავლეთ მონაცემთა ცენტრალური ტენდენციის (ანუ მონაცემთა ცენტრის მდებარეობის) სამი რიცხვითი მაჩასიათებელი: შერჩევითი საშუალო, შერჩევითი მედიანა და შერჩევითი მოდა. გასაგებია, რომ ეს სამი მაჩასიათებელი სრულად ვერ წარმოადგენს მონაცემთა სიმრავლის თვისებებს. მეორე ეტაპზე ბუნებრივად იბადება კითხვა: – რამდენად ტიპურია მონაცემთა ცენტრის მაჩასიათებელი, კერძოდ, საშუალო, მონაცემთა მთელი სიმრავლისათვის? როგორია მონაცემთა გაფანტულობა საერთოდ და კონკრეტულად კი საშუალოს მიმართ? როგორია მონაცემთა ცვალებადობის ხარისხი და რა სიდიდებით შეიძლება მისი დახასიათება?

მაგალითი 4.1.1. დავუშვათ, მოცემულია მონაცემთა ორი მწკრივი:

8	9	10	10	13
1	5	10	16	18.

გამოვთვალოთ თითოეულის საშუალო. გვაქვს:

$$\bar{x}_A = (8 + 9 + 10 + 10 + 13) / 5 = 10 \quad \text{და} \quad \bar{x}_B = (1 + 5 + 10 + 16 + 18) / 5 = 10.$$

აღმოჩნდა, რომ ორივე მონაცემს აქვს ერთი და იგივე საშუალო, მაგრამ ცხადია, რომ B მწკრივის ელემენტთა ცვალებადობა უფრო ძლიერია, ვიდრე A მწკრივის ელემენტების. A მწკრივის ელემენტები უფრო მჭიდროდ არიან თავმოყრილი (კონცენტრირებული) საშუალოს ირგვლივ, ვიდრე B მწკრივის ელემენტები.

ხშირ შემთხვევაში, თვალთ შემჩნევა იმისა, თუ რომელი მწკრივის ცვალებადობა უფრო დიდია, საკმაოდ ძნელია (მაგალითად, მონაცემთა დიდი რაოდენობის დროს). ამიტომ აუცილებელია ცვალებადობის ანუ გაფანტულობის საზომების შემოღება, ე. ი. ისეთი რიცხვითი მაჩასიათებლების შემოღება, რომელიც საშუალებას მოგვცემს შევაფასოთ მონაცემთა გაბნევის ხარისხი.

მონაცემთა გაფანტულობის ერთ-ერთი უმარტივესი რიცხვითი მაჩასიათებელია **გაბნევის დიაპაზონი (range)**, რომელიც წარმოადგენს შერჩევის მაქსიმალური (უდიდესი) და მინიმალური (უმცირესი) მნიშვნელობების სხვაობას. სხვა სიტყვებით – გაბნევის დიაპაზონი წარმოადგენს ვარიაციული მწკრივის უკანასკნელი და პირველი წევრების სხვაობას, თუმცა გაბნევის დიაპაზონის დასადგენად არ არის აუცილებელი (უბრალოდ, მოხერხებულია) ვარიაციული მწკრივის შექმნა

გამოყენებითი სტატისტიკის საფუძვლები

(მონაცემების დალაგება ზრდადობის მიხედვით). გაბნევის დიაპაზონი აღინიშნება R ასოთი. შესაბამისად, თუ $x_1, \dots, x_n$ მოცემული n მოცულობის მქონე შერჩევაა (ხოლო $x_{(1)}, \dots, x_{(n)}$ კი – შესაბამისი ვარიაციული მწკრივი), მაშინ:

$$R = \max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i = x_{(n)} - x_{(1)},$$

სადაც $x_n = \max_{1 \leq i \leq n} x_i$, $(x_{(1)} = \min_{1 \leq i \leq n} x_i)$ – მაქსიმალური (შესაბამისად, მინიმალური) სიდიდის მონაცემია.

ზემოთ მოყვანილ მაგალითში მონაცემთა A მწკრივის გაბნევის დიაპაზონი $R_A = 13 - 5 = 8$, ხოლო B მწკრივის – $R_B = 18 - 1 = 17$.

მონაცემთა გაბნევის დიაპაზონი, გამოთვლისა და ინტერპრეტაციის სიმარტივის გამო ფართოდ გამოიყენება ისეთი ამოცანების გადაწყვეტისას, სადაც გამოთვლის სისწრაფსა და მარტივ შინაარსს, გადამწყვეტი მნიშვნელობა ენიჭება, მაგალითად, ხარისხის კონტროლის სფეროში. სხვა შემთხვევებში კი ამ მახასიათებლის მიმართ ნარმოიშობა სერიოზული პრობლემები. გაბნევის დიაპაზონზე დიდ გავლენას ახდენს იშვიათი ან ე. წ. ამოვარდნილი დაკვირვებები. მაგალითად, დაუშვათ, რომ ჩვენ რომელიმე რეგიონში ვანარმოებთ შერჩევას მამაკაცთა სიმაღლის განაწილების დადგენის მიზნით. ვთქვათ, ამ რეგიონში ცხოვრობს მამაკაცი, რომლის სიმაღლეა 210 სმ. ცხადია, რომ, თუ შერჩევაში შემთხვევით აღმოჩნდება ეს მამაკაცი (რაც იშვიათი ხდომილებაა), გაბნევის დიაპაზონი უფრო დიდი იქნება, ვიდრე იმ შემთხვევაში, როცა მის ნაცვლად შერჩევაში მოხვდება სხვა ტიპური სიმაღლის მქონე მამაკაცი. ანალოგიური სურათი შეიძლება გვქონდეს მოსახლეობის შემოსავლების განაწილების შესწავლისას, როდესაც განსაკუთრებით დიდი შემოსავლების მქონე ადამიანი აღმოჩნდება (იშვიათი ხდომილება) ჩვენს შერჩევაში.

შემდეგი მაგალითი გვიჩვენებს, რომ ერთ ექსტრემალურად მაღალი (სხვა შემთხვევაში შეიძლება – ექსტრემალურად დაბალი) მონაცემის სიდიდეს შეუძლია მნიშვნელოვანი გავლენა მოახდინოს გაბნევის დიაპაზონზე.

მაგალითი 4.1.2. ქვემოთ მოყვანილია ერთ-ერთი წარმოების მუშაკთა წლიური შემოსავალი დოლარებში.

თანამდებობა	შემოსავალი
მეპატრონე	100 000
მენეჯერი	40 000
გაყიდვების აგენტი	30 000
I თანრიგის მუშა	25 000
II თანრიგის მუშა	18 000
III თანრიგის მუშა	15 000

იპოვეთ გაბნევის დიაპაზონი.

ამოხსნა: გაბნევის დიაპაზონი $R = 100000 - 15000 = 85000$.

შერჩევის გაბნევის დიაპაზონის, როგორც მონაცემთა გაფანტულობის საზომის, მთავარი ნაკლი ისაა, რომ ის არ შეიცავს ინფორმაციას იმის შესახებ, თუ როგორაა გაბნეული დანარჩენი (საშუალოდ) მონაცემები მაქსიმალურ და მინიმალურ მნიშვნელობებს შორის. შესაძლოა ეს მონაცემები გროვდებოდნენ საშუალო მნიშვნელობასთან, ან ექსტრემალური მნიშვნელობების მახლობლად, ან თანაბრად იყვნენ განაწილებული მათ შორის და სხვა. საშუალოდ მონაცემების განლაგებისა და, შესაბამისად, მათი გაფანტულობის შესახებ ინფორმაციას იძლევა ე. წ. პროცენტილები (percentiles).

პროცენტილები ერთდროულად წარმოადგენს მონაცემთა განლაგებისა და მათი გაფანტულობის საზომს. ვთქვათ, P რაიმე რიცხვია, მოთავსებული 0-სა და 100-ს შორის $0 < P < 100$. მონაცემთა სიმრავლის P რიგის პროცენტილი (უბრალოდ P -პროცენტილი) ეწოდება ისეთ $\bar{x}_P$ მნიშვნელობას (სიდიდეს), რომელსაც გააჩნია შემდეგი თვისება: მონაცემთა არაუმეტეს $P\%$ -ისა ნაკლებია $\bar{x}_P$ -ზე და არაუმეტეს $(100 - P)\%$ -ისა მეტია $\bar{x}_P$ -ზე.

მიღებულია P რიცხვის წარმოდგენა $P = 100 \cdot \alpha$ სახით, სადაც $0 < \alpha < 1$. მაგალითად, თუ $P = 10$ -ს, მაშინ $\alpha = 0.1$, თუ $P = 25$ -ს, მაშინ $\alpha = 0.25$. ამ შემთხვევაში P -პროცენტულის განსაზღვრება ასე გამოითქმის: $\bar{x}_P$ ისეთი რიცხვია, რომ ერთდროულად უნდა სრულდებოდეს შემდეგი ორი უტოლობა:

$$N\{x_i : x_i < \bar{x}_P\} \leq \alpha n, \quad N\{x_i : x_i > \bar{x}_P\} \leq (1 - \alpha)n,$$

სადაც $N\{x_i : \dots\}$ სიმბოლოთი აღნიშნულია იმ დაკვირვებების რაოდენობა, რომელთათვისაც სრულდება ორი წერტილის შემდეგ მოთავსებული თვისება. $P = 100 \cdot \alpha$ პროცენტილი არის α -კვანტილის (F განაწილების α -კვანტილი x_α განიმარტება შემდეგნაირად: $x_\alpha = \min\{x : F(x) \geq \alpha\}$) შერჩევითი ანალოგი.

სიმარტივისათვის ქვემოთ ვიგულისხმებთ, რომ ვარაიციული მწკრივი მკაცრად ზრდადი მიმდევრობაა: $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

შენიშვნა. თუ ვარაიციულ მწკრივში მონაცემები მეორდება ნაცვლად რაოდენობისა, $N\{x_i : x_i < \bar{x}_P\}$ (შესაბამისად, $N\{x_i : x_i > \bar{x}_P\}$) უნდა დავითვალოთ რაოდენობა იმ პოზიციების, რომელიც მოთავსებულია $\bar{x}_P$ -ის პოზიციის მარცხნივ (შესაბამისად, მარჯვნივ). ამ შეთანხმების გათვალისწინებით, პროცენტულის ცნება ადვილად გავრცელდება დაჯგუფებული მონაცემების შემთხვევაში.

ადვილი მისახვედრია, რომ მედიანა წარმოადგენს 50 პროცენტულს.

თუ $P = 100 \cdot \alpha$, მაშინ მედიანის ანალოგიურად P -პროცენტულის ($\bar{x}_P$ -ს) გამოსათვლელი გამოსახულება იქნება:

$$\bar{x}_p = x_{([n\alpha]+1)}, \text{ როცა } n\alpha \text{ არ არის მთელი რიცხვი და}$$

$$\bar{x}_p = \frac{x_{(n\alpha)} + x_{(n\alpha+1)}}{2}, \text{ როცა } n\alpha \text{ მთელი რიცხვია}$$

(უკანასკნელ შემთხვევაში, ისევე როგორც მედიანის დროს, P -პროცენტილი ცალსახად არ განიზარტება. მის როლში შესაძლებელია ავილოთ ნებისმიერი რიცხვი $(x_{(n\alpha)}, x_{(n\alpha+1)})$ ინტერვალადან. ამიტომ თანხმდებიან, რომ P -პროცენტილი ვუწოდოთ ამ ინტერვალის შუანერტილს).

მაგალითი 4.1.3. თუ მოცემულია $n=15$ მოცულობის მქონე შერჩევა, და $P=10$, მაშინ $\alpha=0.1$. შესაბამისად, $n\alpha=1.5$ და ვინაიდან $n\alpha$ არ არის მთელი რიცხვი, ხოლო მისი მთელი ნაწილია $[1.5] = 1$, ამიტომ ზემოთ მოყვანილი ფორმულის თანახმად: $\bar{x}_{10} = x_{(2)}$, ანუ 10-პროცენტილი ვარაიციული მწკრივის მეორე ნევრის ტოლია. მართლაც,

$$N\{x_i : x_i < x_{(2)}\} = 1 < n\alpha = 1.5 \quad \text{და} \quad N\{x_i : x_i > x_{(2)}\} = 13 < n(1-\alpha) = 13.5.$$

მაგალითი 4.1.4. თუკი $n=20$, $P=10$, მაშინ $\alpha=0.1$, $n\alpha=2$ – მთელი რიცხვია და ისევე ზემოთ მოყვანილი ფორმულის თანახმად:

$$\bar{x}_{10} = \frac{x_{(2)} + x_{(3)}}{2},$$

ანუ 10-პროცენტილი არის $[x_{(2)}, x_{(3)}]$ ინტერვალის შუა წერტილი. ამ შემთხვევაშიც:

$$N\{x_i : x_i < \bar{x}_{10}\} = 2 = n\alpha \quad \text{და} \quad N\{x_i : x_i > \bar{x}_{10}\} = 18 = n(1-\alpha).$$

მედიანა არის 50-პროცენტილი.

10-ის ჯერად პროცენტელებს დეცილები (decile) ეწოდება. პირველი დეცილია – 10-პროცენტილი, მეორე დეცილია – 20-პროცენტილი და ა.შ მე-9 დეცილია – 90-პროცენტილი.

25-პროცენტის პირველი კვარტილი (quartile) ეწოდება. პირველი კვარტილი Q_1 სიმბოლოთი აღინიშნება.

75-პროცენტის მესამე კვარტილი ეწოდება. მესამე კვარტილი Q_3 სიმბოლოთი აღინიშნება.

მეორე კვარტილი არის მედიანა – Q_2 .

კვარტილები მედიანასთან ერთად, ე. ი. სამეული ($Q_1, \bar{x}, Q_2$) გვაძლევს გარკვეულ წარმოდგენას მონაცემთა ცენტრის, გაბნევისა და განაწილების (გაგლუვებული ჰისტოგრამისა და პოლიგონის) ფორმის შესახებ. იმ შემთხვევაში, როცა

$Q_3 - \bar{x} > \bar{x} - Q_1$ (მესამე კვარტილი უფრო შორსაა მედიანისაგან, ვიდრე პირველი კვარტილი), განანილება მარჯვნივ ასიმეტრიულია (მისი მარჯვენა ბოლო უფრო ასიმეტრიულია, ვიდრე მარცხენა).

მანძილი კვარტილებს შორის წარმოადგენს გაფანტულობის მარტივ საზომს, რომელიც ახასიათებს ვარიაციულ მწკრივში კვარტილებს შორის მოთავსებულ მონაცემთა გაფანტულობის ხარისხს (შუა 50%-იანი ზონის გაბნევას). მანძილს მესამე კვარტილსა და პირველ კვარტილს შორის **კვარტილთშორისი გაბნევის დიაპაზონი (Interquartial range)** ეწოდება და აღინიშნება IQR სიმბოლოთი, $IQR = Q_3 - Q_1$.

პირველი და მესამე კვარტილები და კვარტილთშორისი გაბნევის დიაპაზონი შერჩევის მდგრადი მახასიათებლებია. არცერთი მათგანი არ რეაგირებს $[Q_1, Q_3]$ ინტერვალის გარეთ მოთავსებულ ნებისმიერ ისეთ ცვლილებაზე, რომელიც ამ დაკვირვებებს კვლავ $[Q_1, Q_3]$ ინტერვალის გარეთ ტოვებს, თანაც ინტერვალის იმავე მხარეს, სადაც ადრე იმყოფებოდა.

ვინაიდან მედიანა და პირველი და მესამე კვარტილები არ შეიცავენ ინფორმაციას განანილების (გაგლუვებული ჰისტოგრამის) კუდების შესახებ, ამიტომ განანილების ფორმის შესახებ უფრო სრულ ინფორმაციას იძლევა შემდეგი ხუთეული: $x_{\min}$, Q_1 , $\bar{x}$, Q_3 , $x_{\max}$.

§2. რანგები და პროცენტული რანგები

დავუშვათ, რომ მოცემულია n მოცულობის განსხვავებული მნიშვნელობების მქონე შერჩევა $x_1, \dots, x_n$, სადაც $x_i \neq x_j$, თუ $i \neq j$, ე. ი. არცერთი მნიშვნელობა არ მეორდება. შევადგინოთ ვარიაციული მწკრივი (დავალაგოთ მონაცემები ზრდის მიხედვით):

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}.$$

x_i დაკვირვების **რანგი (rank)** ეწოდება ამ დაკვირვების ნომერს ვარიაციულ მწკრივში (ის წარმოადგენს მონაცემთა პოზიციის ერთ-ერთ მახასიათებელს). სხვა სიტყვებით რომ ვთქვათ, x_i -ს რანგი ეწოდება ისეთ მთელ რიცხვს r_i , რომლისთვისაც სრულდება ტოლობა:

$$x_{(r_i)} = x_i.$$

სინამდვილეში x_i დაკვირვების რანგი r_i არის იმ დაკვირვებების რაოდენობა, რომლებიც არ აღემატებიან x_i დაკვირვებას:

$$r_i = N\{x_j : x_j \leq x_i\},$$

ამასთანავე გასაგებია, რომ მაქსიმალური დაკვირვების რანგი დამოთხვევაა შერჩევის მოცულობას.

მაგალითი 4.2.1. განვიხილოთ შემდეგი მონაცემები: 5, 6, 2, 7, 9, 8, 10 და თითოეულ მათგანს მივუწეროთ თავისი რანგი.

ამოხსნა : ვარიაციული მწკრივი არის: 2, 5, 6, 7, 8, 9, 10. ამიტომ $x_1 = 5$ -ის რანგი იქნება $r_1 = 2$ (დაკვირვებას 5 არ აღემატება ორი დაკვირვება: 2 და 5); $x_2 = 6$ -ის რანგი $r_2 = 3$ (დაკვირვებას 6 არ აღემატება სამი დაკვირვება: 2, 5 და 6) და ა. შ. საბოლოოდ, მონაცემთა მოცემული სიმრავლე ქვეშ მიწერილი რანგებით ასე გამოიყურება:

5	6	2	7	9	8	10
↓	↓	↓	↓	↓	↓	↓
2	3	1	4	6	5	7

მაგალითი 4.2.2. მონაცემებია: 18, 13, 15, 12, 16, 20. მივუწეროთ რანგები. გვაქვს:

18	13	15	12	16	20
5	2	3	1	4	6

ალსანიშნავია, რომ რანგები ინვარიანტულია (არ იცვლება) მონაცემთა ნებისმიერი გადალაგების მიმართ. თუ ამ მაგალითის მონაცემებს დავალაგებთ სხვა თანმიმდევრობით, მაგალითად, შემდეგნაირად: 15, 20, 12, 18, 13, მაშინ რანგები იქნება:

15	20	12	18	13	16
3	6	1	4	2	5

ანუ ყოველ წევრს იგივე რანგი მიენერა, რაც ადრე.

ბუნებრივად იზადება კითხვა: თუ მოცემულია შერჩევა $x_1, \dots, x_n$, ისეთი, რომ პირობა $x_i \neq x_j$, თუ $i \neq j$ არ სრულდება, ანუ ზოგიერთი წევრი შერჩევაში მეორდება, მაშინ როგორი წესით მოხდება რანგების მიწერა?

ვთქვათ, ჩვენი მონაცემებია: 18, 28, 23, 29, 32, 18, 21, 14, 18, 14. რანგის განმარტების ძველი წესის თანახმად, ვარიაციული მწკრივის სახით გადაწერილ ამ მონაცემებს რანგები ასე უნდა მიენეროს:

14	14	18	18	18	21	23	28	29	32
1	2	3	4	5	6	7	8	9	10

როგორც ვხედავთ, 14-ის ტოლი დაკვირვება შეგვხვდა ორჯერ და ძველი წესით პირველს მივანერეთ რანგი 1, ხოლო მეორეს – რანგი 2 (ანალოგიურად, 18 შეგვხვდა სამჯერ და, შესაბამისად, მას დაკავებული პოზიციის მიხედვით მიენერა რანგები: 3, 4 და 5). გამოდის, რომ ერთ-ერთ მათგანს მივანიჭეთ უპირატესობა. უმჯობესია განმეორებადი დაკვირვებებისათვის მიგვეწერა ერთი და იგივე რანგი შემდეგი წესით: ტოლი დაკვირვებებიდან თითოეულს მივანიჭოთ მათზე მოსული რანგების საშუალო არითმეტიკული. მაგალითად, დაკვირვებებს, რომელთა მნი-

შენელობებია 14, უნდა მიენიჭოს რანგი 1.5, რადგანაც $(1+2)/2=1.5$, 18-ის ტოლ დაკვირვებებს – რანგი 4, ვინაიდან $(3+4+5)/3=4$).

მეორე მხრივ, იმ შემთხვევაშიც კი, როდესაც შერჩევაში არცერთი ელემენტი არ მეორდება, ელემენტის რანგის მითითება არ იძლევა ამომწურავ ინფორმაციას მის მდებარეობაზე, თუ ამავე დროს ჩვენთვის ცნობილი არ არის შერჩევის მოცულობაც. მაგალითად, თუ ვიცით, რომ x^* ელემენტს n მოცულობის შერჩევიდან და y^* ელემენტს მეორე, m მოცულობის შერჩევიდან გააჩნიათ ერთი და იგივე რანგი, ჩვენ ჯერ კიდევ ვერ დავასკენით, რომ მათი პოზიციები იდენტურია, თუ არ გვექნება დამატებითი ინფორმაცია იმის შესახებ, რომ $m = n$.

საზოგადოდ, როგორ შევადაროთ ერთმანეთს იმ ორი მონაცემის პოზიციები, რომლებიც განსხვავებული მოცულობის სხვადასხვა შერჩევას ეკუთვნის და რანგებიც განსხვავებულები აქვს (ამ ტიპის კითხვები წარმოიშობა, მაგალითად, ეროვნული გამოცდების დროს, როცა შეფასების ცენტრმა უნდა შეადაროს სხვადასხვა აბიტურიენტის პოზიციები (რეიტინგები) სხვადასხვა დისციპლინაში ან მაშინაც კი, როცა ერთი დისციპლინის ფარგლებში მათ შეასრულეს სხვადასხვა საგამოცდო ვარიანტი)? ამ კითხვაზე პასუხს იძლევა ე. წ. **პროცენტული რანგის (percentile rank) ცნება**.

თუ n შერჩევის მოცულობაა, ხოლო დაკვირვების რანგია r , მაშინ ამ დაკვირვების პროცენტული რანგი მოიცემა შემდეგი ფორმულით:

$$\frac{r-1}{n} \cdot 100\% + \frac{0.5}{n} \cdot 100\% = \frac{2r-1}{2n} \cdot 100\%.$$

მოცემული დაკვირვების პროცენტული რანგი წარმოადგენს ამ დაკვირვების ქვევით მდგომ მონაცემებზე მოსული პროცენტებისა $(\frac{r-1}{n} \cdot 100\%)$ და თავად დაკვირვებაზე მოსული პროცენტების ნახევრის $(\frac{0.5}{n} \cdot 100\%)$ ჯამს.

ამრიგად, მოცემული დაკვირვების პროცენტული რანგი მიგვითითებს დაკვირვებულ მნიშვნელობათა რა პროცენტს (მინუს ამ დაკვირვებაზე მოსული პროცენტების ნახევარი) აღემატება ეს მონაცემი. ცხადია, რომ რაც უფრო მაღალია დაკვირვების პროცენტული რანგი, მით უკეთესია მისი პოზიცია (თუ ნიშნები განლაგებულია აღმავლობის მიხედვით).

ფაქტობრივად, პროცენტული რანგი გვიჩვენებს, ამა თუ იმ შერჩევის კონკრეტული მონაცემი, მონაცემების რამდენი პროცენტის წინაა. შესაბამისად, თუ ვიგულისხმებთ, რომ მონაცემებს სხვადასხვა შერჩევაში მოხვედრის თანაბარი შანსები აქვთ, მაშინ ბუნებრივი იქნება ერთ შერჩევაში უკეთეს პოზიციაზე მდგომ მონაცემს მიენიჭოს უპირატესობა სხვა შერჩევაში უარეს პოზიციაზე მდგომ მონაცემთან შედარებით. მაგალითად, ეროვნულ გამოცდებზე ზოგადი უნარების გამოცდის პირველ ვარიანტში 37 პროცენტული რანგის მქონე აბიტურიენტს უპირატესობა ენიჭება მესამე ვარიანტში 36 პროცენტული რანგის მქონე აბიტურიენტთან შედარებით, ან ინგლისურ ენაში 43 პროცენტული რანგის მქონე

აბიტურიენტს უპირატესობა ენიჭება გერმანულ ენაში 42.9 პროცენტული რანგის მქონე აბიტურიენტთან შედარებით.

მაგალითი 4.2.3. ვთქვათ, კლასში არის 10 მოსწავლე და გიორგის ნიშნები უფრო მაღალია, ვიდრე ხუთი მოსწავლის. კოტე სწავლობს 50 მოსწავლისაგან შემდგარ კლასში და მისი ნიშნები უფრო მაღალია, ვიდრე 17 მოსწავლის. რომელს უკავია უკეთესი პოზიცია თავის კლასში –, კოტეს თუ გიორგის?

ამოხსნა: უპირველეს ყოვლისა, შევნიშნოთ, რომ მოსწავლეები დალაგებულია ნიშნების ზრდის მიხედვით და გიორგის რანგია 6, ხოლო კოტესი 18. ჩავატაროთ თითოეულის პოზიციის პროცენტული ანალიზი. გიორგი უკეთესია 5 მოსწავლეზე, რომლებიც შეადგენენ კლასის 50%-ს, თვითონ გიორგიზე მოდის 10%, ხოლო 40%-ს აქვს გიორგიზე უკეთესი ნიშნები. ამ მონაცემებით გიორგის პროცენტული რანგია:

$$50\% + 5\% = 55\%.$$

კოტეს ნიშნები უკეთესია 17 მოსწავლის ნიშნებზე, რომელიც შეადგენს კლასის 34%-ს, კოტეზე მოდის 2%, ამიტომ მისი პროცენტული რანგია:

$$34\% + 1\% = 35\%.$$

ვინაიდან, გიორგის პროცენტული რანგი (55) უფრო მაღალია, ვიდრე კოტესი (35), ამიტომ გიორგის პოზიცია თავის კლასში უკეთესია, ვიდრე კოტესი საკუთარ კლასში.

ეს მაგალითი, ამავე დროს, გვიჩვენებს, რომ რანგის ცნების შემოსაღებად არ არის აუცილებელი გაგვაჩნდეს რიცხვითი მონაცემები, საკმარისია მხოლოდ ობიექტების ან ინდივიდების სიმრავლის რაიმე თვისების ან ნიშნის (მაგალითად, ნიჭიერება, სიმაღლე, პოპულარობა და ა. შ.) გაძლიერების ან ზრდის მიხედვით დალაგება და ყოველი მათგანისათვის რიგობრივი ნომრის მინერა.

პროცენტილებსა და პროცენტულ რანგებს შორის შემდეგი კავშირია: თუ შერჩევის ელემენტის პროცენტული რანგია P , მაშინ ეს ელემენტი წარმოადგენს შერჩევის P -პროცენტულს (თუმცა არ არის სავალდებულო, რომ P -ს ნებისმიერი მნიშვნელობისათვის P -პროცენტილი წარმოადგენდეს შერჩევის ელემენტს).

§3. დისპერსია და სტანდარტული გადახრა

ისევ დავუბრუნდეთ ადრე განხილულ მაგალითს. დავუშვათ, რომ მოცემულია მონაცემთა ორი მწკრივი:

8 9 10 10 13

1 5 10 16 18

როგორც ჩვენ ვნახეთ, $\bar{x}_A = \bar{x}_B = 10$. გარდა ამისა, დავასკვნით, რომ B მწკრივის ელემენტების ცვალებადობა უფრო მაღალია, ვიდრე A მწკრივის ელემენტების ცვალებადობა.

შენიშვნა. აღსანიშნავია, რომ შერჩევითი საშუალო წარმოადგენს პოპულაციის საშუალოს საუკეთესო შეფასებას (კერძოდ, ის წარმოადგენს ე.წ. ჩაუნაცვლებელ შეფასებას), მაშინ, როდესაც შერჩევითი დისპერსია არ წარმოადგენს პოპულაციის დისპერსიის საუკეთესო შეფასებას, ის ე.წ. ჩანაცვლებული შეფასებაა. იმისათვის, რომ მიიღონ ჩაუნაცვლებელი შეფასება, n -ს ცვლიან $n-1$ -ით (ჩაუნაცვლებლობა გულისხმობს შემდეგს: მიუხედავად იმისა, რომ სხვადასხვა შერჩევის მიხედვით გამოთვლილი პარამეტრის შეფასება საზოგადოდ სხვადასხვა შედეგს იძლევა, ის საშუალოდ ემთხვევა პარამეტრის ჭეშმარიტ მნიშვნელობას).

მოვიყვანოთ საილუსტრაციო მაგალითი იმისა, რომ შერჩევითი საშუალო ჩაუნაცვლებელი შეფასებაა, ხოლო შერჩევითი დისპერსია კი – ჩანაცვლებული.

დავუშვათ, რომ პოპულაცია შეიცავს 3 მონაცემს და მისგან ვაკეთებთ ორმონაცემიან ($n=2$ მოცულობიან) შერჩევებს. პოპულაცია იყოს $\{0, 2, 4\}$.

ცხადია, რომ პოპულაციის საშუალო იქნება $\mu = \frac{0+2+4}{3} = 2$, ხოლო დისპერსია კი

$$\sigma^2 = \frac{1}{3}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2] = \frac{1}{3}[(0-2)^2 + (2-2)^2 + (4-2)^2] = \frac{8}{3}.$$

ყველა შესაძლო ორმოცულობიანი შერჩევების რაოდენობა იქნება $3 \times 3 = 9$ (შერჩევის როგორც პირველი, ისე მეორე ელემენტის შესაძლო მნიშვნელობათა რაოდენობა 3-ის ტოლია). ამოვწეროთ ეს შერჩევები და გამოვთვალოთ მათი საშუალოები და დისპერსიები. ვისარგებლოთ ქვემოთმოყვანილი ცხრილით:

შერჩევა $\{x_1, x_2\}$	საშუალო $\bar{x} = \frac{x_1 + x_2}{2}$	დისპერსია $s_x^2 = \frac{1}{2}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2]$
$\{0, 0\}$	0	$\frac{1}{2}[(0-0)^2 + (0-0)^2] = 0$
$\{0, 2\}$	1	$\frac{1}{2}[(0-1)^2 + (2-1)^2] = 1$
$\{0, 4\}$	2	$\frac{1}{2}[(0-2)^2 + (4-2)^2] = 4$
$\{2, 0\}$	1	1
$\{2, 2\}$	2	0
$\{2, 4\}$	3	1
$\{4, 0\}$	2	4
$\{4, 2\}$	3	1
$\{4, 4\}$	4	0

აქედან გამომდინარე, შერჩევითი საშუალოების საშუალო იქნება:

$$\frac{0+1+2+1+2+3+2+3+4}{9} = 2,$$

ხოლო დისპერსიების საშუალო კი

$$\frac{0+1+4+1+0+1+4+1+0}{9} = \frac{4}{3}.$$

როგორც ვხედავთ, შერჩევითი საშუალო პოპულაციის საშუალოს ჩაუნაცვლებელი შეფასებაა, შერჩევითი დისპერსია კი პოპულაციის დისპერსიის ჩანაცვლებული შეფასება. სინამდვილეში, $\frac{4}{3} = \frac{2-1}{2} \cdot \frac{8}{3} = \frac{n-1}{n} \cdot \frac{8}{3}$ და შესწორება ხდება სწორედ $\frac{n-1}{n}$ -ის შებრუნებული სიდიდით $\frac{n}{n-1}$, რათა მივიღოთ ჩაუნაცვლებელი შეფასება.

შევნიშნოთ, რომ თუ პოპულაციის μ საშუალო ცნობილია, მაშინ შერჩევითი დისპერსია განისაზღვრება ფორმულით:

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

უცნობი საშუალოს შემთხვევაში კი μ -ს ნაცვლად უნდა ვისარგებლოთ შერჩევითი $\bar{x}$ საშუალოთი, მაგრამ, როგორც ცნობილია (საშუალოს მეორე თვისება):

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - \mu)^2,$$

ანუ x_i -ები უფრო ახლოსაა $\bar{x}$ -თან, ვიდრე μ -სთან. ამის საკომპენსაციოდ s_n^2 -ის ფორმულაში n -ის ნაცვლად გამოიყენება $n-1$.

მოვიყვანოთ შერჩევითი დისპერსიის თვისებები:

- I. თუ არსებული $x_1, \dots, x_n$ მონაცემებიდან შევქმნით ახალ მონაცემებს $y_i = ax_i + b$, $i = 1, \dots, n$, მაშინ $s_y^2 = a^2 s_x^2$ და $s_y^2 = a^2 s_x^2$.
- II. თუ $x_1, \dots, x_n$ და $y_1, \dots, y_m$ ორი შერჩევაა, მაშინ გაერთიანებული შერჩევისათვის $x_1, \dots, x_n, y_1, \dots, y_m$ შერჩევითი დისპერსია და შესწორებული შერჩევითი დისპერსიები გადაითვლება შემდეგნაირად:

$$s^2 = \frac{n}{m+n} s_x^2 + \frac{m}{m+n} s_y^2 + \frac{mn}{(m+n)^2} (\bar{x} - \bar{y})^2$$

და

$$s'^2 = \frac{n-1}{m+n} s_x'^2 + \frac{m-1}{m+n} s_y'^2 + \frac{mn}{(m+n-1)(m+n)} (\bar{x} - \bar{y})^2$$

(იმ კერძო შემთხვევაში, როცა $m = n$ და $\bar{x} = \bar{y}$, გაერთიანებული შერჩევის შერჩევითი დისპერსია იქნება თავიდან აღებული შერჩევების შერჩევითი დისპერსიების საშუალო არითმეტიკული: $s^2 = (s_x^2 + s_y^2) / 2$).

როგორც ვხედავთ, შერჩევითი დისპერსიის განზომილება უდრის დაკვირვებათა განზომილების კვადრატს. სტატისტიკოსს ხშირად ესაჭიროება გაფანტულობის ისეთი საზომი, რომელიც იმავე ერთეულებით იზომება, რომლითაც სანყისი მონაცემები. ასეთი საზომი მიიღება შერჩევითი დისპერსიიდან კვადრატული ფესვის ამოღებით და მას სტანდარტული გადახრა (standard deviation) ან საშუალო კვადრატული გადახრა (mean square deviation) ეწოდება.

სტანდარტული გადახრა (შესაბამისად, შესწორებული სტანდარტული გადახრა) წარმოადგენს არითმეტიკულ კვადრატულ ფესვს შერჩევითი დისპერსიიდან (შესაბამისად, შესწორებული შერჩევითი დისპერსიიდან):

$$s = +\sqrt{s^2} \quad (\text{შესაბამისად, } s' = +\sqrt{s'^2}).$$

შენიშვნა: შესწორებული სტანდარტული გადახრის ნაცვლად იხმარება აგრეთვე ტერმინი შერჩევითი (შერჩევის) სტანდარტული გადახრა.

შემდგომში შერჩევითი დისპერსიისა და სტანდარტული გადახრის გამოსათვლელად გამოიყენება შედარებით მოხერხებული ფორმულები, რომლებიც მათემატიკურად ეკვივალენტურია ზემოთმოყვანილი ფორმულების, მაგრამ არ საჭიროებს საშუალოს წინასწარ გამოთვლას. ამ ფორმულებს უწოდებენ შერჩევითი დისპერსიისა და სტანდარტული გადახრის შემოკლებულ ფორმულებს და მათ აქვთ შემდეგი სახე:

$$s'^2 = \frac{\sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2 / n]}{n-1} \quad - \text{შერჩევითი დისპერსია,}$$

$$s' = \sqrt{\frac{\sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2 / n]}{n-1}} \quad - \text{შერჩევითი სტანდარტული გადახრა.}$$

საზოგადოდ, თუ x_i მონაცემის სიხშირეა f_i ($i = 1, 2, \dots, k$) და $\sum_{i=1}^k f_i = n$, მაშინ

$$s_n^2 = \frac{1}{n} \cdot \sum_{i=1}^k f_i (x_i - \bar{x})^2 \quad \text{ან} \quad s_n^2 = \frac{1}{n} \cdot \sum_{i=1}^k f_i x_i^2 - \bar{x}^2.$$

$$s_n'^2 = \frac{n}{n-1} s_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^k f_i (x_i - \bar{x})^2 \quad \text{ან} \quad s_n'^2 = \frac{1}{n-1} \cdot \sum_{i=1}^k f_i x_i^2 - \frac{n}{n-1} \bar{x}^2.$$

მაგალითი 4.3.1. ვიპოვოთ დისპერსია და სტანდარტული გადახრა ევროპაში ექვს წელიწადში გაყიდული ავტომანქანების ერთობლივი ღირებულების შერჩევის მიხედვით (მონაცემები მოყვანილია მილიონ დოლარებში):

11.2 11.9 12.0 12.8 13.4 14.3.

ამოხსნა:

ნაბიჯი 1. ვიპოვოთ მნიშვნელობათა ჯამი:

$$\sum_{i=1}^6 x_i = 11.2 + 11.9 + 12.0 + 12.8 + 13.4 + 14.3 = 75.6;$$

ნაბიჯი 2. თითოეული მნიშვნელობა ავიყვანოთ კვადრატში და ვიპოვოთ მათი ჯამი:

$$\sum_{i=1}^6 x_i^2 = 11.2^2 + 11.9^2 + 12.0^2 + 12.8^2 + 13.4^2 + 14.3^2 = 958.94;$$

ნაბიჯი 3. შევიტანოთ შესაბამის ფორმულაში და გამოვთვალოთ:

$$s^2 = \frac{\sum_{i=1}^6 x_i^2 - [(\sum_{i=1}^6 x_i)^2 / 6]}{6-1} = \frac{958.94 - [(75.6)^2 / 6]}{5} = 1.28;$$

ე. ი. შერჩევის დისპერსიაა 1.28. სტანდარტული გადახრა კი იქნება

$$s = \sqrt{1.28} \approx 1.13.$$

ცნობილია, რომ მცირე ინვესტორისათვის, როგორც საინვესტიციო ალტერნატივა, დიდი პოპულარობით სარგებლობს ინვესტიციების ჩადება საინვესტიციო ფონდებში – ტრესტებში. იმისათვის, რომ ინვესტორს გაუადვილდეს გადაწყვეტილების მიღება, თუ რომელ კონკრეტულ ტრესტში მოახდინოს ინვესტირება, ფინანსურ ჟურნალებში რეგულარულად ქვეყნდება ყოველი ცალკეული ტრესტის წლიური საპროცენტო შემოსავლის მონაცემები ბოლო ათი წლის განმავლობაში. პუბლიკაციებში, ამასთან ერთად, მითითებულია ყოველი ტრესტის რისკის დონე: მაღალი, საშუალო და დაბალი. რისკის დონეების კლასიფიკაცია კი ხდება საშუალო წლიური საპროცენტო შემოსავლის ისტორიული ცვალებადობის მიხედვით. რაც უფრო მაღალია წლიური საპროცენტო შემოსავლის ცვალებადობა, მით მაღალია რისკის დონე.

მაგალითი 4.3.2. მოცემულია ორი ტრესტის წლიური საპროცენტო შემოსავლის მონაცემები ბოლო 10 წლის განმავლობაში:

ტრესტი A	8.3	-6.2	20.9	-2.7	33.6	42.8	24.4	5.2	3.1	30.5
ტრესტი B	12.1	-2.8	6.4	12.2	27.8	25.3	18.2	10.7	-1.3	11.4

რომელი ტრესტი მიეკუთვნება რისკის უფრო მაღალ დონეს?

ამოხსნა: რადგან დისპერსია მონაცემთა ცვალებადობის ყველაზე თვალსაჩინო საზომია, ყოველი ტრესტის მონაცემების მიხედვით უნდა გამოვთვალოთ თითოეული მათგანის შესწორებული შერჩევითი დისპერსია s_A^2 და s_B^2 და შევადაროთ ისინი ერთმანეთს. წინასწარ გამოვთვალოთ შერჩევითი საშუალოები. გვაქვს:

$$\bar{x}_A = (8.3 - 6.2 + 20.9 - 2.7 + 33.6 + 42.8 + 24.4 + 5.2 + 3.1 + 30.5)/10 = 16\%,$$

$$\bar{x}_B = (12.1 - 2.8 + 6.4 + 12.2 + 27.8 + 25.3 + 18.2 + 10.7 - 1.3 + 11.4)/10 = 12\%.$$

შესაბამისად:

$$s_A^2 = (8.3^2 + \dots + 10.5^2 - 10 \cdot 16^2)/9 = (5083.36 - 2560)/9 = 280.34 (\%)^2,$$

$$s_B^2 = (12.1^2 + \dots + 11.4^2 - 10 \cdot 12^2)/9 = (2334.36 - 1440)/9 = 99.38 (\%)^2.$$

როგორც ვხედავთ, $s_A^2 > s_B^2$. ეს იმას ნიშნავს, რომ A ტრესტი მიეკუთვნება რისკის უფრო მაღალ დონეს (უფრო რისკიანია), ვიდრე B ტრესტი, იმავდროულად, $\bar{x}_A > \bar{x}_B$, ანუ A ტრესტის ამონაგები საშუალოდ უფრო მაღალია, ვიდრე B ტრესტისა. ეს ფაქტი სავსებით ეთანხმება ჩვენს ინტუიციას: ინვესტიცია, რომელიც დაკავშირებულია რისკის უფრო მაღალ დონესთან, უნდა იძლეოდეს უფრო მაღალ საშუალო ამონაგებსაც.

მაგალითი 4.3.3. დავუშვათ, რომ 10 წლის წინ თქვენ მოახდინეთ ტოლი თანხების ინვესტირება A და B ტრესტებში, ე. ი. შექმენით საინვესტიციო პორტფელი, რომელშიც A და B ტრესტებში ინვესტიციათა წონები ტოლია და უდრის 1/2-ს. ახალი პორტფელი აღვნიშნოთ C-თი. შევადაროთ A, B და C პორტფელების რისკიანობა.

ამოხსნა: წინასწარ შემოვიღოთ შემოსავლიანობის ცნება: დავუშვათ, რომ საინვესტიციო თანხა წლის დასაწყისში უდრის P_0 -ს და ამ თანხის ინვესტირებამ A და B ტრესტებში წლის ბოლოს მოგვცა, შესაბამისად, P_A და P_B თანხები, მაშინ A და B ტრესტების ამონაგები R_A და R_B შემდეგი ფორმულებით მოიცემა (%-ში):

$$R_A = \frac{P_A - P_0}{P_0} \cdot 100\%, \quad R_B = \frac{P_B - P_0}{P_0} \cdot 100\%.$$

ვთქვათ, ახლა A და B ტრესტებიდან თითოეულში ინვესტირებულია $\frac{1}{2}P_0$ -ის ტოლი თანხა, მაშინ წლის ბოლოს C პორტფელისაგან მიღებული თანხაა $\frac{1}{2}P_A + \frac{1}{2}P_B$. ამიტომ მისი ამონაგები იქნება:

$$R_C = \frac{(\frac{1}{2}P_A + \frac{1}{2}P_B) - P_0}{P_0} = \frac{\frac{1}{2}(P_A - P_0) + \frac{1}{2}(P_B - P_0)}{P_0} = \frac{1}{2}R_A + \frac{1}{2}R_B.$$

გასაგებია, რომ წინა მაგალითში მოყვანილი მონაცემები წარმოადგენენ ბოლო 10 წლის დაკვირვებებს R_A -სა და R_B -ზე, ამიტომ ჩვენ შეგვიძლია შევქმნათ C პორტფელის ამონაგებთა მწკრივი. ის იქნება:

10.2 -4.5 13.65 4.75 30.7 34.1 21.3 7.95 0.9 20.95.

სტანდარტული გადახრის მეშვეობით შევადაროთ C პორტფელის რისკიანობა და ტრესტებში ინვესტიციითა რისკიანობას. საშუალოს მე-3 და მე-4 თვისებების საფუძველზე შეგვიძლია დავნეროთ, რომ:

$$\bar{x}_A = \frac{1}{2}(\bar{x}_A + \bar{x}_B) = 14\%.$$

ამიტომ, შესწორებული შერჩევითი დისპერსია იქნება:

$$s_C'^2 = (10.2^2 + \dots + 20.95^2 - 10 \cdot 14^2) / 9 = 159.45(\%)^2,$$

შესაბამისად, $s_C' = 12.63\%$.

როგორც ვხედავთ, რისკიანობის მიხედვით A, B და C ინვესტიციები დალაგდა შემდეგი მიმდევრობით:

$$s_B' < s_C' < s_A',$$

ხოლო მათი ამონაგები კი შემდეგი მიმდევრობით:

$$R_A > R_C > R_B.$$

დისპერსიისა და სტანდარტული გადახრის მოძებნის პროცედურა დაჯგუფებული მონაცემებისათვის ანალოგიურია დაჯგუფებული მონაცემებისათვის საშუალოს მოძებნის პროცედურის და ის იყენებს თითოეული კლასის შუანერტილს.

მაგალითი 4.3.4. კვირის განმავლობაში შემთხვევით შერჩეული 20 მორბენალის მიერ გარბენილი მანძილების ზემოთ მოყვანილი სიხშირული განაწილების ცხრილის მიხედვით გამოვთვალოთ დისპერსია და სტანდარტული გადახრა.

ამოხსნა:

ნაბიჯი 1. გავაკეთოთ შემდეგი სახის ცხრილი, ვიპოვოთ თითოეული კლასის შუანერტილი და შევიტანოთ C სვეტში:

$$x_m = \frac{5.5 + 10.5}{2} = 8, \quad \frac{10.5 + 15.5}{2} = 13 \quad \text{და ა. შ.}$$

ნაბიჯი 2. გავამრავლოთ თითოეული კლასის სიხშირე ამავე კლასის შუანერტილზე და ნამრავლი ჩავწეროთ D სვეტში:

$$1 \cdot 8 = 8, \quad 2 \cdot 13 = 26, \quad \text{და ა. შ.} \quad 2 \cdot 38 = 76.$$

ნაბიჯი 3. გავამრავლოთ თითოეული კლასის სიხშირე ამავე კლასის შუანერტილის კვადრატზე და ნამრავლი ჩავწეროთ E სვეტში:

$$1 \cdot 8^2 = 64, \quad 2 \cdot 13^2 = 338, \quad \text{და ა. შ.} \quad 2 \cdot 38^2 = 2888.$$

ნაბიჯი 4. ვიპოვოთ B, D და E სვეტების ჯამი. B სვეტის ჯამია n, D სვეტის ჯამია $\sum f \cdot x_m$, ხოლო E სვეტის ჯამია $\sum f \cdot x_m^2$. საბოლოოდ, ცხრილს ექნება ასეთი სახე:

A კლასები	B სიხშირე (f)	C შუანერტილი (x_m)	D $f \cdot x_m$	E $f \cdot x_m^2$
5.5-10.5	1	8	8	64
10.5-15.5	2	13	26	338
15.5-20.5	3	18	54	972
20.5-25.5	5	23	115	2645
25.5-30.5	4	28	112	3136
30.5-35.5	3	33	99	3267
35.5-40.5	2	38	76	2888
	$n = 20$		$\sum f \cdot x_m = 490$	$\sum f \cdot x_m^2 = 13310$

ნაბიჯი 5. შევიტანოთ მიღებული მონაცემები შესაბამის ფორმულაში და გამოვთვალოთ დისპერსია:

$$s^2 = \frac{\sum_{m=1}^n f \cdot x_m^2 - [(\sum_{m=1}^n f \cdot x_m)^2 / n]}{n-1} = \frac{13310 - [490^2 / 20]}{20-1} = 68.7.$$

ნაბიჯი 6. სტანდარტული გადახრის გამოსათვლელად ამოვიღოთ კვადრატული ფესვი დისპერსიიდან:

$$s' = \sqrt{68.7} \approx 8.3.$$

ცხადია, რომ $s' = 0$, მაშინ და მხოლოდ მაშინ, როდესაც შერჩევის ყველა მნიშვნელობა ტოლია. საჭიროა აღინიშნოს, რომ სტანდარტული გადახრა არ წარმოადგენს გაფანტულობის მდგრად საზომს, ის გაცილებით უფრო მგრძნობიარეა ექსტრემალური დაკვირვებების მიმართ, ვიდრე საშუალო. მართლაც, განვიხილოთ მონაცემები: 34 46 40 57 50, მაშინ მათთვის საშუალო იქნება $\bar{x} = 46$. გამოვთვალოთ სტანდარტული გადახრა s' . ჯერ გამოვთვალოთ საშუალოდან გადახრები, გვექნება: -9 0 6 11 4. შესაბამისად:

$$s^2 = (81 + 0 + 36 + 121 + 16) / 4 = 63.5. \text{ ამიტომ } s' = \sqrt{63.5} \approx 8.2.$$

ახლა მონაცემთა მოყვანილ სიმრავლეში მონაცემი 50 შევცვალოთ 200-ით ანუ განვიხილოთ ახალი მონაცემები: 34 46 40 57 200. ამ შემთხვევაში საშუალო იქნება $\bar{x} = 76$, ხოლო სტანდარტული გადახრა კი:

$$s' = \sqrt{(39^2 + 30^2 + 36^2 + 19^2 + 124^2) / 4} = \sqrt{4383.5} \approx 66.$$

როგორც ვხედავთ, შედარებით თანაბრად განაწილებული მონაცემებიდან ერთ-ერთი მონაცემის მკვეთრ ცვლილებაზე სტანდარტული გადახრის რეაგირება უფრო ძლიერი აღმოჩნდა, ვიდრე საშუალოსი. ამ ცვლილებით მონაცემები გახდა

მარჯვნივ ასიმეტრიული. მკვეთრად ასიმეტრიული, მაგალითად, მარჯვნივ ასიმეტრიული განაწილებებისათვის, მარჯვენა მხარეს გრძელი კუდით, შერჩევითი დისპერსია დიდია და არ იძლევა სასარგებლო ინფორმაციას გაფანტულობის შესახებ.

იმ შემთხვევაში, როდესაც ორი შერჩევის მონაცემები გაზომილია ერთსა და იმავე ერთეულებში, მაგალითად, სანტიმეტრებში, შერჩევას, რომლისთვისაც $s'=5.9$, ახასიათებს მეტი გაფანტულობა, ვიდრე შერჩევას, რომლისთვისაც $s'=2.1$. მაგრამ, როგორ უნდა მოვიქცეთ, თუ შესაძარებელია სხვადასხვა ერთეულში გაზომილი მონაცემების ცვალებადობა?

მაგალითი 4.3.5. სამთვიანი პერიოდის განმავლობაში გაყიდული მანქანების საშუალოა – 87, სტანდარტული გადახრა – 35, გადახდილი თანხების საშუალო – 3567 ლარი, სტანდარტული გადახრა კი 775 ლარი. სად არის უფრო დიდი ცვალებადობა?

სტატისტიკას, რომელიც სტანდარტული გადახრების შედარების საშუალებას იძლევა, იმ შემთხვევაში, როცა ცვლადები გაზომილია სხვადასხვა ერთეულებში, **ვარიაციის კოეფიციენტი (coefficient of variation)** ეწოდება. ის გამოისახება პროცენტებში, აღინიშნება $CVar$ სიმბოლოთი და განიმარტება შემდეგნაირად:

$$CVar := \frac{s}{x} \cdot 100\%.$$

4.3.5 მაგალითის მონაცემებისათვის გვაქვს:

$$CVar = \frac{35}{87} \cdot 100\% = 40.2\% - \text{გაყიდვების შემთხვევაში,}$$

$$CVar = \frac{775}{3567} \cdot 100\% = 21.7\% - \text{თანხების შემთხვევაში.}$$

შესაბამისად დავასკენით, რომ თანხები ნაკლებად ცვალებადია, ვიდრე გაყიდვები.

სტანდარტული გადახრის სიდიდის ინტერპრეტაცია

სტანდარტული გადახრის სიდიდის ინტერპრეტაციისთვის საჭიროა პოპულაციის განაწილების ფორმის ცოდნა, მაგალითად, ემპირიული წესი.

ემპირიული წესი (empirical rule):

თუ განაწილება ზარისებურია, მაშინ: პოპულაციის მონაცემთა მნიშვნელობების დაახლოებით 68% მოთავსებულია საშუალოდან ერთი სტანდარტული გადახრის ფარგლებში.

მონაცემთა მნიშვნელობების დაახლოებით 95% მოთავსებულია საშუალოდან ორი სტანდარტული გადახრის ფარგლებში.

მონაცემთა მნიშვნელობების დაახლოებით 99.7% მოთავსებულია საშუალოდან სამი სტანდარტული გადახრის ფარგლებში.

დავუშვათ, რომ ეროვნულ გამოცდებზე ქულების საშუალოა – 480, ხოლო სტანდარტული გადახრა – 90. თუ ვიგულისხმებთ, რომ ქულების განაწილების ფორმა ზარისებურია, მათი დაახლოებით 68% მოთავსებული იქნება 390-სა და

570-ს შორის. ქულების დაახლოებით 95% ჩაეარდება 300-სა და 660-ს შორის. ქულათა 99.7% მოხვდება შუალედში (210; 750).

ნებისმიერი სხვა განაწილებისათვის, მონაცემთა მნიშვნელობების, სულ ცოტა 3/4, ანუ 75% ეარდება საშუალოდან ორი სტანდარტული გადახრის მიდამოში და მონაცემთა 88.89% –საშუალოდან სამი სტანდარტული გადახრის ფარგლებში.

ჩებიშევის თეორემა. მონაცემთა სიმრავლიდან იმ მონაცემების წილი, რომლებიც მოთავსებულია საშუალოდან k სტანდარტული გადახრის ფარგლებში, არის, სულ ცოტა, $1 - 1/k^2$, სადაც k არის 1-ზე დიდი რიცხვი (k არ არის აუცილებლად მთელი).

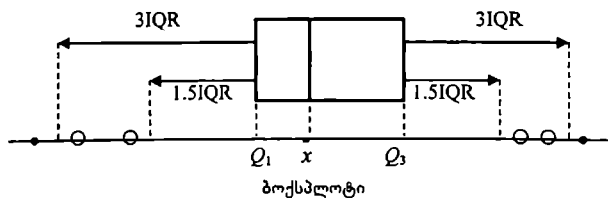
ფორმალურად ეს შედეგი ასე ჩამოყალიბდება: დაკვირვებათა მოცემული $x_1, \dots, x_n$ სიმრავლისათვის სამართლიანია შემდეგი უტოლობა:

$$\frac{N\{x_i : |x_i - \bar{x}| \leq ks\}}{n} \geq 1 - \frac{1}{k^2}.$$

სხვა სიტყვებით: $[\bar{x} - ks, \bar{x} + ks]$ ინტერვალში მოხვედრილ მონაცემთა ფარდობითი სიხშირე (ანუ, ამ ინტერვალში მოხვედრილ მონაცემთა რაოდენობის შეფარდება შერჩევის მოცულობასთან) არანაკლებია $(1 - 1/k^2)$ -ზე.

ბოქსპლოტი

უკანასკნელ პერიოდში განსაკუთრებული პოპულარობით სარგებლობს მონაცემთა განაწილების ვიზუალური წარმოჩენის ახალი საშუალება – ე. წ. **ბოქსპლოტი** (boxplot). ბოქსპლოტი მონაცემთა გაფანტულობას გრაფიკულად წარმოადგენს. მისი მეშვეობით გამოიკვეთება მონაცემთა ისეთი თავისებურება, როგორიცაა განლაგების ცენტრი, გაფანტულობა, სიმეტრიულობიდან გადახრა. როგორც ცნობილია, ფოთლებიანი ღეროების მსგავსი დიაგრამა და ჰისტოგრამა გვაძლევს მონაცემთა სიმრავლის ზოგად დახასიათებას, ბოქსპლოტი კი იძლევა მონაცემთა სიმრავლის გრაფიკულ დახასიათებას. ბოქსპლოტი მდგრადი საზომების – ($Q_1, \bar{x}, Q_3, IQR$) მეშვეობით საშუალებას იძლევა გამოიკვეთოს მონაცემთა სიმრავლის ზოგიერთი თავისებურება: ა). განლაგების ცენტრი; ბ). გაფანტულობა; გ). გავრცობა და სიმეტრიულობიდან გადახრის ბუნება; დაბოლოს, მოხდეს დ). „ამოვარდნილი (outlier)“ დაკვირვებების იდენტიფიკაცია. აღვნიშნოთ, თუ როგორ ხდება ბოქსპლოტის აგება:



1. გავევლოთ პორიზონტალური ღერძი, რომელზედაც მოენიშნოთ Q_1 , $\bar{x}$ და Q_3 ;
2. ავაგოთ მართკუთხედი, რომლის გვერდები მართობულია პორიზონტალური ღერძის და რომლის მარცხენა გვერდი Q_1 -ის, ხოლო მარჯვენა კი Q_3 -ის ზემოთაა მოთავსებული;
3. გავევლოთ ვერტიკალური წრფე მართკუთხედში მედიანის ზევით;
4. მართკუთხედის თითოეული გვერდიდან გავევლოთ პორიზონტალური ღერძის პარალელური ორი წრფე, პირველზე გადავზომოთ $1.5 IQR$, ხოლო მეორეზე – $3 IQR$ სიგრძის მონაკვეთები;
5. ყოველი მონაცემის შესაბამის ნერტილზე, რომელიც მართკუთხედის გვერდიდან გადაზომილ $1.5 IQR$ -სა და $3 IQR$ -ს შორისაა მოთავსებული, შემოვხაზოთ წრეწირი. ამ მონაცემებს ეწოდება ზომიერი „ამოვარდნები“. იმ მონაცემების შესაბამის ნერტილებზე, რომლებიც მოთავსებულია მართკუთხედის გვერდიდან გადაზომილი $3 IQR$ -ის გარეთ, შემოვხაზოთ გამუქებული წრე. ასეთი წრეებით შემოხაზულ მონაცემებს კი ექსტრემალური „ამოვარდნები“ ეწოდება.

ხშირად გრაფიკს აბრუნებენ 90° -იანი კუთხით და გამოსახვენ ოდნავ მოდიფიცირებული სახით. კერძოდ, ვერტიკალურ ღერძზე გადაიზომება $x_{\min}$, Q_1 , $\bar{x}$, Q_3 , $x_{\max}$ და მხოლოდ $1.5 IQR$ მანძილი მართკუთხედის გვერდებიდან და ისიც მხოლოდ იმ შემთხვევაში, როდესაც $x_{\min}$ ან $x_{\max}$ აღმოჩნდება მის გარეთ. ზოგჯერ ამ გრაფიკზე პირველ და მეცხრე დეცილსაც მიუთითებენ.

გარდა ზომიერი და ექსტრემალური „ამოვარდნების“ საიდენტიფიკაციო წესისა, გამოიყენება ე. წ. უხეში წესი იმ მონაცემების აღმოსაჩენად, რომელთა შესახებ შეიძლება გაჩნდეს ეჭვი, რომ ისინი არ არიან ტიპური მონაცემთა მოცემული სიმრავლისათვის. ეს წესი მდგომარეობს შემდეგში: გამოვყოთ ის მონაცემები, რომლებიც აღემატებიან $(Q_3 + 1.5 IQR)$ -ს და ის მონაცემები, რომლებიც ნაკლებია $(Q_1 - 1.5 IQR)$ -ზე. იმის დასადგენად, არიან თუ არა ეს მონაცემები „ამოვარდნილები“ (ანუ მკვეთრად მოშორებულნი მონაცემთა ძირითად მასივს), უნდა აიგოს ფოთლებიანი ღეროების მსგავსი დიაგრამა.

შეჯამება

- ✓ ცენტრალური ტენდენციის, მონაცემთა განლაგების ცენტრის საზომებია: საშუალო, მედიანა, მოდა. ცენტრალური ტენდენციის საზომების გამოთვლა გულისხმობს მონაცემებში „ტიპური“ მონაცემის მოძებნას.
- ✓ საშუალოს გამოსათვლელად უნდა შევკრიბოთ ყველა მონაცემი და მიღებული ჯამი გავყოთ მონაცემთა რაოდენობაზე.
- ✓ მედიანა ზრდის მიხედვით დალაგებულ მონაცემებს ორ ნაწილად ყოფს, მონაცემთა ტოლი რაოდენობით. მონაცემთა ნახევარი ნაკლებია ან ტოლი $\leq$ მედიანაზე, მეორე ნახევარი კი მეტია ან ტოლი $\geq$ მედიანაზე.

- ✓ პროცენტები – მონაცემთა მდებარეობის საზომებია. მონაცემთა მეოთხედი ნაკლებია ან ტოლი 25-პროცენტებზე (პირველ კვარტილზე). მონაცემთა სამი მეოთხედი ნაკლებია ან ტოლი 75-პროცენტებზე (მესამე კვარტილზე.) კვარტილები და მედიანა ყოფს მონაცემებს ოთხ, რაოდენობით ტოლ ნაწილად.
- ✓ მოდა – საზომია, რომელიც გამოითვლება როგორც რაოდენობრივი, ასევე თვისებრივი მონაცემებისათვის.
- ✓ საშუალო ზედმეტად მგრძნობიარეა ექსტრემალური (მონაცემთა ძირითადი მასისგან გამორჩეული) დაკვირვებების მიმართ. მედიანა კი მდგრადია ასეთი დაკვირვებების მიმართ, მაგრამ არ უნდა ვიფიქროთ, რომ მედიანას ყოველთვის უნდა მიენიჭოს უპირატესობა საშუალოსთან შედარებით. საშუალო გამოითვლება ყველა მონაცემის მეშვეობით და შეიცავს მეტ ინფორმაციას. არსებობს მრავალი პოპულაცია, რომელთათვის ექსტრემალური დაკვირვება ნაკლებად მოსალოდნელია.
- ✓ ცვალებადობის საზომები ზომავენ მონაცემთა გაფანტულობას.
- ✓ გაბნევის დიაპაზონი – მანძილია უმცირესი მონაცემიდან უდიდეს მონაცემამდე.
- ✓ კვარტილთშორისი გაბნევის დიაპაზონი – მანძილია პირველ და მესამე კვარტილებს შორის.
- ✓ დისპერსია და სტანდარტული გადახრა ზომავს ცვალებადობას საშუალოს მიმართ.
- ✓ ბოქსპლოტი არის კვარტილების და ამოვარდნილი მონაცემების „პორტრეტი“.

ძირითადი ტერმინები

აღწერითი სტატისტიკა, საშუალო, მედიანა, მოდა, პროცენტილი, კვარტილი, დეცილი, უნიმოდალური განაწილება, ბიმოდალური განაწილება, მულტიმოდალური განაწილება, ცენტრალური ტენდენციის საზომები, ცვალებადობის საზომები, გაბნევის დიაპაზონი, კვარტილთშორისი გაბნევის დიაპაზონი, დისპერსია, სტანდარტული გადახრა, ვარიაციის კოეფიციენტი, ბოქსპლოტი.

კითხვები

1. იპოვეთ გაბნევის დიაპაზონი და 25-პროცენტილი (I კვარტილი) შემდეგი სიმრავლეებისათვის:
ა) 7 9 12 13 8 11;
ბ) 7 8 22 20 15 18 19 13 11.
2. იპოვეთ გაბნევის დიაპაზონი და 75-პროცენტილი (III კვარტილი) 20 სტუდენტის მასებისათვის:
106 114 118 118 122 127 127 135 136 138
139 140 143 146 146 147 153 155 159 175.
3. შეაფასეთ 25-პროცენტილი და 75-პროცენტილი CD დისკების ჩანაწერის ხანგრძლივობა (წუთებში) გვ. 56-ე მოყვანილი №10 სიხშირული განაწილების ცხრილის მიხედვით.

4. იპოვეთ გაბნევის დიაპაზონი და 45-პროცენტილი მონაცემთა შემდეგი ერთობლიობებისათვის:

ა) 7 4 14 9 12 2 19 6 15;

ბ) 7.6 4.8 1.2 6.9 4.8 7.2 8.1 10.3 4.8 6.7.

5. 50 კვირის მანძილზე აკვირდებოდნენ გარკვეული დანადგარის მუშაობას და აფიქსირებდნენ დაზიანებათა რაოდენობას კვირის განმავლობაში. მიღებულ იქნა შემდეგი მონაცემები:

დაზიანებათა რიცხვი 0 1 2 3 4 5 6

კვირათა რიცხვი 2 12 14 8 8 4 2

ა) იპოვეთ გაბნევის დიაპაზონი და 35-პროცენტილი.

ბ) იპოვეთ კვირაში დანადგარის დაზიანებათა რიცხვის 95-პროცენტილი.

6. იპოვეთ მონაცემების საშუალო:

ა) 1 2 3 4 5 6 7; ბ) 4 12 -2 7 0 9.

გამოთვალეთ თითოეული სიმრავლის სტანდარტული გადახრა და გაერთიანებული მონაცემების სტანდარტული გადახრა.

7. მე-10 კლასელი 204 მოსწავლის ძმებისა და დების რიცხვის ზემოთმოყვანილი სიხშირული ცხრილის მიხედვით (იხ. მაგ. 3.1.4) გამოთვალეთ შერჩევითი დისპერსია.

8. გამოთვალეთ სტანდარტული გადახრა მონაცემთა თითოეული სიმრავლისათვის:

ა) 2 1 5.3 -4.2 6.7 3.1;

ბ) 15.2 12.3 5.7 4.3 11.2 2.5 8.7.

9. გრამებში გაზომილი ხიზილალის 25 ქილის მასების ჯამი და მასების კვადრატების ჯამი აღმოჩნდა შესაბამისად:

$$\sum_{i=1}^{25} x_i = 1268.2 \quad \text{და} \quad \sum_{i=1}^{25} x_i^2 = 64585.16.$$

იპოვეთ მასების საშუალო და დისპერსია. რომელი ერთეულით იზომება დისპერსია?

10. ქვემოთ მოყვანილია, კრიკეტის 12 თამაშის განმავლობაში, ტომისა და ვილის მიერ ტაფელით მოგერიებული ბურთების რაოდენობა:

ტომი 23 83 40 0 89 98 71 31 102 48 15 18

ვილი 43 32 61 75 68 92 17 15 25 43 86 12

თქვენი აზრით, რომელი მოთამაშეა: ა) უკეთესი; ბ) უფრო სტაბილური.

11. კლასში 12 ბიჭის სიმაღლეების საშუალო და სტანდარტული გადახრა, შესაბამისად, ტოლია 148.8 სმ და 5.4 სმ. მათ დაემატა ერთი ბიჭი, რომლის სიმაღლეა 153.4 სმ. გამოთვალეთ 13 ბიჭის სიმაღლეების საშუალო და სტანდარტული გადახრა.

12. დიდი პოპულაციიდან აღებული შემთხვევითი შერჩევა შედგება შემდეგი მონაცემებისაგან:

13.2 5.7 8.3 6.7 9.2 11.4 9.7 8.1 6.3

გამოთვალეთ პოპულაციის საშუალოსა და სტანდარტული გადახრის ჩაუნაცვლებელი შეფასებები.

13. კლასში 25 მოსწავლეა, რომელთაგან 12 ვაჟია და 13 ქალი. მათემატიკაში ტესტირების შედეგად 12 ვაჟის საშუალო ქულამ შეადგენა 31, ხოლო სტანდარტული

გადახრა იყო 6.2. გოგონების საშუალო ქულა იყო 36, სტანდარტული გადახრა კი 4. ვიპოვოთ მათემატიკაში ტესტირების ქულების საშუალო და სტანდარტული გადახრა მთელი კლასის 25 მოსწავლისათვის.

14. ცნობილია, რომ $n = 10$, $\sum_{i=1}^n x_i = 410$ და $s_n^2 = 12$. იპოვეთ $\sum_{i=1}^n x_i^2$.
15. 80 სტუდენტი გოგონას სიმაღლეების ჯამი და სიმაღლეების კვადრატების ჯამია, შესაბამისად: $\sum_{i=1}^{80} x_i = 13040$ და $\sum_{i=1}^{80} x_i^2 = 2133520$. იპოვეთ სიმაღლეების საშუალო და სტანდარტული გადახრა.
16. ქვემოთ მოყვანილია 96 დღის განმავლობაში აკადემიური პერსონალის მიერ გაცემილი ლექციების რაოდენობა უნივერსიტეტში:

გაცემილი ლექციების

რაოდენობა 0 1 2 3 4 5

დღეების რაოდენობა 54 24 11 4 2 1

გამოთვალეთ გაცემილი ლექციების რაოდენობის საშუალო და სტანდარტული გადახრა.

17. მენეჯერი სამუშაო დღის ბოლოს ამოწმებს მხატვრის მიერ დღის განმავლობაში მოხატული თეფშების ხარისხს და მათ ნაწილს ინუნებს. ქვემოთ მოყვანილია 30 დღის განმავლობაში მენეჯერის მიერ დანუებული თეფშების რიცხვი:

დანუებული თეფშების

რიცხვი 0 1 2 3 4 5 6

დღეების რიცხვი 18 5 3 1 1 1 1

შეამოწმეთ, რომ ნუნდებული თეფშების რაოდენობის სტანდარტული გადახრა დაახლოებით ტოლია გაბნევის დიაპაზონის მეოთხედის.

18. ქვემოთ მოყვანილია 80 მორბენალის მიერ 20-კილომეტრიან მარათონულ დისტანციაზე დახარჯული დროების (წუთებში) დაჯგუფებული სიხშირული განაწილების ცხრილი:

სირბილის

დრო 60–80 80–100 100–120 120–140 140–160 160–180 180–200

მორბენალთა

რიცხვი 1 4 26 24 10 7 8

შეაფასეთ სირბილის დროის შერჩევითი დისპერსია. თუ ეს 80 მორბენალი შემთხვევით იქნა შერჩეული მარათონში მონაწილე 2000 მორბენალიდან, რა იქნება 2000 მორბენალის დისპერსიის უკეთესი შეფასება.

19. შეარჩიეს ყავის 80 პაკეტი და მათი მასები (გრამებში) განაწილებული აღმოჩნდა შემდეგნაირად:

პაკეტის მასა 244–246 247–249 250–252 253–255 256–258

პაკეტის რიცხვი 10 20 24 18 8

შეაფასეთ მასების საშუალო და სტანდარტული გადახრა. როგორ შეიძლება შეფასების სიზუსტის გაზრდა?

20. საწარმოს 104 მუშის ასაკი განაწილებულია შემდეგნაირად:

ასაკი	16-20	21-25	26-30	31-35	36-40	41-50	51-60	61-70
სიხშირე	5	12	18	14	25	16	8	6

შეაფასეთ მუშების ასაკის საშუალო და სტანდარტული გადახრა. სხვა საწარმოში, სადაც მუშათა რიცხვი ისევ 104-ია, ასაკის საშუალო შეადგენს 28.4 წელს, ხოლო სტანდარტული გადახრა კი 9.9 წელია. შეადარეთ ამ საწარმოების მუშათა ასაკის განაწილებები.

21. გამოთვალეთ, რომელი ჯგუფის მონაცემების ვარიაციის კოეფიციენტი მეტი:

ჯგუფი 1: 1.2 1.7 1.1 2.1 0.9 1.6 1.4 1.6 1.8 0.8;

ჯგუფი 2: 217 287 196 225 276 298 313 356 244 234.

22. სამშენებლო კომპანიამ წარადგინა ინფორმაცია ბოლო ათი ერთნაირი ზომის სახლის ამენების ვადის შესახებ (სთ): 245, 291, 353, 118, 152, 402, 351, 242, 595, 175.

ა. გამოთვალეთ საშუალო და ახსენით რას ნიშნავს ეს სიდიდე;

ბ. გამოთვალეთ მედიანა და ახსენით რას ნიშნავს ეს სიდიდე;

გ. გამოთვალეთ სტანდარტული გადახრა და ახსენით რას ნიშნავს ეს სიდიდე;

დ. თუ გსურთ კონტრაქტის გაფორმება ამ კომპანიასთან, მიაქცევთ თუ არა ყუ-რადლებას ამ სიდიდეებს?

23. ბოტანიკოსმა შეაგროვა მონაცემები ერთი წლის წინ დარგულ ნერგებზე. ცხრილში მოცემულია, ყოველი 10 დარგულიდან რამდენმა გაიხარა:

გახარებული (10-დან)	0	1	2	3	4	5	6	7	8	9	10
სიხშირე	0	0	1	1	1	7	8	10	23	15	12

ა. რამდენი 10-ნერგიანი უბანი იყო შემონმებული?

ბ. რამდენი ნერგი იყო დარგული?

გ. რამდენმა ნერგმა გაიხარა?

დ. საშუალოდ ყოველი 10 დარგულიდან რამდენმა გაიხარა?

24. ქვემოთ მოცემულია 23 პაციენტის განკურნების დრო (თვეები) სტანდარტული მკურნალობის შემთხვევაში და 19 პაციენტის განკურნების დრო ექსპერიმენტული მკურნალობის შემთხვევაში. დააეადება, დააეადების სიმძიმე, პაციენტების ასაკი და სქესი იყო ერთი და იგივე.

სტანდარტული მკურნალობა

ექსპერიმენტული მკურნალობა

9 13 13 16 17 17 21 24

19 27 28 35 38 39

25 25 31 40 43 50 50 51

44 44 46 47 54 57 58

54 56 56 60 62 62 64

58 59 63 65 65 68

ა. თითოეული ჯგუფისათვის იპოვეთ კვარტილები, IQR , დიაპაზონი;

ბ. არის თუ არა ჯგუფებში ექსტრემალური მონაცემები?

გ. ააგეთ ორი ბოქსპლოტი და შეადარეთ ერთმანეთს.

25. ცხრილში მოყვანილია შემთხვევით შერჩეული ხანდაზმული წყვილების წლიური შემოსავალი:

შემოსავალი (1000 დოლარი)	წყვილების რაოდენობა
5 000–9 999	2
10 000–14 999	27
15 000–19 999	59
20 000–24 999	60
25 000–29 999	28
30 000–34 999	18
35 000–39 999	5
40 000–45 999	1

ა. იპოვეთ, დაახლოებით რას უდრის საშუალო შემოსავალი და მედიანა;

ბ. იპოვეთ, დაახლოებით რას უდრის სტანდარტული გადახრა.

26. ცხრილში მოცემულია 30 კომპანიის აქტივები (მილიონი დოლარი):

აქტივები	0-დან 2-მდე	2-დან 4-მდე	4-დან 6-მდე	6-დან 8-მდე	8-დან 10-მდე	10-დან 12-მდე
სიხშირე	3	5	10	6	4	2

გამოთვალეთ დაახლოებით რას უდრის აქტივების მედიანა, საშუალო, სტანდარტული გადახრა.

27. 100 უბანზე დათვლილი იყო ბაქტერიების გარკვეული სახეობის რაოდენობა:

უბანზე რაოდენობა 0 1 2 3 4 5 6

უბნების რაოდენობა 8 17 28 23 16 5 3

იპოვეთ უბანზე ბაქტერიების საშუალო რაოდენობა და სტანდარტული გადახრა.

თავი 5

დანყვილებული მონაცემები

აქამდე ჩვენ ვიხილავდით მონაცემებს, რომლებიც წარმოადგენდა რაიმე ერთი ცვლადის დაკვირვებულ მნიშვნელობებს. მრავალი პრობლემის კვლევისას საქმე გვაქვს ისეთ მონაცემებთან, რომლებიც წარმოადგენენ ორი ან მეტი ცვლადის ერთდროულად დაკვირვებულ (გაზომილ) მნიშვნელობათა სიმრავლეს. ხშირ შემთხვევაში მკვლევარი დაინტერესებულია ორ ან მეტ ცვლადს შორის კავშირის არსებობის დადგენაში. მაგალითად, არსებობს თუ არა კავშირი სტუდენტის მიერ სწავლაზე დახარჯული საათების რაოდენობასა და მის მიერ გამოცდაზე მიღებულ ქულას შორის, ან არის თუ არა ეროვნული გამოცდების ჩაბარების ხარისხი დამოკიდებული აბიტურიენტის სოციალურ მდგომარეობაზე, საცხოვრებელ ადგილზე, სკოლაში სწავლების ხარისხზე.

ამ მიზნის მისაღწევად, მკვლევარმა უნდა შეაგროვოს მონაცემები ორივე მახასიათებელზე ერთდროული დაკვირვებების შედეგად და მიღებული მონაცემები დაანევილოს ერთმანეთთან (უფრო ზუსტად, ორი დამკვირვებელი ერთდროულად ატარებს დაკვირვებებს და მიღებულ მონაცემებს აწყვილებს ერთმანეთთან). მაგალითად, თუ ჩვენ გვინტერესებს კავშირი ჰაერის ტემპერატურასა და ატმოსფერულ წნევას შორის, მაშინ დროის ერთსა და იმავე მომენტში პირველი დამკვირვებელი ზომავს ჰაერის ტემპერატურას, ხოლო მეორე – ატმოსფერულ წნევას. პირველი დაკვირვების ან გაზომვის ობიექტს (ამ შემთხვევაში – ტემპერატურას) უწოდებენ **დამოუკიდებელ სიდიდეს** და აღნიშნავენ X -ით, ხოლო მეორე დაკვირვების ან გაზომვის ობიექტს (ამ შემთხვევაში – ატმოსფერულ წნევას) უწოდებენ **დამოკიდებულ სიდიდეს** და აღნიშნავენ Y -ით.

მაგალითი 5.1. დაკავშირებულია თუ არა ბიოლოგიაში მიღებული ქულა მათემატიკაში მიღებულ ქულასთან?

ცხრილი 5.1

ცდისპირის ნომერი	1	2	3	4	5	6	7	8	9	10	11	12	13	14
მათემატიკაში მიღებული ქულა, (x)	4	8	8	10	13	16	20	23	28	30	33	35	38	42
ბიოლოგიაში მიღებული ქულა, (y)	8	10	32	35	37	47	47	55	60	60	80	70	70	75

5.1 ცხრილში მოყვანილია 14 სტუდენტის დანევილებული მონაცემი. მონაცემების დათვალიერებით შევემჩნევთ, რომ, რაც უფრო მაღალია მათემატიკაში მიღებული ქულა, მით უფრო მაღალია ბიოლოგიის ქულაც. მაგრამ ეს დამოკიდებუ-

ლება არ არის სრულყოფილი. მაგალითად, მე-14 პირს აქვს მათემატიკის ყველაზე მაღალი ქულა – 42, მაგრამ მე-11 პირზე დაბალი ბიოლოგიის ქულა.

გაბნევის დიაგრამა

ამა თუ იმ ორ მახასიათებელს (ან ორ მაჩვენებელს) შორის კავშირის შესწავლის მიზნით ამ მახასიათებლების დაკვირვებული მნიშვნელობების სიმრავლე წარმოადგინოთ წყვილების შემდეგი სიმრავლის სახით: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. ამ მონაცემების მიხედვით აგებენ ე. წ. **გაბნევის დიაგრამას** შემდეგი წესით: სიბრტყეზე, კოორდინატთა მართკუთხა სისტემაში მონიშნავენ წერტილებს, რომელთა კოორდინატებია (x_i, y_i) , $i = 1, 2, \dots, n$.

გაბნევის გრაფიკი ან გაბნევის დიაგრამა წარმოადგენს მონაცემთა მნიშვნელობების დალაგებული წყვილების გრაფიკს (დიაგრამას), რომელიც გამოიყენება ორ სიდიდეს შორის არსებული კავშირის დასადგენად.

მაგალითი 5.2. მკვლევარს აინტერესებს დაადგინოს, არსებობს თუ არა კავშირი ველოსიპედით წვიმიან ამინდში მომხდარ მსუბუქ ავარიათა რიცხვსა და სიკვდილით (ფატალური შედეგით) დამთავრებულ ავარიათა რიცხვს შორის. ქვემოთ მოყვანილია 10 წლის განმავლობაში წვიმიან ამინდში ველოსიპედით მომხდარი ავარიების რიცხვის ცხრილი ორივე შემთხვევაში:

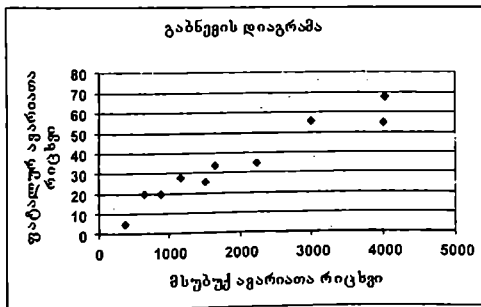
მსუბუქი ავარიების რიცხვი, (x)	376	650	884	1162	1513	1650	2236	3002	4028	4010
ფატალური ავარიების რიცხვი, (y)	5	20	20	28	26	34	35	56	68	55

აევათ გაბნევის დიაგრამა ამ მონაცემებისათვის.

ამოხსნა:

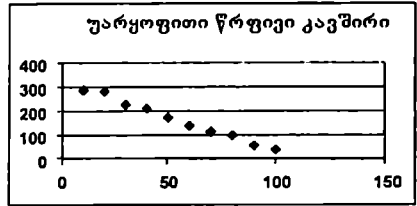
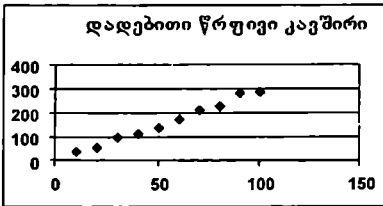
ნაბიჯი 1. სიბრტყეზე აევათ მართკუთხა კოორდინატთა სისტემა დამოუნიშნოთ x და y ღერძები.

ნაბიჯი 2. აევათ წერტილთა წყვილები საკოორდინატო სიბრტყეზე ისე, როგორც ეს ქვემოთაა ნაჩვენები:

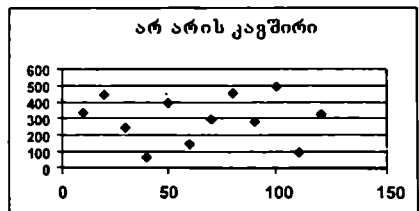
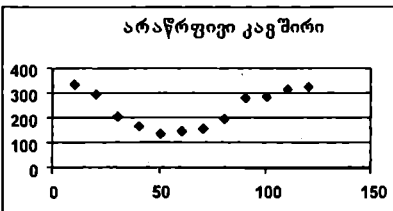


X და Y სიდიდეებს შორის შესაძლებელია არსებობდეს რამდენიმე განსხვავებული ტიპის დამოკიდებულება. ამ დამოკიდებულების გარკვევა (იდენტიფიცირება) შესაძლებელია საკოორდინატო სიბრტყეზე წერტილების განლაგების სტრუქტურაზე დაყრდნობით. ქვემოთ მოყვანილი იქნება წერტილების სტრუქტურული განლაგების სხვადასხვა ტიპი და შესაბამისი დამოკიდებულების სახეები.

1. **დადებითი წრფივი კავშირი** არსებობს (გვაქვს) იმ შემთხვევაში, როდესაც წერტილები დაახლოებით ლაგდება (თავმოყრილია, კონცენტრირებულია) აღმავალი სწორი ხაზის ირგვლივ (მიდამოში) და ერთდროულად ორივე X და Y სიდიდის მნიშვნელობები ზრდადია. ასეთი კავშირი გვაქვს მაშინ, როდესაც X სიდიდის ზრდასთან ერთად იზრდება Y სიდიდეც.
2. **უარყოფითი წრფივი კავშირი** არსებობს (გვაქვს) იმ შემთხვევაში, როდესაც წერტილები დაახლოებით ლაგდება (თავმოყრილია, კონცენტრირებულია) დაღმავალი სწორი ხაზის ირგვლივ. ასეთი კავშირი გვაქვს მაშინ, როდესაც X სიდიდის ზრდასთან ერთად კლებულობს Y სიდიდე, და პირიქით, X სიდიდის კლებასთან ერთად იზრდება Y სიდიდე.



3. **არანრფივი კავშირი** არსებობს (გვაქვს) იმ შემთხვევაში, როდესაც წერტილები დაახლოებით ლაგდება (თავმოყრილია, კონცენტრირებულია) არანრფივი (მრუდი) ხაზის (წირის) ირგვლივ. ეს დამოკიდებულება აღინერება შესაბამისი წირის ბუნებით.
4. **არ არის კავშირი** გვაქვს იმ შემთხვევაში, როდესაც წერტილები უნესრიგოდაა მიმოფანტული, არ ჩანს, რომ ისინი რაიმე წირის ირგვლივაა კონცენტრირებული.



ზემოთ მოყვანილ მაგალითში X (მსუბუქ ავარიათა რიცხვი) და Y (ფატალურ ავარიათა რიცხვი) სიდიდეებს შორის ადგილი აქვს წრფივ დადებით კავშირს (დამოკიდებულებას). სხვა სიტყვებით, როგორც კი სველ ასფალტზე ექვლოსიპედით მსუბუქ ავარიათა რიცხვი იწყებს ზრდას, მაშინ იმ ავარიების რიცხვიც მატულობს, სადაც დადგა სასიკვდილო შედეგი.

წირს, რომელიც აღწერს კავშირს (დამოკიდებულებას) X და Y სიდიდეების მნიშვნელობებს შორის, **მისადაგების წირი** ეწოდება. ამ წირის პოვნის საკითხებს სწავლობს მათემატიკური სტატისტიკის ნაწილი, რომელსაც **რეგრესიული ანალიზი** ეწოდება. აქ მხოლოდ შევნიშნავთ, რომ წირის მოსაძებნად იყენებენ ე. წ. **უმცირეს კვადრატთა მეთოდს**, რომლის თანახმადაც იძებნება ისეთი წირი $y = f(x)$, რომლისთვისაც სრულდება პირობა:

$$\text{გამოსახულება } \sum_{i=1}^n [y_i - f(x_i)]^2 \text{ აღწევს თავის მინიმუმს,}$$

ანუ წირი, რომლიდანაც დამოუკიდებელი X სიდიდის დაკვირვებული x_i მნიშვნელობების შესაბამისი დაკვირვებული y_i მნიშვნელობების თეორიული $f(x_i)$ მნიშვნელობებისაგან ვერტიკალური გადახრების კვადრატების ჯამი მინიმალურია.

თუ, მაგალითად, სიდიდეებს შორის გვაქვს წრფივი კავშირი, მაშინ, ბუნებრივია, მისადაგების წირი უნდა ვეძებოთ წრფივ ფუნქციებს შორის, ანუ $f(x) = a + bx$. წრფივი ფუნქციის მოძებნა ნიშნავს a და b კოეფიციენტების პოვნას. უმცირეს კვადრატთა მეთოდის თანახმად, უნდა ვიპოვოთ ისეთი a და b კოეფიციენტები, რომ გამოსახულებამ $-\sum_{i=1}^n (y_i - a - bx_i)^2$ მიიღოს თავისი მინიმალური მნიშვნელობა. ფუნქციის ექსტრემუმის პოვნის წესის თანახმად, აღნიშნული გამოსახულება უნდა გავანარმოოთ როგორც a -ს, ისე b -ს მიმართ. მიღებული წარმოებულები გავუტოლოთ ნულს და ამოვხსნათ მიღებულ განტოლებათა სისტემა a და b -ს მიმართ. საბოლოოდ, მივიღებთ, რომ:

$$a = \frac{(\sum_{i=1}^n y_i) \cdot (\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n x_i y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad \text{და} \quad b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}.$$

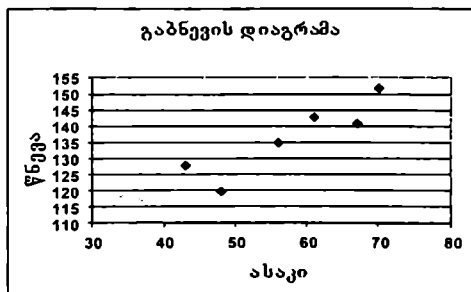
მაგალითი 5.3. ავადოთ გაბნევის დიაგრამა მონაცემებისათვის, რომელიც მიიღება 6 შემთხვევით შერჩეული პიროვნების ასაკსა და არტერიულ წნევას შორის დამოკიდებულების შესწავლისას. მონაცემების ცხრილი ასეთია:

პერსონა	ასაკი, x	წნევა, y
A	43	128
B	48	120
C	56	135
D	61	143
E	67	141
F	70	152

ამოხსნა:

ნაბიჯი 1. დაეხაზოთ საკოორდინატო სიბრტყე და მოენიშნოთ x და y ღერძები.

ნაბიჯი 2. მოენიშნოთ თითოეული წერტილი გრაფიკზე, როგორც ეს ქვემოთაა ნაჩვენები:



კორელაციის კოეფიციენტი

კორელაცია (წრფივი კავშირი). როცა ხელთ გვაქვს დანყვილებული მონაცემები და აგებულია მათი გაბნევის დიაგრამა, ბუნებრივია, დაისმის კითხვა: „რა ტიპისაა ეს კავშირი და რამდენად ძლიერია ის?“. სიტყვა „კორელაცია“ ნიშნავს, რომ ცვლადებს შორის არსებობს არა მიზეზობრივი, არამედ სტატისტიკური კავშირი, როცა ერთი ცვლადის მნიშვნელობების ზრდა იწვევს მეორე ცვლადის მნიშვნელობების შეთანხმებული ზრდის ან კლების ტენდენციას.

კორელაციის კოეფიციენტი (correlation coefficient) – წარმოადგენს სტატისტიკას, რომელიც ზომავს ცვლადებს შორის წრფივი დამოკიდებულების ხარისხს და ამ კავშირის მიმართულებას შერჩევითი მონაცემების მიხედვით. არსებობს სხვადასხვა სახის კორელაციის კოეფიციენტი. ყველაზე უფრო ცნობილია პირსონის კორელაციის კოეფიციენტი, რომელიც შემოიღო ბრიტანელმა სტატისტიკოსმა კარლ პირსონმა (1857–1936). პირსონის შერჩევითი კორელაციის კოეფიციენტი აღინიშნება r ასოთი, ხოლო პოპულაციის მიხედვით გამოთვლილი კორელაციის კოეფიციენტი აღინიშნება ბერძნული ρ (რო) ასოთი.

თავდაპირველად განვმარტოთ შერჩევითი კოვარიაციის კოეფიციენტი. ორ - X და Y სიდიდეს შორის კოვარიაციის კოეფიციენტი აღინიშნება $\text{cov}(X, Y)$ სიმბოლოთი და განიშარტება შემდეგი თანაფარდობით:

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

სადაც $\bar{x}$ და $\bar{y}$ არის, შესაბამისად, X და Y სიდიდეების შერჩევითი საშუალო ($\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$).

შერჩევითი კორელაციის კოეფიციენტი გამოითვლება შემდეგი ფორმულით:

$$r = \frac{\text{cov}(X, Y)}{s_x s_y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

კოვარიაციის კოეფიციენტის გამოსათვლელად მოხერხებულია ვისარგებლოთ შემდეგი ეკვივალენტური და გამარტივებული ფორმულით:

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y}.$$

თუ მოცემულია მონაცემთა ორი სიმრავლე $x_1, \dots, x_n$ და $y_1, \dots, y_n$, მაშინ მონაცემთა ახალი $z_i = x_i \pm y_i$ ($i = 1, \dots, n$) სიმრავლისათვის შერჩევითი დისპერსია ასე გამოითვლება:

$$s_z^2 = s_x^2 + s_y^2 \pm 2 \text{cov}(x, y)$$

(თუ $\text{cov}(x, y) = 0$, მაშინ $s_z^2 = s_x^2 + s_y^2$). ანალოგიურად:

$$s_z'^2 = s_x'^2 + s_y'^2 \pm 2 \text{cov}'(x, y),$$

სადაც

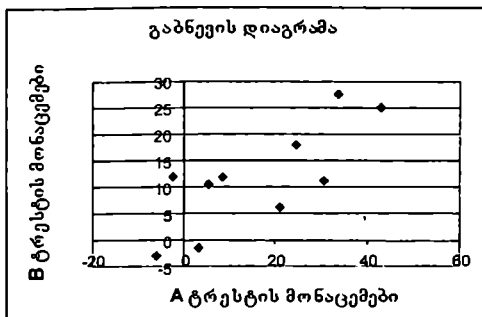
$$\text{cov}'(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n}{n-1} \text{cov}(x, y).$$

მაგალითი 5.4. მოცემულია ორი ტრესტის წლიური საპროცენტო შემოსავლის მონაცემები ბოლო 10 წლის განმავლობაში:

ტრესტი A	8.3	-6.2	20.9	-2.7	33.6	42.8	24.4	5.2	3.1	30.5
ტრესტი B	12.1	-2.8	6.4	12.2	27.8	25.3	18.2	10.7	-1.3	11.4

ავაგოთ გაბნევის დიაგრამა და გამოვთვალოთ კორელაციის კოეფიციენტი.

ამოხსნა. A ტრესტის მონაცემები გადავზომოთ x ლერძზე, B ტრესტისა კი – y ლერძზე.



ამ შემთხვევაში $\bar{x} = 16$, $\bar{y} = 12$, $s_x = 17.65$ და $s_y = 10.51$. ამიტომ გვაქვს:

$$\begin{aligned} \text{cov}(x, y) &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} = \\ &= \frac{1}{10} \cdot [8.3 \cdot 12.1 + (-6.2) \cdot (-2.8) + \dots + 30.5 \cdot 11.4] - 16 \cdot 12 = 71.48, \end{aligned}$$

შესაბამისად:

$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{71.48}{17.65 \cdot 10.51} \approx 0.39.$$

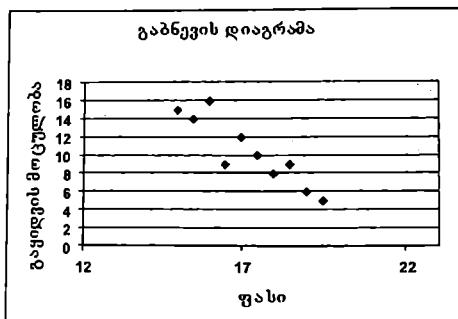
კორელაციის კოეფიციენტის ეს მნიშვნელობა გვარწმუნებს, რომ არსებობს სუსტად გამოხატული პოზიტიური წრფივი კავშირი A და B ტრესტების ამონაგებს შორის, რასაც გაბნევის დიაგრამაც ადასტურებს.

მაგალითი 5.5. ახალი პროდუქციის ფასის დასადგენად კომპანიამ შეარჩია 10 ქალაქი, რომელთაც, არსებითად, ერთნაირი ყიდვის პოტენციალი გააჩნია და გამოიტანა პროდუქცია ბაზარზე, ყოველ ქალაქში სხვადასხვა ფასად. კომპანიის მიერ მიღებული შედეგები მოყვანილია ცხრილში:

ქალაქის ნომერი	1	2	3	4	5	6	7	8	9	10
ფასი, x	15.0	15.5	16.0	16.5	17.0	17.5	18.0	18.5	19.0	19.5
გაყიდვის მოცულობა, y	15	14	16	9	12	10	8	9	6	5

დავადგინოთ გაყიდვის მოცულობის ფასზე დამოკიდებულების სახე.

ამოხსნა: ავავთ გაბნევის დიაგრამა



როგორც დიაგრამა გვიჩვენებს, ჩვენ წინაშეა უარყოფითი წრფივი კავშირის ტენდენცია: მონაცემები თავმოყრილია უარყოფითი დახრილობის წრფის მიდამოში. რადგანაც კორელაციის კოეფიციენტი ზომავს ორ სიდიდეს შორის არსებული წრფივი კავშირის სიძლიერეს, უნდა მოველოდეთ, რომ კორელაციის კოეფიციენტი საკმაოდ ახლოს იქნება -1-თან. გამოვთვალოთ კორელაციის კოეფიციენტი. გვაქვს:

$$\bar{x}=17.25, s_x=1.44, \bar{y}=10.4, s_y=3.56, \text{cov}(x,y)=-4.75 \text{ და}$$

$$r = \frac{\text{cov}(x,y)}{s_x s_y} = \frac{-4.75}{5.13} \approx -0.93.$$

კორელაციის კოეფიციენტის ეს მნიშვნელობა ადასტურებს გაბნევის დიაგრამის დახმარებით მიღებულ დასკვნას. მაგრამ ეს მნიშვნელობა არაფერს გვეუბნება იმ წრფის შესახებ, რომლის მიდამოშიც თავმოყრილია დაკვირვებული წყვილები. ამ წრფის მოსაძებნად იყენებენ უმცირეს კვადრატთა მეთოდს, რომლის თანახმადაც ეძებენ ისეთ წრფეს $y = \bar{a} + \bar{b}x$, რომლის კოეფიციენტებისთვისაც სრულდება პირობა:

$$\min_{a,b \in (-\infty, +\infty)} \sum_{i=1}^n (y_i - a - bx_i)^2 = \sum_{i=1}^n (y_i - \bar{a} - \bar{b}x_i)^2,$$

ანუ, წრფე რომლიდანაც y_i -ების ვერტიკალური გადახრების კვადრატების ჯამი მინიმალურია. ამ წრფის კოეფიციენტების მოსაძებნად უნდა ვიპოვოთ ორი ცვლადის $\varphi(a,b) = \sum_{i=1}^n (y_i - a - bx_i)^2$ ფუნქციის მინიმუმის ნერტილები, რისთვისაც, თავის მხრივ, უნდა ამოიხსნას განტოლებათა სისტემა:

$$\frac{\partial}{\partial a} \varphi(a,b) = 0 \text{ და } \frac{\partial}{\partial b} \varphi(a,b) = 0,$$

სადაც $\frac{\partial}{\partial a} \varphi(a, b)$ და $\frac{\partial}{\partial b} \varphi(a, b)$ სიმბოლოებით აღნიშნულია $\varphi(a, b)$ ფუნქციის

კერძო წარმოებულები, შესაბამისად, a -სა და b -ს მიმართ.

ძნელი დასანახი არ არის, რომ უკანასკნელი სისტემის ამონახსნია:

$$\bar{b} = \frac{\text{cov}(x, y)}{s_x^2} \quad \text{და} \quad \bar{a} = \bar{y} - \bar{b}\bar{x}.$$

ამ მაგალითში მოყვანილი მონაცემებისათვის გვექნება:

$$\bar{b} = \frac{-4.75}{1.44^2} = \frac{-4.75}{2.07} \approx -2.29 \quad \text{და} \quad \bar{a} = 10.4 + 2.29 \cdot 17.25 \approx 49.90.$$

ამიტომ საძებნი წრფის განტოლება იქნება: $y = -2.29x + 49.9$.

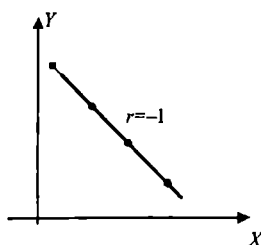
ეს განტოლება კომპანიას საშუალებას აძლევს ახალი პროდუქციის ამა თუ იმ ფასის შემთხვევაში გააკეთოს პროგნოზი გაყიდვის შესაძლო მოცულობის შესახებ.

კორელაციის კოეფიციენტის მნიშვნელობა მოთავსებულია -1 -სა და 1 -ს შორის. თუ ორ სიდიდეს შორის ადგილი აქვს მკაცრად დადებით წრფივ დამოკიდებულებას, მაშინ r -ის მნიშვნელობა ახლოსაა $+1$ -თან. თუ ორ სიდიდეს შორის ადგილი აქვს მკაცრად უარყოფით წრფივ დამოკიდებულებას, მაშინ r -ის მნიშვნელობა ახლოსაა -1 -თან. როდესაც ორ სიდიდეს შორის არა გვაქვს წრფივი დამოკიდებულება, ან ამ სიდიდეებს შორის სუსტი კავშირია, მაშინ r -ის მნიშვნელობა ახლოსაა 0 -თან. r -ის მცირე სიდიდე ყოველთვის არ ნიშნავს, რომ ცვლადებს შორის არ არის კავშირი. ცვლადებს შორის შეიძლება არ იყოს წრფივი კავშირი, მაგრამ იყოს ზუსტი არაწრფივი კავშირი.

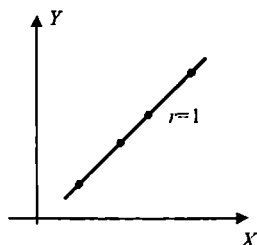
პირსონის კოეფიციენტის გამოთვლა არ ღირს, თუ ცვლადებს შორის არ არის მოსალოდნელი წრფივი კავშირი, ან როცა ცვლადები გაზომილია სახელდების ან რიგის სკალაზე.

ქვემოთ მოყვანილია კორელაციის ხარისხის სიტყვიერი დახასიათებისათვის მიღებული ტერმინოლოგია და შესაბამისი დიაგრამები:

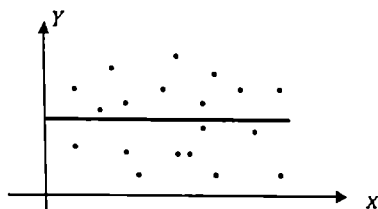
სრულყოფილი უარყოფითი კორელაცია	საშუალო უარყოფითი კორელაცია	არ არის კორელაცია	საშუალო დადებითი კორელაცია	სრულყოფილი დადებითი კორელაცია
ძლიერი უარყოფითი კორელაცია	სუსტი უარყოფითი კორელაცია		სუსტი დადებითი კორელაცია	ძლიერი დადებითი კორელაცია
-1	-0.5	0	0.5	1
უარყოფითი კორელაცია			დადებითი კორელაცია	



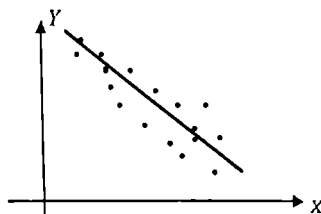
სრულყოფილი უარყოფითი
კორელაცია



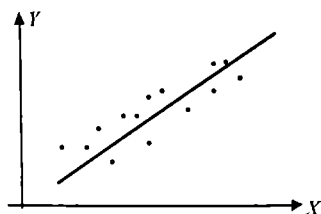
სრულყოფილი დადებითი
კორელაცია



არ არის კორელაცია



ძლიერი უარყოფითი
კორელაცია



ძლიერი დადებითი
კორელაცია

კორელაციის კოეფიციენტის პირდაპირი გამოთვლისათვის (როცა წინასწარ არ არის გამოთვლილი სიდიდეების საშუალო და სტანდარტული გადახრა) იყენებენ შემდეგ ფორმულას:

$$r = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{[n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2] \cdot [n(\sum_{i=1}^n y_i^2) - (\sum_{i=1}^n y_i)^2]}}$$

სადაც n მონაცემთა წყვილების რაოდენობაა.

ეს ფორმულა საკმაოდ რთულად გამოიყურება. გამოთვლების გამარტივების მიზნით ჩვენ ვიხელმძღვანელებთ ქვემოთ მოყვანილი ცხრილით.

მაგალითი 5.6. გამოვთვალოთ კორელაციის კოეფიციენტის მნიშვნელობა იმ მონაცემებისათვის, რომელიც მიღებული იყო 6 შემთხვევით შერჩეულ პიროვნების ასაკსა და არტერიულ წნევას შორის დამოკიდებულების შესწავლისას.

ამოხსნა:

ნაბიჯი 1. გავაკეთოთ ცხრილი ისე როგორც ეს ქვემოთაა ნაჩვენები:

ნაბიჯი 2. გამოვთვალოთ xy , x^2 და y^2 გამოსახულებათა მნიშვნელობები და შევიტანოთ ცხრილის შესაბამის სვეტებში. შევკრიბოთ სვეტებში მოთავსებული მნიშვნელობები. დასრულებულ ცხრილს ექნება შემდეგი სახე:

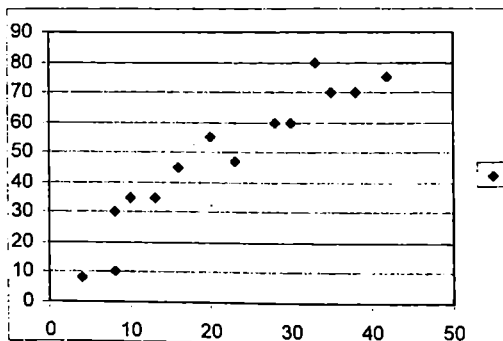
პერსონა	ასაკი, x	წნევა, y	xy	x^2	y^2
A	43	128	5.504	1.849	16.384
B	48	120	5.760	2.304	14.400
C	56	135	7.560	3.136	18.225
D	61	143	8.723	3.721	20.449
E	67	141	9.447	4.489	19.881
F	70	152	10.640	4.900	23.104
	$\sum x = 345$	$\sum y = 819$	$\sum xy = 47.634$	$\sum x^2 = 20.399$	$\sum y^2 = 112.443$

ნაბიჯი 3. მიღებული მნიშვნელობები შევიტანოთ კორელაციის კოეფიციენტის ფორმულაში და გამოვთვალოთ r :

$$r = \frac{6 \cdot 47.634 - 345 \cdot 819}{\sqrt{(6 \cdot 20.399 - 345^2) \cdot (6 \cdot 112.443 - 819^2)}} = 0.897.$$

დასკვნა. კორელაციის კოეფიციენტის მიღებული მნიშვნელობა გვეუბნება, რომ ადამიანის ასაკსა და არტერიულ წნევას შორის არსებობს მკაცრად დადებითი წრფივი დამოკიდებულება.

გამოვთვალოთ კორელაციის კოეფიციენტი 5.1 მაგალითის მონაცემებისათვის. აქ მისადაგების წირს აქვს შემდეგი სახე:



x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
4	8	-18	-41	738	324	1681
8	10	-14	-39	546	196	1521
8	32	-14	-17	238	196	289
10	35	-12	-14	168	144	196
13	37	-9	-12	108	81	144
16	47	-6	-2	12	36	4
20	47	-2	-2	4	4	4
23	55	1	6	6	1	36
28	60	6	11	66	36	121
30	60	8	11	88	64	121
33	80	11	31	341	121	961
35	70	13	21	273	169	441
38	70	16	21	336	256	441
42	75	20	26	520	400	676
$\bar{x} = 22$	$\bar{y} = 49$			$\sum = 3444$	$\sum = 2028$	$\sum = 6636$

$$\bar{x} = 22, \quad \bar{y} = 49, \quad \sum (x - \bar{x})(y - \bar{y}) = 3444,$$

$$\sum (x - \bar{x})^2 = 2028, \quad \sum (y - \bar{y})^2 = 6636.$$

$$r = 3444 / \sqrt{2028 \cdot 6636} = 0.939.$$

დასკვნა. ცვლადებს შორის აღინიშნება ძლიერი დადებითი კორელაცია.

სპირმენის რანგობრივი კორელაციის კოეფიციენტი

თუ მონაცემები მიღებულია რიგის სკალაზე, ან გაზომილია ინტერვალის სკალაზე, მაგრამ გაბნევის დიაგრამა გვიჩვენებს, რომ მათ შორის კავშირი არ არის წრფივი, ცვლადებს შორის კავშირის დასადგენად უნდა გამოვიყენოთ რანგული კორელაციის კოეფიციენტი. არსებობს ძალიან ბევრი რანგული კორელაციის კოეფიციენტი. ჩვენ განვიხილავთ სპირმენის რანგობრივი კორელაციის კოეფიციენტს, რომელიც მოიცემა ფორმულით:

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_A - R_B)^2,$$

სადაც n – შერჩევის მოცულობაა, ხოლო R_A და R_B – ე.წ. რანგებია.

განვიხილოთ ინდივიდთა ან ობიექტთა ერთობლიობა, რომლის ყველა წევრი ხასითდება რაიმე ორი ნიშნით. ვიგულისხმობთ, რომ ორი ნიშნით დახასიათების შედეგები ჩანერილია შემდეგი სახით:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix},$$

სადაც x_i და y_i არის i -ური ობიექტის დამახასიათებელი ნიშნები. ამ ნიშნებს შეიძლება არც ჰქონდეთ ზუსტი რიცხობრივი მნიშვნელობები, მაგრამ შესაძლებელი იყოს მათი ისე დალაგება, რომ თითოეული ნიშნის მიხედვით ყოველ ობიექტს მიეცეს რიგითი ნომერი – რანგი.

მაგალითი 5.7. ქვემოთ მოყვანილია სხვადასხვა კომპანიის მიერ გამოშვებული ერთი და იმავე სახის პროდუქციის რანჟირების შედეგები, ჩატარებული ორი A და B ექსპერტის მიერ. პირველ სვეტში მოცემულია შესამონშებელი პროდუქციის პირობითი ნომერი, ხოლო II და III სვეტში A და B ექსპერტების შეფასებები. მაგალითად, A ექსპერტი საუკეთესოდ მიიჩნევს პროდუქციის 3 ნიმუშს, მაშინ, როცა B ექსპერტი მას მეოთხედ ადგილს მიანერს. თუ რამდენიმე ნიმუშს ერთნაირად აფასებენ, მაშინ თითოეულს მიენერება მინერილი რანგების საშუალო არითმეტიკული. მაგალითად, A ექსპერტი ერთნაირად აფასებს პროდუქციის 5 და 7 ნიმუშებს, ისინი იყოფენ მეორე-მესამე ადგილს, ამიტომ ორივეს მიენერება $(2+3)/2 = 2.5$ -ის ტოლი რანგი. IV სვეტში A ექსპერტის რანგები დალაგებულია ზრდის მიხედვით, ხოლო მომდევნო V სვეტში მას მინერილი აქვს პროდუქციის იმავე ნიმუშისათვის B ექსპერტის მიერ მიკუთვნებული რანგი.

პროდუქციის პირობითი ნომერი	A ექსპერტის შეფასება	B ექსპერტის შეფასება	R_A	R_B	$R_A - R_B$	$(R_A - R_B)^2$
1	4	8	1	4	-3	9
2	9	7	2.5	5	-2.5	6.25
3	1	4	2.5	1	1.5	2.25
4	7	6	4	8	-4	16
5	2.5	5	5	3	2	4
6	8	10	6	2	4	16
7	2.5	1	7	6	1	1
8	6	2	8	10	-2	4
9	5	3	9	7	2	4
10	10	9	10	9	1	1
						$\Sigma = 63.5$

იმისათვის, რომ შევისწავლოთ A და B ექსპერტების კრიტერიუმებს შორის ურთიერთდამოკიდებულება, გამოვიყენოთ სპირმენის რანგობრივი კორელაციის კოეფიციენტი:

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_A - R_B)^2.$$

თუ ექსპერტების მიერ გამოყენებულ კრიტერიუმებს შორის სრული თანხვედრაა, მაშინ $\sum_{i=1}^n (R_A - R_B)^2 = 0$ და $r_s = 1$; თუ უარყოფითი კავშირია, მაშინ $r_s = -1$ და თუ კავშირი არ არსებობს, მაშინ $r_s = 0$.

$$\text{განხილულ მაგალითში } r_s = 1 - \frac{6}{10 \cdot (10^2 - 1)} \cdot 63.5 = 1 - \frac{381}{990} = \frac{609}{990} \approx 0.62 .$$

დასკვნა. ექსპერტების შეფასებებს შორის ძლიერი დადებითი კავშირია.

სპირმენის რანგობრივი კორელაციის კოეფიციენტი გამოიყენება შემდეგი პიპოთეზების შესამოწმებლად:

H_0 : ნიშნები დამოუკიდებელია, ე. ი. $r_s = 0$,

H_1 : ნიშნებს შორის დადებითი კავშირია, ე. ი. $r_s > 0$.

ცნობილია, რომ დიდი მოცულობის შემთხვევაში სტატისტიკა

$$t = \sqrt{n-2} / \sqrt{1-r_s^2}$$

განაწილებულია ე.წ. სტიუდენტის კანონით, თავისუფლების ხარისხით $n-2$. ამიტომ კრიტერიუმი ასე ყალიბდება:

თუ t სტატისტიკის დაკვირვებული t_g მნიშვნელობისათვის და მოცემული α მნიშვნელოვნების დონისათვის სრულდება პირობა

$$t_g > t_{n-2, \alpha} ,$$

მაშინ H_0 პიპოთეზას უარყოფთ, წინააღმდეგ შემთხვევაში ამის საფუძველი არა გვაქვს.

გასაგებია, რომ ორმხრივი ალტერნატივის შემთხვევაში - $H_1: r_s \neq 0$, კრიტიკული არე იქნება: $U_1 = \{ |t| > t_{n-2, \alpha/2} \}$.

მაგალითი 5.8. დღისა და საღამოს სწავლების ეკონომისტ სტიუდენტებს ეკითხებოდნენ, თუ როგორ აფასებენ ისინი სხვადასხვა პროფესიის პრესტიჟს მბალანსი სისტემით. მიიღეს შემდეგი ცხრილი:

პროფესია	დღის	საღამოს
ბუღალტერი	6	3
პროგრამისტი	7	2
ბანკის მენეჯერი	2	6
ადმინისტრატორი	5	4
სტატისტიკოსი	1	7
მარკეტინგის მენეჯერი	4	8
ფინანსისტ-ანალიტიკოსი	3	5
წარმოების მენეჯერი	8	1

იპოვეთ კორელაციის კოეფიციენტი.

ამოხსნა: ამ შემთხვევაში გვაქვს პროფესიებისაგან შედგენილი ერთობლიობა, ხოლო ნიშნების როლში განიხილება დღისა და საღამოს სწავლების სტუდენტთა მიერ ამ პროფესიების შეფასებები. შევადგინოთ შემდეგი ცხრილი:

პროფესია	დღის	სალამოს	$R_A - R_R$	$(R_A - R_R)^2$
ბუღალტერი	6	3	3	9
პროგრამისტი	7	2	5	25
ბანკის მენეჯერი	2	6	-4	16
ადმინისტრატორი	5	4	1	1
სტატისტიკოსი	1	7	-6	36
მარკეტინგის მენეჯერი	4	8	-4	16
ფინანსისტი-ანალიტიკოსი	3	5	-2	4
წარმოების მენეჯერი	8	1	7	49
Σ				156

$$r_s = 1 - \frac{6 \cdot 156}{8 \cdot (8^2 - 1)} = -0.857.$$

დასკვნა. შეფასებებს შორის ძლიერი უარყოფითი კავშირია.

მაგალითი 5.9. არის თუ არა კავშირი ასაკსა და სტიმულზე რეაქციის დროს (წმ) შორის?

ცდისპირი	ასაკი, x	რეაქციის დრო, y	რანგი, R_x	რანგი, R_y	$R_x - R_y$	$(R_x - R_y)^2$
1	18	0.75	1	2	-1	1
2	20	1.0	3	5	-2	4
3	23	1.05	6	6	0	0
4	19	0.5	2	1	1	1
5	27	0.9	9	4	5	25
6	32	1.1	10	7	3	9
7	21	1.5	4	8	-4	16
8	26	0.8	8	3	5	25
9	24	1.8	7	10	-3	9
10	22	1.6	5	9	-4	16
						106

$$r_s = 1 - \frac{6 \cdot 106}{10(10^2 - 1)} = 1 - \frac{636}{990} \approx 0.36.$$

დასკვნა. ასაკსა და რეაქციის დროს რანგებს შორის სუსტი დადებითი კავშირია.

რეგრესია

წირს, რომელიც აღწერს კავშირს ცვლადების მნიშვნელობებს შორის *მიხა-დაცების წირი* ეწოდება. ამ წირის პოვნის საკითხებს სწავლობს სტატისტიკის ნაწილი, რომელსაც *რეგრესიული ანალიზი* ეწოდება. წრფივი რეგრესია – არის

ტექნიკა, რომელიც გვაძლევს საშუალებას ფორმულის სახით გამოვსახოთ კავშირი X და Y ცვლადებს შორის. მოითხოვება, რომ ორივე ცვლადი იყოს რაოდენობრივი.

რეგრესიის წრფის განტოლება მოიცემა ფორმულით: $y = a + bx$, სადაც a და b – მუდმივი რიცხვებია. რეგრესიის წრფის ასაგებად პირველ რიგში უნდა ავაგოთ გაბნევის დიაგრამა. რეგრესიის წრფის აგება ღირს იმ შემთხვევაში, თუ წერტილები გაბნევის დიაგრამაზე კონცენტრირებულია რომელიმე წრფის მიდამოში. ჩვენ შეგვიძლია ავაგოთ უამრავი ასეთი წრფე. რომელია საუკეთესო?

განვიხილოთ მაგალითი. ვთქვათ, შევაგროვეთ შემდეგი მონაცემები:

X	1	2	3	4
Y	1.5	3.4	3.7	5.4

განვიხილოთ სამი წრფე: $A: y = 1.2x + 0.5$, $B: y = x + 1$ და $C: y = 3.5$. გამოვთვალოთ, რას უდრის y -ების მნიშვნელობები, თუ ჩავსვამთ x -ების მნიშვნელობებს ამ სამ განტოლებაში:

X	1	2	3	4
A	1.7	2.9	4.1	5.3
B	2	3	4	5
C	3.5	3.5	3.5	3.5

გამოვთვალოთ, რას უდრის სხვაობა დაკვირვებულ და გამოთვლილ y -ებს შორის სამივე წრფისათვის. ამ სხვაობას ეწოდება ნაშთი ან შეცდომა.

მონაცემები		ნაშთი (შეცდომა)			შეცდომა კვადრატში		
X	Y	A	B	C	A	B	C
1	1.5	-0.2	-0.5	-2.0	0.04	0.25	4.00
2	3.4	0.5	0.4	-0.1	0.25	0.16	0.01
3	3.7	-0.4	-0.3	0.2	0.16	0.09	0.04
4	5.4	0.1	0.4	1.9	0.01	0.16	3.61
ჯამი		0	0	0	0.46	0.66	7.66

საუკეთესო წრფისათვის ამ ნაშთების კვადრატების ჯამი უნდა იყოს მინიმალური და შესაბამის წრფეს უმცირეს კვადრატთა მეთოდით დადგენილი რეგრესიის წრფე ეწოდება.

ორი ცვლადის უმცირეს კვადრატთა რეგრესიის წრფის a და b კოეფიციენტები გამოითვლება ფორმულებით:

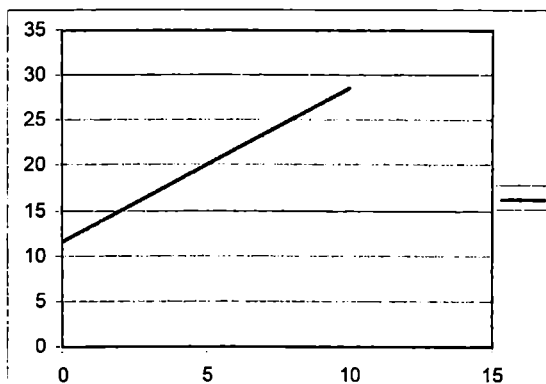
$$b = \frac{\text{cov}(X, Y)}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}.$$

ავაგოთ რეგრესიის წრფე 5.6 მაგალითის მონაცემებისათვის. ზემოთ ჩატარებული გამოთვლების თანახმად:

$$b = \frac{3444}{2028} \approx 1.7 \quad \text{და} \quad a = 49 - 22 \cdot 1.7 = 11.6.$$

შესაბამისად, რეგრესიის წრფის განტოლება იქნება:

$$y = 1.7x + 11.6.$$



რას გვიჩვენებს b კოეფიციენტი? b გვიჩვენებს რამდენი ერთეულით შეიცვლება Y , თუ X -ს შევცვლით ერთი ერთეულით.

განვიხილოთ ისევ განტოლება $y = 1.7x + 11.6$. ვთქვათ, $x_1 = 10$, მაშინ $y_1 = 28.6$, შევცვალოთ X ერთი ერთეულით ანუ ავიღოთ $x_2 = x_1 + 1 = 11$, მაშინ $y_2 = 30.3$ და $y_2 - y_1 = 30.3 - 28.6 = 1.7 = b$. რას გვიჩვენებს a კოეფიციენტი? a გვიჩვენებს Y -ის საპროგნოზო მნიშვნელობას, როცა $X = 0$.

არსებობს პირსონის კორელაციის კოეფიციენტის გამოთვლის სხვა მეთოდი, კერძოდ:

$$r = \frac{s_x}{s_y} \cdot b,$$

სადაც s_x - X -ის სტანდარტული გადახრაა და s_y - Y -ის სტანდარტული გადახრა. r გამოითვლება b კოეფიციენტის გამრავლებით სტანდარტული გადახრების შეფარდებაზე. ამის გამო კომპიუტერულ პროგრამებში პირსონის კოეფიციენტს ხშირად აღნიშნავენ სიტყვით „ეტა“ (b არის პოპულაციის β („ბეტა“) კოეფიციენტის ნერტილოვანი შეფასება).

წინასწარმეტყველება (პროგნოზირება)

რეგრესიის წრფე შეიძლება გამოყენებულ იქნეს წინასწარმეტყველებისათვის. მაგალითად, გავაკეთოთ პროგნოზი: რას შეიძლება უდრიდეს ბიოლოგიის ქულა, თუ მათემატიკაში მიღებულია ქულა 25. განტოლებაში $y = 1.7x + 11.6$ ჩავსვათ $x = 25$, მივიღებთ $y = 54.1$.

პროგნოზის გასაკეთებლად უნდა გვახსოვდეს:

1. თუ X და Y ცვლადები კორელაციაში არ არიან, X -ის ცოდნა არ გვაძლევს ინფორმაციას Y -ის შესახებ. Y -ის საუკეთესო პროგნოზი არის დაკვირვებული Y -ების საშუალო.

2. პროგნოზის გაკეთება არ შეიძლება იმ მნიშვნელობებისათვის, რომლებიც X შერჩევის საზღვრებს მიღმა არიან. აქ კავშირი შეიძლება იყოს შეცვლილი, მაგალითად, არანრფივი.

დეტერმინაციის კოეფიციენტი

X და Y ცვლადებს ახასიათებს საკუთარი ცვალებადობა. ამ ცვალებადობის საერთო ნილს ზომავს დეტერმინაციის კოეფიციენტი, რომელიც უდრის $r^2 \times 100\%$. კავშირი კორელაციის კოეფიციენტსა და რეგრესიის წრფეს შორის გამოიხატება იმაში, რომ, ერთის მხრივ, Y -ების ცვალებადობა რეგრესიის წრფის მეშვეობით აიხსნება X -ების ცვალებადობით და, მეორე მხრივ, დეტერმინაციის კოეფიციენტი ზომავს ამ ცვალებადობის საერთო ნაწილს.

თუ r^2 მცირეა, ე.ი. რეგრესიით ახსნილია სრული გაბნევის მცირე წილი, მაშინ მოსაძებნია სხვა ალტერნატიული მოდელი (მაგალითად, არანრფივი ან მრავლობითი რეგრესიის მოდელი), რომელიც უფრო ეფექტურად ახსნიდა Y ცვლადის დაკვირვებული მნიშვნელობების სრულ გაბნევას თავისი საშუალო მნიშვნელობების მიმართ.

შეჯამება

რეგრესია და კორელაცია შეისწავლის ორ ცვლადს შორის სტატისტიკურ კავშირს. კორელაციის კოეფიციენტი რიცხვია, რომელიც ახასიათებს წრფივი კავშირის სიმჭიდროვეს და მიმართულებას. რეგრესიის წრფე აიგება მხოლოდ რაოდენობრივი მონაცემებისათვის. რეგრესიის საუკეთესო წრფის განსაზღვრავად გამოიყენება უმცირეს კვადრატთა მეთოდი. ის მდგომარეობს კოეფიციენტების ისეთ მნიშვნელობათა განსაზღვრაში, რომლებიც უზრუნველყოფენ დაკვირვებულ და გამოთვლილ Y -ების სხვაობათა ჯამის მინიმალურობას (ნამთათა კვადრატების ჯამის მინიმალურობის თვალსაზრისით).

ძირითადი ტერმინები

დანეცილებული მონაცემები, კორელაცია, პირსონის კორელაციის კოეფიციენტი, სპირმენის რანგობრივი კორელაციის კოეფიციენტი, რეგრესიის წრფე, დეტერმინაციის კოეფიციენტი, ნაშთი.

კითხვები

1. არის თუ არა კავშირი ადამიანის ასაკსა და არტერიულ წნევას შორის?

ასაკი	42	48	56	64	70
წნევა	120	125	130	145	150

- ა. ააგეთ გაბნევის დიაგრამა;
 ბ. გამოთვალეთ პირსონის კორელაციის კოეფიციენტი;
 გ. ააგეთ რეგრესიის წრფე და მისი მეშვეობით გააკეთეთ პროგნოზი: საშუალოდ რა წნევას მოსალოდნელი 60 წლის ასაკისთვის. მიზანშეწონილია თუ არა პროგნოზის გაკეთება 75 წლის ადამიანისათვის?
2. ახალი პროდუქციის ფასის დასადგენად კომპანიამ შეარჩია 5 ერთნაირი მარკეტი და გასაყიდ პროდუქციას სხვადასხვა ფასი დაადო. ქვემოთ მოყვანილია ერთი დღის გაყიდვის მოცულობები:

ფასი	15	15.5	16	16.5	17
გაყიდვების მოცულობა	16	14	18	12	10

- ა. დაადგინეთ, არსებობს თუ არა კავშირი პროდუქციაზე დადებულ ფასსა და გაყიდვის მოცულობას შორის?
 ბ. ააგეთ რეგრესიის წრფე.
3. დამოკიდებულია თუ არა პარკინსონით დაავადების ხანგრძლივობაზე დაავადების სიმძიმის დონე. დაავადების სიმძიმე იზომებოდა რიგის სკალაზე, სადაც 1 – აღნიშნავდა უმნიშვნელო სიმძიმეს, 2 – საშუალოს, 3 – მნიშვნელოვანს, 4 – ძალიან მაღალს.

დაავადების ხანგრძლივობა (წელი)	6	1	3.5	1.5	2.5	18	11	8	3
დაავადების სიმძიმე	2	2	2	2	3	4	3	3	1

- ა. რომელი კორელაციის კოეფიციენტი უნდა გამოეთვალათ ამ კვლევაში? რატომ?
 ბ. დაადგინეთ, არის თუ არა კორელაციური კავშირი ამ ორ ცვლადს შორის.
4. ცხრილში მოყვანილია დისტრიბუტორების მიერ ერთ კვირაში მობილური ტელეფონით დარეკილი ზარების რაოდენობა და გაყიდული პროდუქციის რაოდენობა.

ზარების რაოდენობა	66	43	57	32	18	59	61	32
გაყიდვები	20	15	18	12	2	21	18	8

- ა. ააგეთ გაბნევის დიაგრამა და მიიღეთ გადაწყვეტილება, ღირს თუ არა რეგრესიის წრფის აგება;
 ბ. ააგეთ რეგრესიის წრფე;
 გ. გააკეთეთ პროგნოზი: საშუალოდ რა იქნება გაყიდვების მოცულობა, თუ დაფიქსირდება 20 ზარი? 60 ზარი?
 დ. გამოთვალეთ 18 და 32 ზარების შესაბამისი ნაშთები და მოახდინეთ ამ რიცხვების ინტერპრეტაცია.

5. მოქმედებს თუ არა რეკლამა პროდუქციის გასაღების დონეზე? ცხრილში მოყვანილია 5 ფირმის რეკლამაზე დახარჯული თანხა X (ათასი ლარი) და გასაღების დონე Y (ათასი ლარი).

X	2	4	5	7	3
Y	10	40	30	50	20

- ა. ააგეთ გაბნევის დიაგრამა;
 ბ. გამოთვალეთ კორელაციის კოეფიციენტი და მიეცით მას ინტერპრეტაცია;
 გ. გამოთვალეთ დეტერმინაციის კოეფიციენტი და მიეცით მას ინტერპრეტაცია;
 დ. ააგეთ რეგრესიის წრფე და მისი მეშვეობით შეაფასეთ საშუალოდ რამდენ ათას ლარს შეადგენს გასაღების დონე, თუ რეკლამაზე დახარჯავენ 6000 ლარს.
6. მძიმე მეტალების დონეს ტბებში მიიჩნევენ წყლის ხარისხის მნიშვნელოვან მაჩვენებლად. 10 ტბაში გაზომილი იყო წყალში სტრონციუმის პროცენტული შემადგენლობა და იმავე დღეს დაჭერილი ერთი და იმავე ჯიშის თევზის საშუალო წონა:

სტრონციუმი %	4.0	5.5	4.5	4.3	6.0	5.7	4.5	4.0
თევზის წონა (კგ)	1.4	1.7	2.0	1.3	1.8	2.2	1.6	1.6

- ა. ააგეთ გაბნევის დიაგრამა;
 ბ. გამოთვალეთ კორელაციის კოეფიციენტი;
 გ. გამოთვალეთ დეტერმინაციის კოეფიციენტი;
 დ. როგორ შეაფასებთ მიღებულ კავშირს?
7. ფილტვების ფუნქციონირების სტანდარტულ ზომად იყენებენ FEV სიდიდეს. ეს სიდიდე, როგორც წესი, დამოკიდებულია ადამიანის სიმაღლეზე და ასაკზე. იმისათვის, რომ გამოეკვლიათ მხოლოდ სიმაღლეზე დამოკიდებულების საკითხი, შეარჩიეს ათი წლის ბიჭები და მიიღეს ქვემოთ მოყვანილი მონაცემები:

სიმაღლე (სმ)	134	138	142	146	150	154
FEV სიდიდე	1.7	1.9	2.0	2.1	2.3	2.6

- ა. ააგეთ რეგრესიის წრფე;
 ბ. საშუალოდ, რისი ტოლი იქნება FEV სიდიდე, თუ სიმაღლე 140-ის ტოლია;
 გ. გამართლებულია თუ არა პროგნოზის გაკეთება, თუ ბიჭის სიმაღლე 160 სმ-ია?
8. არის თუ არა საბავშვო ბაღის აღსაზრდელ გოგონებში ფიზიკური მომზიბვლელობა და პოპულარობა კორელაციაში? ბიჭებში? ორივე ეს ცვლადი იზომებოდა რიგის სკალებზე

გოგონები

მომზიბვლელობა	3	6	2	9	12	1	7	8	4	10	5	11
პოპულარობა	1	7	3	10	11	3	8	7	5	9	6	12

ბიჭები

მომზიბვლელობა	5	11	3	9	4	8	12	1	14	10	6	13	2	7
პოპულარობა	9	1	10	4	6	12	5	13	8	14	2	11	7	3

- ა. კორელაციის რომელ კოეფიციენტს გამოთვლით? რატომ?
- ბ. რა დასკვნას გააკეთებთ გოგონებში ამ ორი ცვლადის კავშირის შესახებ? ბიჭებში – ამ ორი ცვლადის კავშირის შესახებ?
9. რა გავლენას ახდენს კომპიუტერის ეკრანის ხანძრლივად ყურება ალქმის უნარზე? გაზომილი იყო ცდისპირის მიერ კომპიუტერთან გატარებული საათების რაოდენობა და იმ შეცდომების სიდიდეები, რასაც ცდისპირი უშვებდა სილრმის ალქმის ტესტში.

კომპიუტერთან მუშაობის ხანგრძლივობა (სთ)	0.5	1.7	4.2	6.0	2.8	5.3	0.9	3.4	1.4	5.1	1.9	4.8	2.6
სილრმის ალქმის შეცდომა	4.3	3.1	2.1	0.4	6.3	3.7	1.6	1.2	2.4	5.2	0.6	4.1	5.4

მონაცემების დამუშავების შედეგად რა დასკვნას აკეთებთ კომპიუტერთან მუშაობის საათებსა და სილრმის ალქმის შეცდომის კავშირის შესახებ?

10. ცხრილში მოყვანილია ოჯახში ერთ კვირაში შემოსული ზარების რაოდენობა და ოჯახის წევრების რაოდენობა:

ზარების რაოდენობა	37	23	67	21	10	33	46	67	44	38
წევრების რაოდენობა	4	2	3	1	2	3	4	5	3	4

- ა. ააგეთ გაბნევის დიაგრამა;
- ბ. გამოთვალეთ პირსონის კორელაციის კოეფიციენტი;
- გ. ააგეთ რეგრესიის განტოლება.
11. არსებობს თუ არა კავშირი აბიტურიენტის SAT ქულასა და GPA ქულას შორის? 10 სტუდენტის ქულებია:

GPA (Y)	2.5	3.2	3.5	2.8	3.0	2.4	3.4	2.9	2.7	3.8
SAT (X)	640	700	550	540	620	490	710	600	505	710

- ა. ააგეთ გაბნევის დიაგრამა;
- ბ. არსებობს თუ არა კავშირი ამ ორ ცვლადს შორის?
- გ. ააგეთ რეგრესიის განტოლება. რამდენით შეიცვლება GPA -ს საპროგნოზო მნიშვნელობა, თუ SAT ქულას შევცვლით ერთი ერთეულით?
12. ფეხბურთის გუნდის მენეჯერმა დათვალა ერთი წლის განმავლობაში რამდენი დღე (X) იყო ფეხბურთელი დაზიანების გარეშე და ჩანერა ამ ფეხბურთელის წონა Y (ფუნტებში):

X	17	9	0	3	26	18	2	8	19	38	20	0	2
Y	221	198	234	256	278	197	206	224	234	278	220	199	234

არსებობს, თუ არა კავშირი ამ ორ სიდიდეს შორის?

13. ქვემოთ მოცემულია ამონაწერი SPSS პროგრამიდან. გასარკვევი იყო არსებობს თუ არა კავშირი ბინის გასაყიდ ფასსა და სახლის ზომას შორის.

Part of a SPSS Computer Printout for Regression Analysis of Seling Price and Sise of House

Variable	N	Mean	Std. Dev
PRICE	93	99.533	44.184
SISE	93	1650	0.525

Variable	B	Beta	R Square
SISE	75.6068	0.8988	0.8079
Constant	-25.1936		

„B“ – აღნიშნავს შესაფასებელ რეგრესიის კოეფიციენტს b -ს. Constant აღნიშნავს a კოეფიციენტს. Beta – პირსონის კორელაციის კოეფიციენტი. R Square – დეტერმინაციის კოეფიციენტი. განხილული იყო 93 ბინის ფასი (საშუალო ფასი 99533 დოლარი, სტანდარტული გადახრა 44 184 დოლარი) და ამ ბინების ფართობები საშუალოთი 1 650 კვ.მ და სტანდარტული გადახრით 0.525 კვ.მ.

- ამოწერეთ რეგრესიის განტოლება;
 - საშუალოდ რამდენი დოლარით გაიზრდება ბინის ფასი, თუ ფართობი გაიზრდება 1 ათასი კვ.მ-ით?
14. დამოკიდებულია თუ არა შვილების რაოდენობა დედის განათლებაზე? ტეხასში გამოკითხული იყო 49 კათოლიკე დედა: X – სწავლის წლების რაოდენობა, Y – ბავშვების რაოდენობა. მიღებული იყო: $\bar{x} = 9.88$, $s_x = 3.77$, $\bar{y} = 3.35$, $s_y = 2.19$. საპროგნოზო განტოლება: $y = -0.207x + 5.40$.
- საშუალოდ რამდენი შვილი ჰყავთ დედებს, რომელთაც არ მიუღიათ განათლება?
 - იპოვეთ შვილების საპროგნოზო რაოდენობა, თუ დედას აქვს 8-წლიანი განათლება? 16-წლიანი განათლება?
 - რას უდრის პირსონის კორელაციის კოეფიციენტი?
15. გამოიყენეთ სპირმენის კორელაციის კოეფიციენტი ორ ცვლადს შორის კავშირის დასადგენად. X – რელიგიურობა, Y – დამოკიდებულება სამუშაო კოლექტივის მიმართ. რელიგიურობა იზომებოდა რიგის სკალაზე (ეკლესიაში სიარული) კატეგორიებით: არასდროს (0), წელიწადში ერთხელ (1), წელიწადში ორჯერ (2), წელიწადში რამდენჯერმე (3), თვეში ერთხელ (4), თვეში 2-3 ჯერ (5), კვირაში ერთხელ (6), ყოველ დღე (7), დღეში ორჯერ (8). დამოკიდებულება კოლექტივის მიმართ იზომებოდა 100-ქულიან სკალაზე, სადაც უფრო დიდი ქულა უფრო კარგ განწყობას ნიშნავს.

რელიგიურობა	8	5	7	4	3	4	5	3	4	6
დამოკიდებულება	10	70	50	30	90	60	40	80	50	40

16. წამლის მოქმედების ეფექტურობის შესამოწმებლად ჩატარდა 20 ცდა (X წამლის დოზაა, ხოლო Y ავადმყოფობის მახასიათებელი):

X	8	6	5	1	0	2	4	3	8	10	9	9	5	1	3	2	6	4	8	5
Y	13	15	16	21	23	19	17	17	14	12	13	12	18	20	18	17	14	16	13	18

ააგეთ გაბნევის დიაგრამა, იპოვეთ კორელაციის კოეფიციენტი. მისი საშუალებით დაადგინეთ, როგორ კავშირში იმყოფებიან ეს ცვლადები. შეადგინეთ რეგრესიის წრფე და მისი საშუალებით გააკეთეთ პროგნოზი: რისი ტოლი იქნება ავადმყოფობის მახასიათებელი, თუ წამლის დოზა იქნება 7-ის ტოლი.

17. ლექციებზე აქტიურობასა და საგნის წარმატებით ათვისების შესამოწმებლად ჩატარდა 20 ცდა (X აქტიურობაა, ხოლო Y წარმატებულობა):

X	2	16	18	10	8	6	9	13	15	12	17	16	6	4	11	14	9	17	14	12
Y	4	17	17	11	7	5	11	12	13	13	16	17	4	5	10	12	11	15	15	13

ააგეთ გაბნევის დიაგრამა, იპოვეთ კორელაციის კოეფიციენტი მისი საშუალებით დაადგინეთ, როგორ კავშირში იმყოფებიან ეს ცვლადები. შეადგინეთ რეგრესიის წრფე და მისი საშუალებით გააკეთეთ პროგნოზი: რისი ტოლი იქნება წარმატებულობა, თუ აქტიურობა იქნება 20-ის ტოლი.

თავი 6

ალბათობა

ალბათობა გვაცხადებს საშუალებას, აღვწეროთ ამა თუ იმ მოვლენის განხორციელების შანსი. ინტუიციურად ყველას ესმის:

- რას უდრის მონეტის აგდებისას გერბის მოსვლის შანსი;
- რას უდრის კამათლის გაგორებისას ექვსიანის მოსვლის შანსი;
- ლატარეაში მოგების შანსი;
- მწველისათვის ფილტვის კიბოს აღმოჩენის შანსი.

ალბათ, ბევრი ადამიანი შეაფასებს გერბის მოსვლის შანსს, როგორც ორმოცდაათი ორმოცდაათზე, ექვსიანის მოსვლას, როგორც ერთი ექვსთან. ბევრს ესმის, რომ ლატარეაში მოგების შანსი ძალიან პატარაა, ზოგიერთ ლატარეაში უდრის 1:1 000 000-თან. თუ ადამიანი ენევა, უფრო მოსალოდნელია, რომ მას აღმოაჩნდეს ფილტვის კიბო.

რა არის ალბათობა?

ალბათობა გვეხმარება შემთხვევითი ცდის (მოვლენის) შედეგების აღწერაში. ჩვენ ვამბობთ, რომ ცდა შემთხვევითია, თუ ზუსტად არ ვიცით ის რა შედეგით დამთავრდება. მაგალითად, მონეტის აგდებისას მოვა გერბი თუ საფასური.

არსებობს ალბათობის განსაზღვრის რამდენიმე გზა:

- ✓ განსაზღვრეთ ალბათობა, როგორც ცდის შედეგის განხორციელების შანსის რიცხვითი მნიშვნელობა.
- ✓ როგორც ცდის შედეგის ფარდობითი სიხშირე ცდათა გრძელ სერიაში.

ალბათობა – რიცხვია, რომელიც შეესაბამება ცდის შედეგს. ცდის შედეგებს აღნიშნავენ ლათინური ასოებით: A , B , C , ... პროგნოზ უნდა გავიგოთ მტკიცებულება, რომ A -ს ალბათობა უდრის 0.25-ს? ცდათა გრძელ სერიაში ცდათა 25% დამთავრდება A შედეგით.

მაგალითი 6.1. ვაგდებთ სამ მონეტას. გერბის მოსვლა აღვნიშნოთ სიმბოლოთი – „გ“, საფასურის მოსვლა კი სიმბოლოთი – „ს“. ჩამოვწეროთ ამ ცდის ყველა შესაძლო შედეგი:

{გგგ, გგს, გსგ, სგგ, სსგ, სგს, გსს, სსს}.

თითოეულ შედეგს ეწოდება ელემენტარული ხდომილება. ხდომილება $A = \{\text{სსს}\}$ – ელემენტარული ხდომილებაა, რომელიც აღნიშნავს, რომ სამივე მონეტაზე მოვიდა საფასური. ხდომილება $B = \{\text{გგს, გსგ, სგგ}\}$ – ნიშნავს, რომ სამი მონეტისაგან დებისას მივიღებთ ორ გერბს და ერთ საფასურს და ეს ხდომილება შეიძლება განხორციელდეს მაშინ, თუ ცდა დამთავრდა ერთ-ერთი შედეგით სამიდან. ეს ხდომილება მოიცავს სამ ელემენტარულ ხდომილებას.

ვამბობთ, რომ განხორციელება ხდომილება A ან B , თუ ცდა დამთავრდება რომელიმე შედეგით A -დან ან B -დან:

$$(A \text{ ან } B) = \{სსს, გგს, გსგ, სგგ\}$$

ხდომილება $C = \{გგგ, გგს, გსგ, სგგ\}$ ნიშნავს, რომ ცდაში ორი გერბი მაინც მოვა. ხდომილებები A და C ერთდროულად ვერ მოხდება, რადგანაც მათი არცერთი შედეგი არ არის საერთო.

შეუძლებელია ხდომილება, თუ მისი განხორციელება შეუძლებელია.

შეუძლებელი ხდომილება არ განხორციელდება, მისი ფარდობითი სიხშირე ნულია. ეს კი ნიშნავს, რომ მისი განხორციელების ალბათობა ნულის ტოლია. თუ სპორტსმენი შეჯიბრებაში არ მონაწილეობს, მისი გამარჯვების ალბათობა ნულის ტოლია.

აუცილებელია ხდომილება, რომელიც აუცილებლად განხორციელდება. მისი ალბათობა ერთის ტოლია.

ორი ხდომილება უთავსებადია, თუ მათი ერთდროული მოხდენა შეუძლებელია.

A და B ხდომილებები უთავსებადია, რადგანაც შეუძლებელია სამი მონეტის აგდებისას ერთდროულად ორი გერბი და სამი საფასური მოვიდეს.

ალბათობის თვისებები

ალბათობა – ხდომილებასთან დაკავშირებული რიცხვია, რომელიც P სიმბოლოთი აღინიშნება (პირველი ასო ინგლისური შესატყვისისა – Probability).

თვისება 1. $0 \leq P(A) \leq 1$, რადგანაც ფარდობითი სიხშირე არ შეიძლება იყოს ნულზე ნაკლები და ერთზე მეტი;

თვისება 2. ცდის ყველა შესაძლო შედეგის (აუცილებელი ხდომილება) ალბათობის ჯამი უდრის ერთს, რადგან ყველა შესაძლო შედეგის ფარდობით სიხშირეთა ჯამი უდრის ერთს.

თვისება 3. $P(\bar{A}) = 1 - P(A)$ ($\bar{A}$ – ხდომილება: A არ მოხდა).

მაგალითად, თუ მიზანში მოხვედრის ალბათობა 20%-ია, მიზანში არმოხვედრის ალბათობა 80%-ია.

თვისება 4. თუ A და B უთავსებადი ხდომილებებია, მაშინ

$$P(A \text{ ან } B) \equiv P(A \cup B) = P(A) + P(B).$$

მაგალითად, თუ აუდიტორიაში სტუდენტების 15% პირველკურსელია და 10% მეორეკურსელია, მაშინ აუდიტორიაში სტუდენტების 25% პირველ ორ კურსზე სწავლობს. საზოგადოდ, გვაქვს **ჯამის ალბათობის** შემდეგი ფორმულა:

$$P(A \text{ ან } B) \equiv P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

თვისება 5. ხდომილების ალბათობა უდრის ამ ხდომილების ყველა შესაძლო შედეგის (ელემენტარული ხდომილებების) ალბათობათა ჯამს.

განვიხილოთ ისევ მაგალითი 6.1. სამი მონეტის აგდებისას მოსალოდნელია 8 შედეგი. ყველა ამ შედეგის ალბათობის ჯამი უდრის ერთს (თვისება 2). ყველა შედეგი არის ტოლშესაძლო, ანუ ერთნაირად მოსალოდნელი. თითოეული შედეგის ალბათობა უდრის $1/8 = 0.125$. ამიტომ:

$$P(A) = 0.125, \quad P(B) = 0.375, \quad P(C) = 0.5.$$

A და B უთავესბადი ხდომილებებია, ეს კი ნიშნავს, რომ

$$P(A \text{ ან } B) \equiv P(A \cup B) = P(A) + P(B) = 0.125 + 0.375 = 0.5.$$

A და B -ს ერთდროული მოხდენა შეუძლებელია, ამიტომ,

$$P(A \text{ და } B) \equiv P(A \cap B) = 0.$$

ვთქვათ, თქვენ ჩაკეტილი ხარტ სარდაფში თბილისში და გამოგიშვებენ მხოლოდ მაშინ, თუ სწორად უპასუხებთ კითხვას: „ხუთ წუთში განვიმდება?“ თქვენ შეიძლება მიიღოთ ასეთი ინფორმაცია:

ა. ერთი საათის წინ წვიმდა;

ბ. პარიზში წვიმს.

„ა“ ალბათ იმოქმედებს თქვენ პასუხზე, „ბ“ კი – არა.

თქვენი პასუხი „ბ“-ზე არ არის დამოკიდებული, ანუ „ბ“-სგან დამოუკიდებელია.

თვისება 6. A და B ხდომილებები დამოუკიდებელია, თუ

$$P(A \text{ და } B) \equiv P(A \cap B) = P(A) \cdot P(B).$$

დენდროგრამა (ხისებრი დიაგრამა)

დენდროგრამა – გრაფიკი, რომელიც გვეხმარება ალბათობების გამოთვლაში.

მაგალითი 6.2. ვთქვათ, 120 ადამიანიდან ერთი არის დაავადებული. ადამიანებს უტარდებათ ტესტირება ამ დაავადებაზე. ბევრი სამედიცინო ტესტი არ არის სრულყოფილი. ვიგულისხმობთ, რომ, როცა ტესტირება უტარდება ადამიანს, რომელსაც აქვს დაავადება, მაშინ 0.9 არის ალბათობა იმისა, რომ მიიღება სწორი დადებითი შედეგი. ჯანმრთელ ადამიანებში სწორი უარყოფითი შედეგის მიღება ხდება ტესტების 80%-ში. დენდროგრამის დახმარებით დავთვალოთ ალბათობა იმისა, რომ:

1. თუ ადამიანი ჯანმრთელია, მისი ტესტი იქნება დადებითი – „+“.
2. შემთხვევით შერჩეული ადამიანი იქნება ჯანმრთელი და მისი ტესტირების შედეგი იქნება „+“.
3. შემთხვევით შერჩეული ადამიანის ტესტირების შედეგი იქნება „+“.
4. შემთხვევით შერჩეული ადამიანის ტესტირების შედეგი იქნება „-“. შემოვიღოთ აღნიშვნები:

„ა“ – ადამიანი დაავადებულია.

„ჯ“ – ადამიანი ჯანმრთელია.

ცდის ჩატარებამდე:

$$P("ა") = 1/120 = 0.0083, \quad P("ჯ") = 1 - P("ა") = 1 - 0.0083 = 0.9917.$$

მოცემულია, რომ, თუ ადამიანი ავადია, ტესტი დადებით შედეგს აჩვენებს 90% ცდაში ანუ $P(+ \text{ თუ } "ა") = 0.9$ და თუ ადამიანი ჯანმრთელია, ტესტი ადასტურებს ამას 80% ცდაში ანუ $P(- \text{ თუ } "ჯ") = 0.8$. დიაგრამიდან:

$$P(- \text{ თუ } "ა") = 1 - 0.9 = 0.1 \quad \text{და} \quad P(+ \text{ თუ } "ჯ") = 1 - 0.8 = 0.2.$$

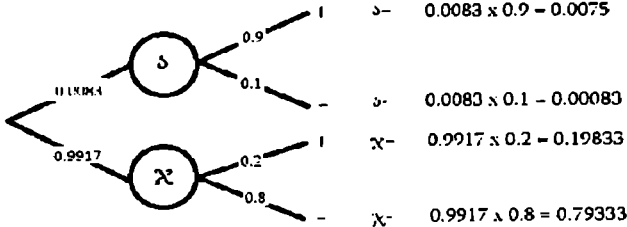
დანარჩენი შედეგების ალბათობებს პოულობენ შესაბამისი ალბათობების გადამრავლებით:

$$P("ა" \text{ და } +) = P(a+) = 0.0083 \times 0.9 = 0.00747;$$

$$P("ა" \text{ და } -) = P(a-) = 0.0083 \times 0.1 = 0.00083;$$

$$P("ჯ" \text{ და } +) = P(b+) = 0.9917 \times 0.2 = 0.19834;$$

$$P("ჯ" \text{ და } -) = P(b-) = 0.9917 \times 0.8 = 0.79336.$$



დიაგრამიდან ასევე ვიპოვით 1, 2, 3 და 4 კითხვებზე პასუხებს:

- $P(\text{თუ ადამიანი ჯანმრთელია, მისი ტესტი იქნება } +) = 0.2;$
- $P(\text{ადამიანი ჯანმრთელია და მისი ტესტი } +) =$
 $= P("ჯ" \text{ და } +) = 0.9917 \times 0.2 = 0.19834;$
- $P(\text{ტესტი } +) = P("ა" \text{ და } + \text{ ან } "ჯ" \text{ და } +) =$
 $= P("ა" \text{ და } +) + P("ჯ" \text{ და } +) = 0.00747 + 0.19834 = 0.20581 \text{ (თვისება 4);}$
- $P(\text{ტესტი } -) = P("ა" \text{ და } - \text{ ან } "ჯ" \text{ და } -) =$
 $= P("ა" \text{ და } -) + P("ჯ" \text{ და } -) = 0.00083 + 0.79336 = 0.79419 \text{ (თვისება 4).}$

იმავე ალბათობას მივიღებთ, თუ გამოვიყენებთ თვისება 3-ს:

$$P(\text{ტესტი } -) = 1 - P(\text{ტესტი } +) = 1 - 0.20581 = 0.79419.$$

შემთხვევითი სიდიდეები და მათი განაწილებები

ალბათობის თეორიაში ხდომილების ცნებასთან ერთად გამოიყენება შემთხვევითი სიდიდის ცნება. ცვლად სიდიდეს, რომლის მნიშვნელობები დამოკიდებულია ცდის შესაძლო შედეგებზე, შემთხვევითი სიდიდე ეწოდება. მაგალითად: აუდიტორიაში სტუდენტების რაოდენობა, ოჯახში ბავშვების რაოდენობა, ოთახში შემოსული ადამიანის სიმაღლე და ა.შ.

შემთხვევით სიდიდეებს აღნიშნავენ $X, Y, X, \dots$ ასოებით. შემთხვევითი სიდიდე შეიძლება იყოს დისკრეტული ან უწყვეტი. დისკრეტული შემთხვევითი სიდიდე ღებულობს განცალკევებულ (იზოლირებულ) მნიშვნელობებს, ხოლო უწყვეტი შემთხვევითი სიდიდე არ გამოტოვებს არცერთ რიცხვით მნიშვნელობას თავის ორ შესაძლო მნიშვნელობას შორის. განხილული პირველი ორი სიდიდე დისკრეტულია, მესამე – უწყვეტი.

მაგალითი 6.3. X შემთხვევითი სიდიდე იყოს მონეტის სამჯერ აგდებისას მოსულ გერბთა რიცხვი. ამ ცდაში 8 შესაძლო შედეგია: {გგგ, გგს, გსგ, სგგ, გსს, სსგ, სსს}.

X -ს შეუძლია მიიღოს მნიშვნელობები 0, 1, 2, 3.

$$P(X=0) = P\{სსს\} = 0.5 \times 0.5 \times 0.5 = 0.125;$$

$$P(X=1) = P\{გსს \text{ ან } სგს \text{ ან } სსგ\} = 0.125 + 0.125 + 0.125 = 0.375;$$

$$P(X=2) = P\{გგს \text{ ან } გსგ \text{ ან } სგგ\} = 0.125 + 0.125 + 0.125 = 0.375;$$

$$P(X=3) = P\{გგგ\} = 0.125.$$

შესაბამისობას შემთხვევითი სიდიდის შესაძლო მნიშვნელობებსა და მათ შესაბამის ალბათობებს შორის, დისკრეტული ტიპის შემთხვევითი სიდიდის განაწილების კანონი ეწოდება. განაწილების კანონი შეიძლება იყოს მოცემული ცხრილის, ფორმულის ან გრაფიკის სახით.

ცხრილით მოცემული შემთხვევითი სიდიდის განაწილების კანონია:

x	0	1	2	3
$P(X=x)$	0.125	0.375	0.375	0.125

ვიპოვოთ $P(X \geq 1)$. გვაქვს:

$$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3) = 0.875$$

ან ალტერნატიული გზით

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X=0) = 1 - 0.125 = 0.875.$$

სხვადასხვა შემთხვევითი სიდიდე და განაწილებები სხვადასხვა ცდასთანაა დაკავშირებული. ერთ-ერთი მნიშვნელოვანი დისკრეტული განაწილება არის ბინომიალური განაწილება. უწყვეტი განაწილების მაგალითებია ე.წ ნორმალური განაწილება, t -განაწილება, ზი-კვადრატ განაწილება, F -განაწილება.

ბაიესის ფორმულა

დენდროგრამა გვაძლევს საშუალებას მივიღოთ ე.წ. ბაიესის ფორმულა.

თვისება 7. თუ მოხდა B ხდომილება, ჩვენ შეგვიძლია გამოვთვალოთ სხვა A ხდომილების მოხდენის ალბათობა:

$$P(A \text{ თუ } B) \equiv P(A|B) = \frac{P(A \text{ და } B)}{P(B)} = \frac{P(A \cap B)}{P(B)},$$

თუ $P(B) \neq 0$.

ამ ალბათობას ეწოდება A -ს პირობითი ალბათობა B -ს პირობაში.

ადგილი აქვს დამოუკიდებლობის ეკვივალენტურ განმარტებას: A და B ხდომილებები დამოუკიდებელია, თუ $P(A|B) = P(A)$.

თვისება 8. პირობითი ალბათობის განმარტებიდან მიიღება ნამრავლის ალბათობის ფორმულა:

$$P(A \cap B) = P(B) \cdot P(A|B),$$

რასაც, თავის მხრივ, მივყავართ ბაიესის ფორმულამდე:

თვისება 9. ბაიესის ფორმულა:

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}.$$

დენდროგრამის დახმარებით ჩვენ შეგვიძლია ვიპოვოთ როგორც $P(A \text{ და } B)$, ასევე $P(B)$.

განვიხილოთ ისევ მაგალითი 6.2 და ვიპოვოთ ალბათობა იმისა, რომ ადამიანი დაავადებულია, თუ მისი ტესტირების შედეგია „+“.

$$P(\text{"ა" თუ } +) = P(\text{"ა" და } +) / P(+) = 0.00747 / 0.20581 = 0.036 \quad (=3.6\%).$$

როგორც ვხედავთ, ეს საკმაოდ პატარა ალბათობაა. ტესტი არ არის კარგი იმის გამო, რომ ჯანმრთელი ადამიანებისათვის ტესტების 20% გვაძლევს მცდარ შედეგს.

იგივე ალბათობა ჩვენ შეგვიძლია ვიპოვოთ ბაიესის ფორმულის გამოყენებითაც:

$$P(\text{"ა" თუ } +) = \frac{P(\text{"ა"}) \cdot P(+ \text{ თუ } \text{"ა"})}{P(+)} = \frac{0.0083 \cdot 0.9}{0.20581} = 0.036.$$

თვისება 10. თუ აუცილებელი ხდომილებად აყოფილია n ცალ წყვილ-წყვილად უთავსებად ხდომილებად $A_1, A_2, \dots, A_n$, მაშინ ნებისმიერი B ხდომილების ალბათობა გამოითვლება ე.წ. სრული ალბათობის ფორმულით:

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

შედეგი. ნებისმიერი A და B ხდომილებებისათვის ($P(A) \neq 0, P(\bar{A}) \neq 0$):

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A}).$$

შეჯამება

ალბათობა გვაძლევს საშუალებას აღვწეროთ ხდომილების მოხდენის შანსი. დენდროგრამა გვჩვენებს სხვადასხვა ხდომილების ალბათობების დათვლაში. შემთხვევითი სიდიდე გვჩვენებს ცდის შედეგების და ამ შედეგების შესაბამისი ალბათობების აღწერაში. ბაიესის ფორმულა გვაძლევს საშუალებას გამოვთვალოთ A ხდომილების ალბათობა, თუ მოხდა B ხდომილება შებრუნებული რიგით აღებული პირობითი ალბათობის საშუალებით.

ძირითადი ტერმინები

ალბათობა, ხდომილება, უთავსებადი ხდომილებები, დამოუკიდებელი ხდომილებები, შემთხვევითი სიდიდე, დისკრეტული შემთხვევითი სიდიდის განაწილება, უწყვეტი შემთხვევითი სიდიდე.

კითხვები

1. თქვენი აზრით, ქვემოთ ჩამოთვლილი ხდომილებებიდან, რომელია ერთმანეთისაგან დამოუკიდებელი:
 - ა. სტუდენტი სწავლობს საზღვარგარეთ და მისი მშობლების შემოსავალი მეტია 20 000 ლარზე;
 - ბ. წვიმს თბილისში და წვიმს ლონდონში;
 - გ. ადამიანი დაავადებულია და ტესტი ამ დაავადებაზე დადებითია;
 - დ. ქუჩაში მოძრაობა გადატვირთულია და ჰაერის დაბინძურების დონე მაღალია;
 - ე. ფასები ნათობზე მაღალია და შუა აღმოსავლეთში მდგომარეობა არასტაბილურია;
 - ვ. საქართველოში ყურძნის კარგი მოსავალია და აშშ-ში ვალუტის კურსი დაბალია.
2. დაავადება ტახემია შეიძლება განვითარდეს ორსულობის დროს. ტახემია უფრო ხშირად ვითარდება დიაბეტით დაავადებულ ქალებში. ცნობილია, რომ ქალების დაახლოებით 2%-ს აქვს დიაბეტი და ამ ქალების 25%-ს შეიძლება განუვითარდეს ტახემია. ჯანმრთელი ორსული ქალების დაახლოებით 4%-ში ფიქსირდება ტახემია.
 - ა. დახაზეთ დენდროგრამა;
 - ბ. იპოვეთ ყველა შესაძლო შედეგის ალბათობა;
 - გ. რას უდრის ალბათობა იმისა, რომ ორსულ ქალს არ განუვითარდება ტახემია?
 - დ. რას უდრის ალბათობა იმისა, რომ ორსულ ქალს განუვითარდება ტახემია?
3. შესწავლილი იყო დიპერსიის დონე ასაკის მიხედვით. რესპონდენტები განაწილებული იყო სამ ასაკობრივ ჯგუფში. 50% მოხვდა პირველ ჯგუფში და ამ რესპ-

ონდენტების 15%-ის დეპრესიის დონე იყო მაღალი. 36% იყო მეორე ასაკობრივი ჯგუფიდან და აქედან 96%-ის დეპრესიის დონე იყო დაბალი. დანარჩენი რესპონდენტების 8%-ის დეპრესიის დონე იყო მაღალი.

- ა. დახაზეთ დენდროგრამა;
 - ბ. გამოთვალეთ ყველა შესაძლო შედეგის ალბათობა;
 - გ. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული რესპონდენტის დეპრესიის დონე იქნება დაბალი;
 - დ. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული რესპონდენტი არის პირველი ასაკობრივი ჯგუფიდან, თუ ცნობილია, რომ მისი დეპრესიის დონე დაბალია.
4. ყოველწლიური დაკვირვების მიხედვით, 0 – 4 წლის ბავშვების დედების 15% ეწევა სიგარეტს. მწველი დედების ამ ასაკის ბავშვების 5%-ში აღირიცხება ასთმა. თუ დედა არ ეწევა, ასთმა აღირიცხება ბავშვების 1%-ში.
- ა. დახაზეთ დენდროგრამა;
 - ბ. გამოთვალეთ ყველა შესაძლო შედეგის ალბათობა;
 - გ. იპოვეთ ალბათობა იმისა, რომ დედა ეწევა, თუ ბავშვს აქვს ასთმა.
5. ფირმაში მომუშავე თანამშრომლების 60% კაცია, მათგან 70% დაოჯახებულია. ქალების 65% არ არის დაოჯახებული;
- ა. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული თანამშრომელი არ არის დაოჯახებული;
 - ბ. იპოვეთ ალბათობა იმისა, რომ თანამშრომელი ქალია, თუ ცნობილია, რომ ის არ არის დაოჯახებული.
6. კაცებში უფრო ხშირად აღირიცხება შიზოფრენია, ვიდრე ქალებში. ალბათობა იმისა, რომ შემთხვევით შერჩეული კაცი დაავადებულია შიზოფრენიით, ტოლია 0.003-ს, ხოლო იგივე ალბათობა ქალებისათვის არის 0.002. ალბათობა იმისა, რომ გარკვეული წამალი დადებითად იმოქმედებს დაავადებულ კაცზე, არის 0.7, ქალზე კი – 0.4.
- ა. ცნობილია, რომ ქალაქის მოსახლეობის 56% ქალია. შემთხვევით შეარჩიეს ერთი ადამიანი. დახაზეთ დენდროგრამა და აღნიშნეთ ყველა შედეგი ინდივიდის სქესის, დაავადების და მკურნალობის მიხედვით;
 - ბ. გამოთვალეთ ყველა შედეგის ალბათობა;
 - გ. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული ინდივიდი შიზოფრენიით დაავადებული კაცია და მასზე წამალი დადებითად არ იმოქმედებს.
7. ვაშლის პარტიების ხარისხის ინსპექტორი მისაღებად თვლის „ცუდი“ პარტიების 5%-ს და ინუნებს „კარგი“ პარტიების 1%-ს. გამოგზავნილი პარტიების 95% „კარგია“.
- ა. განსაზღვრეთ ალბათობა იმისა, რომ:
 1. მიიღება „ცუდი“ პარტია;
 2. დანუნებული იქნება „კარგი“ პარტია.
 - ბ. გამოთვალეთ ალბათობა იმისა, რომ ინსპექტორი მცდარ გადანყვეტილებას მიიღებს.

8. მენეჯერს სურს შეაფასოს მისი გუნდების მუშაობა. მან განიხილა ასეთი ელემენტარული ხდომილებები:

A – პროექტი დროზე ადრე დასრულდა;

B – პროექტი დროზე დასრულდა;

C – პროექტი დროზე არ დასრულდა.

თუ შესაფასებელია ორი გუნდის მუშაობა, მაშინ განსახილველია შემდეგი ელემენტარული ხდომილებები:

ელემენტარული ხდომილება	პროექტი 1	პროექტი 2
A_1	ადრე	ადრე
A_2	ადრე	დროზე
A_3	ადრე	გვიან
A_4	დროზე	ადრე
A_5	დროზე	დროზე
A_6	დროზე	გვიან
A_7	გვიან	ადრე
A_8	გვიან	დროზე
A_9	გვიან	გვიან

- ა. გამოისახება თუ არა ხდომილება: „ერთი მაინც გვიან“ შემდეგი ელემენტარული ხდომილებებით: A_3, A_6, A_7, A_8, A_9 ?
- ბ. როგორ განმარტავთ ხდომილებას: (A_1, A_2, A_4, A_5) ?
9. იწვორმაცია თანამშრომლების გადაადგილების და მივლინებების შესახებ მენეჯერმა გამოსახა ცხრილით:

მივლინება წელიწადში	სქესი		სულ
	მდედრ. E_4	მამრობ. E_5	
E_1 1-2-ჯერ	F_1 $f = 450$	F_2 $f = 500$	$\Sigma = 950$
E_2 3-10-ჯერ	F_3 $f = 300$	F_4 $f = 800$	$\Sigma = 1100$
E_3 10-ზე მეტჯერ	F_5 $f = 100$	F_6 $f = 350$	$\Sigma = 450$
სულ	$\Sigma = 850$	$\Sigma = 1650$	$\Sigma = 2500$

- ა. განმარტეთ, რას ნიშნავს E_1, E_2, E_3, E_4, E_5 ხდომილებები?

- ბ. რა ხდომილებებს აღნიშნავენ $F_1, F_2, F_3, F_4, F_5, F_6$;

გ. გამოთვალეთ ალბათობები: $P(E_2)$, $P(E_3)$, $P(F_1)$, $P(E_2 \text{ და } F_4)$, $P(E_2 \text{ თუ } E_4)$, $P(E_4 \text{ თუ } E_2)$.

10. გამოიყენეთ სიხშირეთა განაწილების ცხრილი და გამოთვალეთ: $P(B)$, $P(C)$, $P(E)$, $P(A \text{ და } B)$, $P(D \text{ და } E)$, $P(E \text{ ან } B)$, $P(A \text{ თუ } B)$.

	A	B	C	სულ
D	20	32	18	70
E	12	28	40	80
F	8	20	22	50
სულ	40	80	80	200

11. ყუთში 10 წითელი, 5 თეთრი და 4 ლურჯი ბურთია. შემთხვევით იღებენ ერთ ბურთს. ვიპოვოთ ალბათობა იმისა, რომ: ა) ამოღებულია ლურჯი ბურთი. ბ) არაა ამოღებული წითელი ბურთი. გ) არაა ამოღებული თეთრი ბურთი. დ) ამოღებულია წითელი ან თეთრი ბურთი.
12. რვა ადამიანისაგან, რომელთა შორის 3 დოქტორია და 5 კანდიდატი, უნდა შევადგინოთ სამკაცრიანი კომისია. კომისიის წევრები შეირჩევა შემთხვევით. როგორია ალბათობა იმისა, რომ კომისიაში მოხვდება: ა) სამი დოქტორი? ბ) ორი დოქტორი და ერთი კანდიდატი? გ) ერთი დოქტორი და ორი კანდიდატი? დ) სამი კანდიდატი?
13. 9 წიგნიდან, რომელთა შორის 4 ხელოვნებაზე და 5 დეტექტივია, შემთხვევით ირჩევენ სამს. როგორია ალბათობა იმისა, რომ: ა) სამივე წიგნი იქნება ხელოვნებაზე? ბ) სამივე იქნება დეტექტივი? გ) ერთი ან ორი წიგნი იქნება ხელოვნებაზე? დ) ზუსტად ორი იქნება დეტექტივი?
14. მონეტას აგდებენ სამჯერ. როგორია ალბათობა იმისა, რომ გერბი მოვა: ა) კენტ რიცხვეჯერ? ბ) ლუნ რიცხვეჯერ? გ) არცერთხელ? დ) ორჯერ მაინც?
15. მონეტას აგდებენ 4-ჯერ. როგორია ალბათობა იმისა, რომ გერბი მოვა: ა) კენტ რიცხვეჯერ? ბ) ლუნ რიცხვეჯერ? გ) არანაკლებ ორჯერ? დ) ერთჯერ მაინც?
16. მონეტას აგდებენ ოთხჯერ. როგორია ალბათობა იმისა, რომ საფასური მოვა: ა) არანაკლებ სამჯერ? ბ) არანაკლებ ორჯერ? გ) არანაკლებ ერთჯერ? დ) არცერთხელ?
17. მონეტას აგდებენ ხუთჯერ. როგორია ალბათობა იმისა, რომ: ა) გერბი მოვა კენტ რიცხვეჯერ, ხოლო საფასური ლუნ რიცხვეჯერ? ბ) გერბი მოვა კენტ რიცხვეჯერ? გ) გერბი მოვა ლუნ რიცხვეჯერ? დ) გერბი მოვა ლუნ რიცხვეჯერ, ხოლო საფასური კენტ რიცხვეჯერ?
18. მონეტას აგდებენ ხუთჯერ. როგორია ალბათობა იმისა, რომ გერბი მოვა: ა) არანაკლებ სამჯერ? ბ) არანაკლებ ოთხჯერ? გ) ან ერთჯერ ან ორჯერ? დ) ან ერთჯერ ან ოთხჯერ? ე) ან ორჯერ ან სამჯერ?
19. სათამაშო კამათელს აგორებენ ორჯერ. რა უფრო ალბათურია – ჯამში მოვა: ა) 6 ქულა თუ 8 ქულა? ბ) 5 ქულა თუ 9 ქულა? გ) 5 ქულა თუ 6 ქულა? დ) 5 ქულა თუ 8 ქულა? ე) 6 ქულა თუ 7 ქულა? ვ) 7 ქულა თუ 5 ქულა?

20. როგორია ალბათობა იმისა, რომ ორი სათამაშო კამათლის გაგორებისას მოვა ისეთი ქულები, რომლებიც ორჯერ განსხვავდებიან ერთმანეთისაგან?
21. როგორია ალბათობა იმისა, რომ სათამაშო კამათლის ორჯერ გაგორებისას ერთჯერ მაინც მოვა: ა) ერთიანი? ბ) ორიანი? გ) სამიანი? დ) ოთხიანი? ე) ხუთიანი? ვ) ექვსიანი?
22. რომელია მეტი და რამდენით: ალბათობა იმისა, რომ სათამაშო კამათლის ერთჯერ გაგორებისას მოვა 6 ქულა თუ ორჯერ გაგორებისას ჯამში მოვა 6 ქულა?
23. რომელია მეტი და რამდენით: ალბათობა იმისა, რომ სათამაშო კამათლის ერთჯერ გაგორებისას მოვა ლუნი ქულა თუ ორჯერ გაგორებისას ჯამში მოვა ლუნი ქულა?
24. სათამაშო კამათელს აგორებენ ორჯერ. როგორია ალბათობა იმისა, რომ მოსულ ქულათა ჯამი აღმოჩნდება: ა) კენტი რიცხვი? ბ) 4-ზე მეტი და 9-ზე ნაკლები? გ) 3-ის ჯერადი? დ) 4-ის ჯერადი? ე) 5-ის ჯერადი? ვ) 6-ის ჯერადი?
25. რომელია მეტი და რამდენით: ალბათობა იმისა, რომ სათამაშო კამათლის ერთჯერ გაგორებისას მოვა 3-ის ჯერადი ქულა თუ ორჯერ გაგორებისას ჯამში მოვა 3-ის ჯერადი ქულა?
26. სათამაშო კამათელს აგორებენ სამჯერ. როგორია ალბათობა იმისა, რომ მოვა სამი ნახნავი: ა) ერთი და იმავე ქულებით? ბ) სხვადასხვა ქულით? გ) 9-ის ტოლ ქულათა ჯამით? დ) 10-ის ტოლ ქულათა ჯამით? ე) 5-ის ჯერად ქულათა ჯამით?
27. სათამაშო კამათელს აგორებენ სამჯერ. როგორია ალბათობა იმისა, რომ მოვა: ა) ერთჯერ მაინც ორიანი? ბ) ერთჯერ მაინც სამიანი? გ) ერთჯერ მაინც ოთხიანი? დ) ერთჯერ მაინც ხუთიანი? ე) ერთჯერ მაინც ექვსიანი?
28. რომელი ალბათობაა მეტი და რამდენით: ა) 5 ქულის მოგროვება სათამაშო კამათლის ორჯერ გაგორებისას თუ სამჯერ გაგორებისას? ბ) 6 ქულის მოგროვება სათამაშო კამათლის ორჯერ გაგორებისას თუ სამჯერ გაგორებისას? გ) 11 ქულის მოგროვება სათამაშო კამათლის ორჯერ გაგორებისას თუ სამჯერ გაგორებისას?
29. A იყოს ხდომილება, რომ შაბათს იწვიმებს, ხოლო B – კვირას იწვიმებს. დაეუშვათ, რომ $P(A) = P(B) = 0.5$. p იყოს ალბათობა, რომ იწვიმებს ორივე დღეს. იპოვეთ ალბათობა იმისა, რომ: ა) იწვიმებს შაბათს, მაგრამ არ იწვიმებს კვირას; ბ) იწვიმებს ერთ დღეს, მაგრამ არ იწვიმებს ორივე დღეს; გ) არ იწვიმებს არცერთ დღეს.
30. განვითარებულ ქვეყნებში საზოგადოდ ახალშობილის გადარჩენის შანსი შეადგენს 99.3%-ს. ახალშობილთა 15% იბადება საკეისრო კვეთის გამოყენებით და მათი გადარჩენის შანსია 98.7%. რას უდრის ახალშობილის გადარჩენის შანსი, თუ ცნობილია, რომ ის არ დაბადებულა საკეისრო კვეთის გამოყენებით.
31. ალბათობა იმისა, რომ დაოჯახებული მამაკაცი (შესაბამისად, ქალი) უყურებს სერიალს, არის 0.4 (შესაბამისად, 0.5). ალბათობა იმისა, რომ მამაკაცი უყურებს სერიალს პირობაში, რომ მისი ცოლი უყურებს მას, ტოლია 0.7-ის. იპოვეთ ალბათობა იმისა, რომ: ა) დაოჯახებული წყვილი უყურებს სერიალს; ბ) ცოლი უყურებს სერიალს პირობაში, რომ მისი ქმარი უყურებს მას; გ) სულ ცოტა, ერთი მეუღლე-თავანი უყურებს სერიალს.

32. აშშ-ის მოსახლეობა სისხლის ჯგუფების მიხედვით განაწილებულია შემდეგნაირად: $O - 45\%$, $A - 40\%$, $B - 11\%$, $AB - 4\%$. სისხლი აგრეთვე ხასიათდება რეზუსის ტიპით, რომელიც შეიძლება იყოს დადებითი ან უარყოფითი და რომელიც დამოუკიდებელია სისხლის ჯგუფისაგან. აშშ-ის მოსახლეობის 84%-ს გააჩნია რეზუსუარყოფითი სისხლი. შემთხვევით შეარჩიეს ორი ადამიანი. იპოვეთ ალბათობა იმისა, რომ: ა) ორივეს აქვს A ჯგუფის რეზუსდადებითი სისხლი; ბ) ერთ-ერთს აქვს O ჯგუფის რეზუსდადებითი სისხლი, ხოლო მეორეს – არა; გ) სულ ცოტა, ერთს მაინც აქვს O ჯგუფის რეზუსდადებითი სისხლი; დ) ერთ-ერთს აქვს რეზუსდადებითი, ხოლო მეორეს არა აქვს AB ჯგუფის სისხლი; ე) მათ აქვთ ერთი და იმავე ჯგუფის სისხლი; ვ) მათ აქვთ ერთი და იმავე ჯგუფის, მაგრამ განსხვავებული რეზუსის მქონე სისხლი.
33. A ყუთი შეიცავს 8 წითელ და 7 მწვანე ბურთს, B ყუთი შეიცავს 6 წითელ და 4 მწვანე ბურთს. აგორებენ წესიერ კამათელს. თუ მოვიდა 3-ის ჯერადი ქულა, მაშინ A ყუთიდან შემთხვევით იღებენ 1 ბურთს, ხოლო, წინააღმდეგ შემთხვევაში, B ყუთიდან შემთხვევით იღებენ 1 ბურთს. იპოვეთ ალბათობა იმისა, რომ: ა) ამოღებული ბურთი მწვანეა; ბ) ბურთი ამოღებული იყო A ყუთიდან, თუ ცნობილია, რომ ამოღებული ბურთი მწვანეა.
34. გარკვეული დაავადება გვხვდება ადამიანთა პოპულაციის 0.1%-ში. დიაგნოსტიკის მეთოდი კორექტულ პასუხს იძლევა ალბათობით 0.99. პაციენტმა გაიარა შემოწმება და გამოკვლევამ აჩვენა დადებითი შედეგი. რას უდრის ალბათობა იმისა, რომ პაციენტი დაავადებულია?
35. ქალაქგარეთ მცხოვრებს აქვს ორი ავტომობილი: ერთი კომპაქტური და მეორე სტანდარტული. ოთხიდან სამ შემთხვევაში ის სამსახურში მიდის კომპაქტური ავტომობილით, ხოლო ერთ შემთხვევაში კი სტანდარტულით. კომპაქტური ავტომობილით მგზავრობისას ის სახლიდან გადის 6:30-ზე 75% შემთხვევაში და 6:35-ზე სხვა დროს, ხოლო სტანდარტულით მგზავრობისას კი 6:35-ზე 60% შემთხვევაში და 6:30-ზე სხვა დროს. იპოვეთ ალბათობა იმისა, რომ: ა) ქალაქგარეთ მცხოვრები სახლიდან გავა 6:35-ზე; ბ) ქალაქგარეთ მცხოვრებმა გამოიყენა კომპაქტური ავტომობილი, თუ ცნობილია, რომ ის სახლიდან გავიდა 6:35-ზე.

თავი 7

პროპორციები და ბინომური განაწილება

ხშირად, ცდის შედეგად ფიქსირდება მხოლოდ ორი შედეგი: „ნარმატება“ ან „მარცხი“. მაგალითად: ფირმის მოგება ან წაგება, გამოცდის ჩაბარება ან არჩაბარება, ახალშობილი იქნება ბიჭი ან გოგო. ყველა ასეთ შედეგს აღწერენ შემთხვევითი სიდიდით, რომელსაც „ნარმატების“ შემთხვევაში შეუსაბამებენ 1-ს, ხოლო „მარცხის“ შემთხვევაში 0-ს. ამავე დროს ნარმატებას მიაწერენ p ალბათობას, ხოლო მარცხს $1 - p$ ალბათობას. ტერმინი „ნარმატება“ არ ნიშნავს ნარმატებას ჩვეულებრივი აზრით, ის უბრალოდ აღნიშნავს ექსპერიმენტატორისათვის საინტერესო ერთ-ერთ შედეგს და შეიძლება აღნიშნავდეს, მაგალითად, სტუდენტის „ჩაჭრას“ გამოცდაზე.

ვთქვათ, ვატარებთ განმეორებით დამოუკიდებელ ცდებს და თითოეული ცდის შედეგია 1 („ნარმატება“), ან 0 („მარცხი“). ჩვენ განვიხილავთ ცდათა ისეთ მიმდევრობას, როცა ნარმატებათა ალბათობა ცდიდან ცდამდე უცვლელია და p -ს ტოლია.

დენდროგრამა ვერ დაგვეხმარება ასეთი ცდების შედეგების და ალბათობების გამოთვლაში, რადგან შედეგთა რაოდენობა შეიძლება ძალიან დიდი აღმოჩნდეს, მაგალითად, როცა ვაგდებთ 10 მონეტას, შესაძლო შედეგთა რაოდენობაა $-2^{10} = 1024$; 15 მონეტის აგდებისას კი $-2^{15} = 32768$.

ბინომური განაწილების თვისებები

ბინომური შემთხვევითი სიდიდე იღებს მნიშვნელობებს: $0, 1, 2, \dots, n$, ანუ უდრის ნარმატებათა რაოდენობას n ცდაში. მნიშვნელობათა შესაძლო რაოდენობაა $n + 1$.

X შემთხვევითი სიდიდეს აქვს ბინომური განაწილება, თუ:

- n – ცდათა რაოდენობა ფიქსირებულია,
- თითოეულ ცდას გააჩნია მხოლოდ ორი შესაძლო შედეგი,
- ნარმატების ალბათობა არ იცვლება ცდიდან ცდამდე, ის მუდმივია,
- ცდები არიან დამოუკიდებელნი (ერთი ცდის შედეგი არ მოქმედებს სხვა ცდების შედეგებზე).

ტერმინი „ბინომური“ აღნიშნავს ფაქტს, როდესაც ყოველ ცდაში მოსალოდნელია მხოლოდ ორი შესაძლო შედეგი (ბი – ორი).

მონეტის აგდებისას მოსალოდნელია ორი შედეგი: „გერბი“ ან „საფასური“. სიმეტრიული მონეტის შემთხვევაში გერბის მოსვლის ალბათობა არ იცვლება ცდიდან ცდამდე და უდრის 0.5-ს. კამათლის აგდებისას „ნარმატებად“ შეიძლება

ჩაითვალოს ექვსი ქულის მოსვლა, მაშინ ყველა დანარჩენი შედეგი ჩაითვლება „მარცხად“. ამ ცდაში $p=1/6$, $1-p=5/6$. სტუდენტის სიმაღლის გაზომვისას „წარმატებად“ შეიძლება ჩაითვალოს შედეგი, თუ სიმაღლე მეტია ან ტოლი 165 სმ-ზე და „მარცხად“, თუ სიმაღლე ნაკლებია 165-ზე.

ცდათა რიცხვს n -ს და წარმატების ალბათობას p -ს ბინომური განაწილების პარამეტრები ეწოდება.

თუ X შემთხვევითი სიდიდეს აქვს ბინომური განაწილება პარამეტრებით n და p , ეს ფაქტი შემდეგნაირად აღინიშნება: $X \equiv Bin(n, p)$.

მაგალითი 7.1. ქვემოთ ჩამოთვლილიდან რომელ ცდას აღწერს ბინომური შემთხვევითი სიდიდე:

ა. შემთხვევით შეარჩიეს 10 სტუდენტი, X – სტუდენტების რაოდენობა, ვისაც აქვს მანქანის ტარების უფლება.

დიახ, ამ ცდას აღწერს ბინომური შემთხვევითი სიდიდე. ცდათა რიცხვი n – ფიქსირებულია და ათის ტოლია, თითოეული ცდა მოიცავს ორ შესაძლო შედეგს. ქალაქში ცნობილია მანქანის ტარების უფლების მქონე ადამიანთა პროპორცია (p), ცდები დამოუკიდებელია.

ბ. შემთხვევით შეარჩიეს პიროვნება და აღნიშნეს რომელ ასაკობრივ ჯგუფს ეკუთვნის ის: ≤ 20 , $20-30$, ≥ 30 .

არა, ამ ცდას არ აღწერს ბინომური შემთხვევითი სიდიდე. ამ ცდაში სამი შესაძლო შედეგია.

გ. X – გოლის შეგდებაზე დარტყმათა რაოდენობა.

არა, ამ ცდას არ აღწერს ბინომური შემთხვევითი სიდიდე. ცდათა რიცხვი ფიქსირებული არ არის.

დ. შემთხვევით გამოკითხეს 50 კაცი 25 ოჯახიდან – არიან თუ არა ისინი მანქანის მფლობელები.

არა, ცდები არ არის დამოუკიდებელი, რადგან ოჯახიდან შეიძლება გამოკითხვაში მოხვდეს ორი ან მეტი ადამიანი და მათი პასუხები შეიძლება დამოკიდებული იყოს ერთმანეთზე.

ე. შეამოწმეს სტუდენტების რეაქციის დრო 100 სტიმულზე. „წარმატება“ – როცა რეაქციის დრო ≤ 0.1 წმ-ზე.

არა, ამ ცდას არ აღწერს ბინომური შემთხვევითი სიდიდე. აქ შეიძლება, ცდიდან ცდამდე, სტუდენტის დაღლის გამო, შეიცვალოს წარმატების ალბათობა.

ბინომური ალბათობების გამოთვლა

ბინომურმა შემთხვევითმა სიდიდემ X – n ცდაში წარმატებათა რაოდენობამ შეიძლება მიიღოს ერთ-ერთი მნიშვნელობა ქვემოთ ჩამოთვლილიდან:

$$0, 1, \dots, n-1, n.$$

მინიმალური – 0 ნიშნავს, რომ წარმატება არ დაფიქსირდა, მაქსიმალური – n ნიშნავს, რომ ყველა ცდა წარმატებით დამთავრდა.

ამ შედეგების ალბათობების ერთობლიობას $P(X = x)$, როცა $x = 0, 1, \dots, n-1, n$ ბინომური განაწილება ეწოდება.

სამართლიანია ბერნულის ფორმულა:

$$P(X = x) = C_n^x p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n-1, n,$$

სადაც

$$C_n^x = \frac{n!}{x!(n-x)!}.$$

კუმულატური (დაგროვილი) ბინომური განაწილება ეწოდება ალბათობებს:

$$P\{Bin(n, p) \leq x\} = \sum_{k=0}^x C_n^k p^k (1-p)^{n-k}.$$

ცხადია, რომ

$$P\{Bin(n, p) = x\} = C_n^x p^x (1-p)^{n-x} = P\{Bin(n, p) \leq x\} - P\{Bin(n, p) \leq x-1\}.$$

წარმატებათა იმ x_0 რაოდენობას, რომელსაც ყველაზე დიდი ალბათობა შეესაბამება (ყველაზე მეტად მოსალოდნელია), უალბათესი რიცხვი ეწოდება. უალბათესი რიცხვი წარმოადგენს ისეთ მთელ რიცხვს, რომელიც აკმაყოფილებს უტოლობას:

$$np - (1-p) \leq x_0 \leq np + p.$$

ბინომური შემთხვევითი სიდიდის ალბათობები შესაძლებელია ვიპოვოთ კუმულატური ბინომური განაწილების ცხრილის (იხილეთ დანართი2) საშუალებით. იმისათვის, რომ გამოვიყენოთ ეს ცხრილი $P(X = x)$ ალბათობის გამოსათვლელად, პირველ რიგში, ცხრილში უნდა ვიპოვოთ n -ის შესაბამისი ბლოკი, შემდეგ სვეტების თავზე p -ს მნიშვნელობა და p -სა და x -ის გადაკვეთაზე მდგომ რიცხვს გამოვკლოთ p -სა და $(x-1)$ -ის გადაკვეთაზე მდგომი რიცხვი. მაგალითად:

$$P\{Bin(10, 0.25) = 7\} = 0.99958 - 0.99649 = 0.00309.$$

მაგალითი 7.2. ვთქვათ, $X \equiv Bin(4, 0.3)$, ვიპოვოთ X -ის განაწილება. ამისათვის ბინომური განაწილების ცხრილში ვიპოვოთ $n=4$ -ისა და $p=0.3$ -ის შესაბამისი ბლოკი. მაშინ დავინახავთ, რომ საძიებელი ალბათობები ტოლი იქნება:

$$P(X = 0) = P(X \leq 0) = 0.2401, \quad P(X = 1) = 0.6517 - 0.2401 = 0.4116,$$

$$P(X = 2) = 0.2646, \quad P(X = 3) = 0.0756, \quad P(X = 4) = 0.0081.$$

მაგალითი 7.3. 12-ჯერ აგდებენ მონეტას. ვიპოვოთ ალბათობა იმისა, რომ:

- გერბი მოვა 6-ჯერ;
- გერბი არცერთხელ არ მოვა;
- გერბი მოვა 8-ზე მეტი;
- გერბი მოვა არაუმეტეს 2-ჯერ.

ამოხსნა: იგულისხმება, რომ მონეტა სიმეტრიულია, $p = 0.5$; $n = 12$. $X \equiv \text{Bin}(12, 0.5)$ და ბინომური განაწილების ცხრილის საშუალებით ვპოულობთ:

$$P(X = 6) = 0.2256; \quad P(X = 0) = 0.0002;$$

$$\begin{aligned} P(X > 8) &= P\{(X = 9) \cup (X = 10) \cup (X = 11) \cup (X = 12)\} = \\ &= P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) = \\ &= 0.0537 + 0.0161 + 0.0029 + 0.0002 = 0.0729; \end{aligned}$$

$$\begin{aligned} P(X \leq 2) &= P\{(X = 2) \cup (X = 1) \cup (X = 0)\} = P(X = 2) + P(X = 1) + P(X = 0) = \\ &= 0.0161 + 0.0029 + 0.0002 = 0.0192. \end{aligned}$$

მაგალითი 7.4. ვთქვათ, ალბათობა იმისა, რომ სტუდენტი ნახევარ საათში სწორად ვერ ამოხსნის ამოცანას, 0.05-ის ტოლია. ჯგუფში 15 სტუდენტი. X – სტუდენტების რაოდენობა, ვინც სწორად ვერ ამოხსნის ამოცანას ნახევარ საათში. $n = 15$, $p = 0.05$. მაშინ ბინომური განაწილების ცხრილიდან გვექნება:

$$P(X = 0) = 0.4633; \quad P(X = 1) = 0.3658; \quad P(X = 2) = 0.1348;$$

$$P(X = 3) = 0.0307; \quad P(X = 4) = 0.0049; \quad P(X = 5) = 0.0006; \quad P(X = 6) = ?.$$

ბინომური განაწილების ცხრილში ცარიელი ადგილი ნიშნავს, რომ შესაბამისი ალბათობა 0.00005-ზე ნაკლებია. ეს ასეა, როცა $X = 6, 7, \dots, 15$.

ზოგიერთი ამოცანის ამოხსნისას საჭირო ხდება „წარმატებას“ და „მარცხს“ ადგილი შევუცვალოთ. ამის აუცილებლობა აიხსნება იმით, რომ ბინომური განაწილების ცხრილში არ არის მითითებული ალბათობები $p \geq 0.5$ -თვის. როცა წარმატების ალბათობაა p , მაშინ მარცხის ალბათობაა $1 - p$. ამიტომ, თუ $p \geq 0.5$, მაშინ $1 - p \leq 0.5$.

თუ $X \equiv \text{Bin}(n, p)$, მაშინ n ცდაში მარცხის რაოდენობის Y -ის განაწილებაა

$$Y \equiv \text{Bin}(n, 1 - p) \quad \text{და} \quad X + Y = n.$$

მაგალითი 7.5. ფირმის თანამშრომლების 70%-ის ასაკი ≤ 30 -ზე. შემთხვევით შეარჩიეს 20 თანამშრომელი. რას უდრის ალბათობა, რომ არაუმეტეს 10 მათგანის ასაკი იქნება ≤ 30 -ზე.

გვაქვს: $X \equiv \text{Bin}(n, p)$. ცდათა რიცხვი $n = 20$ ფიქსირებულია, $p = 70/100 = 0.7$. ცდა სრულდება ორი შესაძლო შედეგით, დამოუკიდებელი ცდებია, მაგრამ ბინომური განაწილების ცხრილში არ არის $p = 0.7$. გამოვთვალოთ სხვა Y სიდიდის ალბათობა. Y – თანამშრომლების რიცხვი 30-წელზე მეტი ასაკით. $Y \equiv \text{Bin}(20, 0.3)$ და $X + Y = 20$.

ამოცანა X -ის შესახებ გადავიყვანოთ ამოცანაში Y -ის შესახებ. $X \leq 10$ ნიშნავს, რომ $X = 0, 1, \dots, 10$ და ამიტომ $Y = 20, 19, \dots, 11$. $X \leq 10$ -ზე ეკვივალენტურია $Y \geq 11$ -ის; ბინომური განაწილების ცხრილის გამოყენებით $P(Y \geq 11) = 0.0171$.

თუ n საკმაოდ დიდიდანარმატების p ალბათობა არ არის ახლოს 1-თან ან 0-თან (მაგალითად, $np > 10$ და $n(1-p) > 10$), მაშინ რეკომენდებულია ნორმალური მიახლოების გამოყენება:

$$P\{a \leq \text{Bin}(n, p) \leq b\} \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right).$$

ზომიერად დიდი n -ისათვის უფრო ზუსტ მიახლოებას იძლევა ფორმულა:

$$P\{a \leq \text{Bin}(n, p) \leq b\} \approx \Phi\left(\frac{b - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np + 0.5}{\sqrt{np(1-p)}}\right).$$

თუ $n > 50$ და $np < 5$, მაშინ რეკომენდებულია პუასონის მიახლოება პარამეტრით $\lambda = np$. პუასონის შემთხვევითი სიდიდე ($Po(\lambda)$) პარამეტრით $\lambda > 0$, ლეზლობს არაუარყოფით მთელ მნიშვნელობებს ალბათობებით:

$$P\{Po(\lambda) = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$$

კუმულატური პუასონის განაწილების ($P\{Po(\lambda) \leq x\}$ ალბათობების) ცხრილი მოყვანილია დანართში 2, საიდანაც:

$$P\{Po(\lambda) = x\} = P\{Po(\lambda) \leq x\} - P\{Po(\lambda) \leq x-1\}.$$

ბინომური შემთხვევითი სიდიდის საშუალო და დისპერსია

ცდის მრავალჯერ განმეორებისას, ბუნებრივია, დავსვათ კითხვა: საშუალოდ რამდენი "წარმატებაა" მოსალოდნელი?

მაგალითი 7.6. ვთქვათ, ბევრი სტუდენტი 100-ჯერ აგდება მონეტას. გერბის მოსვლის ალბათობა 0.5-ის ტოლია. ცხადია, ადამიანი ელოდება, რომ 100 აგდებაში გერბი საშუალოდ 50-ჯერ დაჯდება (ზოგჯერ მეტ-ნაკლებად). მიაქციეთ ყურადღება, რომ $50 = 100 \cdot 0.5$.

წარმატების ალბათობა უდრის ცდათა ნილს, რომლებშიც ჩვენ ველოდებით წარმატებას. წარმატებათა რიცხვის საშუალო არის ცდათა რიცხვის ნილი.

თუ $X \equiv \text{Bin}(n, p)$, მაშინ X -ის საშუალო არის $\mu = np$. X -ის დისპერსია არის

$$\sigma^2 = np(1-p).$$

მაგალითი 7.7. საშუალოდ რამდენჯერ დაჯდება გერბი, თუ მონეტას 30-ჯერ ავაგდებთ და ამ ცდას ბევრჯერ გავიმეორებთ? რას უდრის სტანდარტული გადახრა?

საშუალო $\mu = 30 \cdot 0.5 = 15$, დისპერსია უდრის $\sigma^2 = 30 \cdot 0.5 \cdot (1 - 0.5) = 7.5$, სტანდარტული გადახრა კი იქნება $\sigma = \sqrt{7.5} = 2.74$.

ფარდობითი სიხშირეები

ფარდობითი სიხშირის მოსაძებნად წარმატებათა რაოდენობა უნდა გავყოთ ცდათა რაოდენობაზე.

მაგალითი 7.8. ვთქვათ, პოპულაციის 10% ცაციაა. შემთხვევით შეარჩიეს 50 სტუდენტი და დათვალეს რამდენი ცაციაა მათ შორის.

$X \equiv \text{Bin}(50, 0.1)$. X -ის საშუალო $50 \cdot 0.1 = 5$ -ის ტოლია, ამიტომ, თუ შევარჩევთ 50-კაციანი ჯგუფების დიდ რაოდენობას, უნდა ველოდოთ, რომ საშუალოდ ცაციების რიცხვი 5-ის ტოლი იქნება, ცაციების ფარდობითი სიხშირე კი $5 / 50 = 0.1$ -ის ტოლია.

შერჩევის ფარდობითი სიხშირის საშუალოა p და დისპერსია $p(1 - p) / n$.

მაგალითი 7.9. ცაციების ფარდობითი სიხშირის საშუალოა $p = 0.1$, დისპერსია კი:

$$0.1 \cdot 0.9 / 50 = 0.0018.$$

შეჯამება

ბინომური განაწილება დისკრეტული განაწილებაა, რომელიც აღწერს ორშედეგიან დამოუკიდებელ ცდათა სერიას, სადაც ცდათა რიცხვი ფიქსირებული რიცხვია, შედეგების ალბათობა კი უცვლელი.

ბერნულის ფორმულა:

$$P(X = x) = C_n^x p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n - 1, n.$$

უალბათესი რიცხვი x_0 წარმოადგენს შემდეგი უტოლობის მთელ ამონახსნს:

$$np - (1 - p) \leq x_0 \leq np + p.$$

ნორმალური მიახლოება:

$$P\{a \leq \text{Bin}(n, p) \leq b\} \approx \Phi\left(\frac{b - np}{\sqrt{np(1 - p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1 - p)}}\right).$$

პუასონის შემთხვევითი სიდიდე პარამეტრით $\lambda > 0$, ლეზულობს არაუარყოფით მთელ მნიშვნელობებს ალბათობებით:

$$P\{Po(\lambda) = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$$

ძირითადი ტერმინები:

ბინომური განაწილება, წარმატება, მარცხი, პუასონის შემთხვევითი სიდიდე, ბინომური შემთხვევითი სიდიდის საშუალო, დისპერსია, უალბათესი რიცხვი, ნორმალური მიახლოება.

კითხვები

1. ალბათობა იმისა, რომ მსხვილფეხა რქოსანი პირუტყვი ტუბერკულოზითაა დაავადებული, უდრის 0.05-ს. სხვადასხვა ფერმიდან შემთხვევით შეარჩიეს 5 ძროხა. აღვნიშნოთ X -ით ტუბერკულოზით დაავადებული ძროხების რაოდენობა. ააგეთ X -ის განაწილება.
 - ა. რატომ არის აუცილებელი, რომ სხვადასხვა ფერმის ძროხები იყოს შემონ-მებული?
 - ბ. რატომ არის აუცილებელი, რომ ფერმიდან ერთზე მეტი ძროხა არ უნდა იყოს შემონმებული?
2. 2003 წლის მონაცემებით, მუშების 45%-ს ქალები შეადგენდნენ. რას უდრის ალ-ბათობა იმისა, რომ შემთხვევით შერჩეული 20 მუშიდან:
 - ა. 6 ქალია?
 - ბ. 12 ან მეტი კაცია?
3. დეპრესიაში მყოფი პაციენტების 60% კლასიფიცირებულები იყვნენ როგორც ინ-ტრავერტები, 40% კი – როგორც ექსტრავერტები. X -ით აღვნიშნოთ 15 შემ-თხვევით შერჩეულ პაციენტში ექსტრავერტების რაოდენობა.
 - ა. რა პირობები უნდა იყოს შესრულებული, რომ X -ს ბინომური განაწილება ჰქონდეს?
 - ბ. იპოვეთ ალბათობები: $P(X = 5)$, $P(X \leq 5)$ და $P(X \geq 12)$.
4. ცნობილია, რომ ლოკოკინების (*Limicolaria martensiana*) 60% ზოლებიანია. იპოვეთ: შემთხვევით შერჩეულ 20 ლოკოკინაში რამდენი შეიძლება აღმოჩნდეს ზოლები-ანი და რა ალბათობით?
5. ბანკის მიერ გაცემული კრედიტების 5% არ ბრუნდება. ვიპოვოთ ალბათობა იმი-სა, რომ ბანკის მიერ გაცემულ 20 კრედიტიდან დაბრუნების პრობლემა შეიქმნება არანაკლებ ორ შემთხვევაში. იგულისხმება, რომ კრედიტები გაიცემა და ბრუნ-დება ერთმანეთისაგან დამოუკიდებლად.
6. გარკვეული პროდუქციის დაზიანებულის წილი შეადგენს 1%-ს. პარტია შედგება 300 ნაწარმისგან.
 - ა. რამდენი შეიძლება იყოს დაზიანებული?
 - ბ. გამოთვალეთ რისი ტოლია დისპერსია და სტანდარტული გადახრა.
7. ბანკის კლიენტების 15% კორპორატიულ ბიზნესშია ჩართული. იგულისხმება, რომ ბინომური განაწილებისთვის საჭირო ყველა პირობა დაცულია.
 - ა. იპოვეთ ბანკის 1000 კლიენტიდან კორპორატიული ბიზნესის კლიენტების მო-სალოდნელი რაოდენობა, დისპერსია და სტანდარტული გადახრა;
 - ბ. ბინომური განაწილების ცხრილის გამოყენებით ააგეთ კორპორატიული ბიზ-ნესის კლიენტების ალბათური განაწილება 10 კლიენტის შემთხვევაში.
8. სკოლის ადმინისტრაციის ცნობით, სასკოლო სკამების 15% ყოველ წელიწადს შესაკეთებელია. სკოლას სთავაზობენ ახალი ტიპის სკამებს. მენეჯერის ცნობით, ახალი ტიპის სკამების მხოლოდ 10% იქნება ერთ წელიწადში შესაკეთებელი. თუ სკოლაში 1000 სკამია, საშუალოდ რამდენი სკამი იქნება შესაკეთებელი თითოეუ-ლი ტიპის სკამების პირობებში. თუ ერთი სკამის შეკეთება 5 ლარი ღირს, რა თან-ხა უნდა იყოს გამოყოფილი თითოეულ შემთხვევაში?
9. კოლეჯის ადმინისტრაციის ცნობით, ყოველ წელიწადს მათი ფრიადოსანი კურ-სდამთავრებულებს 90% სჯიმდება.
 - ა. ვთქვათ, კურსზე 5 წარჩინებული სტუდენტია, ააგეთ დასაქმებული სტუდენ-ტების ალბათური განაწილება;

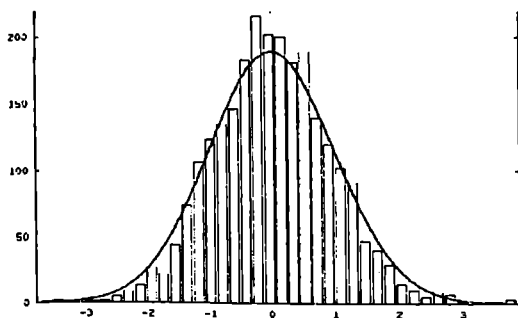
- ბ. იპოვეთ ამ განაწილების საშუალო და სტანდარტული გადახრა;
- გ. თუ შარშან 5 სტუდენტიდან 4-მა მიიღო სამსახური, ეთანხმება თუ არა ეს ფაქტი იმ მოსაზრებას, რომ კურსდამთავრებულთა 90% საქმდება.
10. ფირმა ყიდულობს რძის ბოთლებს შემდგომი გამოყენებისათვის. ბოთლების 80% არ საჭიროებს სკრაბით გარეცხვას. ვიგულისხმობ, რომ საქმე გვაქვს ბინომურ განაწილებასთან. იპოვეთ:
- ა. 200 ბოთლიდან საშუალოდ რამდენს არ დასჭირდება სკრაბით გარეცხვა?
- ბ. 100 ბოთლიდან საშუალოდ რამდენს დასჭირდება სკრაბით გარეცხვა?
- გ. ერთი ბოთლის სკრაბით გარეცხვა ჯდება 5 თეთრი. საშუალოდ რა თანხაა საჭირო ასეთი ბოთლების გასარეცხად, თუ შერჩევამ 1000 ბოთლია?
11. როგორია ალბათობა იმისა, რომ სათამაშო კამათლის ოთხჯერ გაგორებისას ერთიანი მოვა არაუმეტეს ორჯერ?
12. როგორია ალბათობა იმისა, რომ სათამაშო კამათლის ოთხჯერ გაგორებისას სამჯერ მოვა წახნაგი ლუნი ქულით?
13. როგორია ალბათობა იმისა, რომ სათამაშო კამათლის ოთხჯერ გაგორებისას ორჯერ მოვა წახნაგი 3-ის არაჯერადი ქულით?
14. ალბათობა იმისა, რომ მოცემულ სანარმოში ელექტროენერგიის დღიური დანახარჯი არ გადააჭარბებს დადგენილ ნორმას, ტოლია 0.8-ის. როგორია ალბათობა იმისა, რომ კვირის განმავლობაში ელექტროენერგიის დანახარჯი არ გავა ნორმიდან ოთხი დღის განმავლობაში?
15. მსროლელის მიერ სამიზნის დაზიანების ალბათობაა 0.8. როგორია ალბათობა იმისა, რომ 80 გასროლის შედეგად მსროლელი დაზიანებს სამიზნეს: ა) ზუსტად 60-ჯერ? ბ) ზუსტად 64-ჯერ? გ) 60-დან 70-მდე? დ) არანაკლებ 60-ჯერ?
16. თესლის მოცემული პარტიიდან დათესეს 10000 თესლი და ნახეს, რომ აღმოცენდა 8498 თესლი. ამის შედეგად გაკეთდა დასკვნა, რომ თესლის მოცემული პარტიის აღმოცენებადობა შეადგენს 85%-ს ($p = 0.85$). ეიპოვოთ ალბათობა იმისა, რომ მოცემული პარტიიდან დათესილი 100 თესლიდან აღმოცენდება: ა) ზუსტად 85 თესლი? ბ) 60-დან 80 თესლამდე? გ) 70-დან 90 თესლამდე? დ) არანაკლებ 80 თესლი? ე) არანაკლებ 85 თესლი? ვ) არანაკლებ 90 თესლი? ზ) არანაკლებ 92 თესლი?
17. დავუშვათ, რომ გარკვეული ეპიდემიის დროს დაავადების ალბათობაა $p = 0.005$. როგორია ალბათობა იმისა, რომ 1000 ადამიანიდან დაავადდება: ა) არა უმეტეს 2 ადამიანი? ბ) არა უმეტეს 5 ადამიანი? გ) არა უმეტეს 10 ადამიანი? დ) არანაკლებ 3 ადამიანი?
18. პროფილაქტიკური ღონისძიებების გატარებამ მოსალოდნელი ეპიდემიის დროს დაავადების ალბათობა შეამცირა $p_1 = 0.005$ -დან $p_2 = 0.003$ -მდე. ამის შედეგად რამდენჯერ გაიზარდა ალბათობა იმისა, რომ 1000 ადამიანიდან დაავადდება: ა) არა უმეტეს 5 ადამიანი? ბ) არა უმეტეს 2 ადამიანი? გ) არავინ არ დაავადდება?
19. წყალქვეშა ნავი ესვრის კრეისერს 4-ჯერ თითო-თითო ტორპედოს. თითოეული გასროლისას ტორპედოს მოხვედრის ალბათობაა $3/4$. ყოველ ტორპედოს ერთნაირი ალბათობით შეუძლია გახვრიტოს კრეისერის 10 განყოფილებიდან რომელიმე განყოფილება, რომელიც დაზიანების შედეგად წყლით ივსება. კრეისერი იძირება, თუ წყლით აივსო ორი განყოფილება მაინც. გამოვთვალოთ კრეისერის ჩაძირვის ალბათობა.

ჩინა თავში ჩვენ განვიხილეთ ბინომური განაწილება – განაწილება, რომელიც აღწერს ბინომურ შემთხვევით სიდიდეს X' , სადაც X' – „წარმატებათა“ რაოდენობაა „დამოუკიდებელ ცდაში“. „ცდათა რაოდენობა – ფიქსირებული რიცხვია. თითოეულ ცდაში მოსალოდნელია მხოლოდ ორი შედეგი – „წარმატება“ ან „მარცხი“. „წარმატების“ ალბათობა p – მუდმივი რიცხვია, არ იცვლება ცდიდან ცდამდე. X – დისკრეტული შემთხვევითი სიდიდე იღებს მხოლოდ მთელ მნიშვნელობებს $0, 1, 2, \dots, n$.

ბევრი შემთხვევითი სიდიდე, როგორიცაა: წონა, დრო, მანძილი, ასაკი არის უწყვეტი. უწყვეტი შემთხვევითი სიდიდის მნიშვნელობა დამოკიდებულია იმაზე, თუ რა სიზუსტით ვზომავთ მას. ასე, მაგალითად, გაზომილი მანძილი – 4 კმ, თუ მას გავზომავთ მეტრების სიზუსტით, შეიძლება აღმოჩნდეს – 4012 მეტრი = 4.012კმ და 401248სმ = 4.01248კმ.

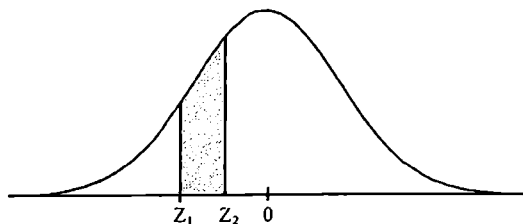
ერთ-ერთ მნიშვნელოვან განაწილებას, რომელიც აღწერს უწყვეტ შემთხვევით სიდიდეს, წარმოადგენს ნორმალური განაწილება. ეს განაწილება შესწავლილი იყო გერმანელი მათემატიკოს გაუსის მიერ და ზოგჯერ მოხსენებულია გაუსის განაწილების სახელით. ნორმალური განაწილების გრაფიკს აქვს ზარისებური ფორმა, ის სიმეტრიულია საშუალოს მიმართ და აქვს მხოლოდ ერთი, საშუალოს ტოლი, მოდა და მედიანა.

თუ ჩვენ გავზომავთ ნორმალურად განაწილებულ შემთხვევით სიდიდეს და ავაგებთ მიღებული მნიშვნელობების ფარდობითი სიხშირეების ჰისტოგრამას, შესაბამის სიხშირეთა პოლიგონს ექნება ზარისებული ფორმა. ამ სიტუაციაში პოლიგონი წარმოადგენს ნორმალური განაწილების გრაფიკს. ჩვენ ვიცით, რომ ჰისტოგრამის ყველა მართკუთხედის ფართობების ჯამი უდრის ერთს (ფარდობით სიხშირეთა ჯამი უდრის ერთს), ეს კი ნიშნავს, რომ პოლიგონის ქვეშ მდებარე ფართობიც უნდა უდრიდეს ერთს (ნახ. 8.1).



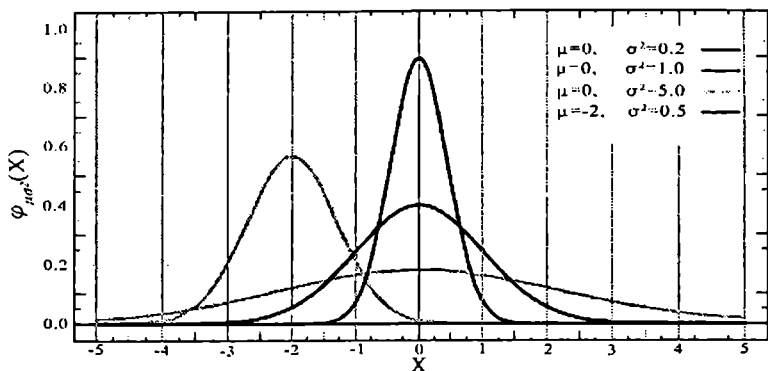
ნახ. 8.1

ჩვენ შეგვიძლია გამოვიყენოთ ნორმალური განაწილების გრაფიკი იმისათვის, რომ გამოვთვალოთ $P(z_1 \leq Z \leq z_2)$ (ნახ. 8.2). გამუქებული ფართობი გრაფიკის ქვეშ, შემოსაზღვრული ვერტიკალური წრფეებით z_1 -სა და z_2 -ში, უდრის $P(z_1 \leq Z \leq z_2)$. თუ X -ს აქვს ნორმალური განაწილება საშუალოთი μ და დისპერსიით σ^2 , ეს ფაქტი აღინიშნება როგორც $X \equiv N(\mu, \sigma^2)$.



ნახ. 8.2

საშუალო μ და დისპერსია σ^2 – ნორმალური განაწილების პარამეტრებია. ნორმალურად განაწილებული შემთხვევითი სიდიდის სიშირეთა განაწილება შეიძლება სრულად აღინეროს, თუ ცნობილია განაწილების საშუალო μ და სტანდარტული გადახრა σ .



ნახ. 8.3

ალბათობების გამოსათვლელად არ არის აუცილებელი ბევრი ცხრილის შექმნა μ და σ სხვადასხვა წყვილისათვის, რადგანაც ნებისმიერი ნორმალური განაწილება შეიძლება დაყვანილ იქნეს ე.წ. სტანდარტულ ნორმალურ განაწილებაზე $N(0,1)$, ანუ ნორმალურ განაწილებაზე საშუალოთი 0 და სტანდარტული გადახ-

რით 1. ამისათვის საჭიროა მოვახდინოთ X -ის ე.წ. სტანდარტიზაცია ფორმულით:

$$Z = \frac{X - \mu}{\sigma}.$$

თუ $X \cong N(\mu, \sigma^2)$, მაშინ $Z = \frac{X - \mu}{\sigma} \cong N(0, 1)$. Z -ს აქვს სტანდარტული ნორმალური განაწილება ნულის ტოლი საშუალოთი და ერთის ტოლი დისპერსიით.

ნორმალური განაწილების თვისებები:

- ✓ ინტერვალი $(\mu - 1\sigma, \mu + 1\sigma)$ მოიცავს მონაცემთა 68.3%;
- ✓ ინტერვალი $(\mu - 2\sigma, \mu + 2\sigma)$ მოიცავს მონაცემთა 95.4%;
- ✓ ინტერვალი $(\mu - 3\sigma, \mu + 3\sigma)$ მოიცავს მონაცემთა 99.7%.

სტანდარტული ნორმალური განაწილების თვისებები: ($\mu = 0, \sigma^2 = 1$)

- ✓ ინტერვალი $(-1, 1)$ მოიცავს მონაცემთა 68.3%;
- ✓ ინტერვალი $(-2, 2)$ მოიცავს მონაცემთა 95.4%;
- ✓ ინტერვალი $(-3, 3)$ მოიცავს მონაცემთა 99.7%.

სტანდარტული ნორმალური განაწილების ცხრილის გამოყენება

მონაცემთა სტანდარტიზაცია ანუ z („ზეტ“) ქულებში გადაყვანის უპირატესობა ისაა, რომ z -ის მნიშვნელობებს შორის მოხვედრის ალბათობები გამოთვლილია და მოცემულია დანართის სტანდარტული ნორმალური განაწილების ცხრილში.

განაწილების მრუდის ქვეშ მდებარე ფართობი უდრის 1-ს, განაწილების გრაფიკი სიმეტრიულია, ამიტომ, ფართობი მოდის (საშუალო, მედიანა, მოდა ერთმანეთის ტოლია) მარცხნივ და მარჯვნივ (სიმეტრიულობის გამო) 0.5-ის ტოლია. სტანდარტული ნორმალური განაწილების ცხრილში – $\Phi(z)$ ფუნქციის ცხრილში მოცემულია $P(-\infty < Z \leq z)$ (შესაბამისად, $\Phi_c(z)$ ფუნქციის ცხრილში მოცემულია $P(0 \leq Z \leq z)$) ალბათობები $z = 0.00, 0.01, \dots, 3.99$ მნიშვნელობებისათვის. თუ $z > 0$, მაშინ

$$\Phi(z) = 0.5 + \Phi_c(z) \text{ და } \Phi(-z) = 1 - \Phi(z).$$

მაგალითი 8.1. გამოვთვალოთ $P(0 \leq Z \leq 1.56)$, $P(0 \leq Z \leq 1.8)$.

ამოხსნა: იმისათვის, რომ ვიპოვოთ ალბათობა იმისა, რომ Z არის 0-სა და 1.56-ს შორის, ცხრილის პირველ სვეტში უნდა ვიპოვოთ 1.5, ხოლო პირველ სტრიქონში 0.06, $(1.5 + 0.06 = 1.56)$ და გადაკვეთაზე მოცემული რიცხვი 0.4406 იქნება საძიებელი ალბათობა – $P(0 \leq Z \leq 1.56) = 0.4406$. მეორე შემთხვევაში, ცხრილის პირველ სვეტში უნდა ვიპოვოთ 1.8, ხოლო პირველ სტრიქონში 0.00 და გადაკვეთაზე მოცემული რიცხვი 0.4641 იქნება საძიებელი ალბათობა – $P(0 \leq Z \leq 1.8) = 0.4641$.

ხშირად მოსაძებნია $P(-z \leq Z \leq 0)$ ალბათობები. ნორმალური მრუდის სიმეტრიულობის გამო $P(-z \leq Z \leq 0) = P(0 \leq Z \leq z)$. შევნიშნავთ აგრეთვე, რომ არამკაცრი უტოლობის მკაცრი უტოლობით შეცვლა არ ცვლის შესაბამისი ფიგურის ფართობს და, მაშასადამე, არ ცვლის შესაბამის ალბათობას, მაგალითად, $P(0 \leq Z \leq z) = P(0 < Z < z)$.

მაგალითი 8.2. $P(-2 < Z < 0) = P(0 < Z < 2) = 0.4772$;

$$P(Z > 1.3) = P(0 \leq Z < \infty) - P(0 \leq Z \leq 1.3) = 0.5 - 0.4032 = 0.0968$$
;

$$P(-1 < Z < z) = P(-1 < Z \leq 0) + P(0 < Z < 2) = P(0 \leq Z < 1) + 0.4772 = 0.3413 + 0.4772 = 0.8185$$
;

$$P(Z < 1.57) = P(-\infty < Z \leq 0) + P(0 < Z < 1.57) = 0.5 + 0.4418 = 0.9418$$
;

$$P(1 < Z < 2) = P(0 \leq Z < 2) - P(0 \leq Z \leq 1) = 0.4772 - 0.3413 = 0.1359$$
.

ეთქვათ, $X \equiv N(12, 16)$, ანუ X ნორმალურადაა განაწილებული 12-ის ტოლი საშუალოთი, 16-ის ტოლი დისპერსიით და სტანდარტული გადახრით $\sqrt{16} = 4$. როგორ ვიპოვოთ $P(X > 14)$?

პირველ რიგში უნდა მოვახდინოთ სტანდარტიზაცია და შემდეგ მიღებული Z -ის მეშვეობით ვიპოვოთ საძიებელი ალბათობა:

$$P(X > 14) = P\left(\frac{X - 12}{4} > \frac{14 - 12}{4}\right) = P(Z > 0.5) = 0.5 - 0.1915 = 0.3085$$
.

ყველა განხილულ მაგალითში მოცემული z -თვის სტანდარტული ნორმალური განაწილების ცხრილის დახმარებით ვპოულობდით ალბათობებს. ხშირად საჭიროა შებრუნებული ამოცანა: მოცემული ალბათობით z სიდიდის პოვნა.

მოცემული ალბათობით z სიდიდის პოვნა.

მაგალითი 8.3. ვიპოვოთ ისეთი z , რომ $P(Z < z) = 0.75$ (ან, რაც იგივეა $P(Z \geq z) = 1 - 0.75 = 0.25$).

ამოხსნა: ცხადია, რომ $0.75 = P(Z < z) = P(-\infty < Z < 0) + P(0 \leq Z < z) = 0.5 + \Phi_*(z)$ (შესაბამისად, $0.25 = P(Z \geq z) = P(0 \leq Z < \infty) - P(0 \leq Z < z) = 0.5 - \Phi_*(z)$). ამიტომ, ამოცანა დაიყვანება ისეთი z -ის მოძებნაზე, რომ $\Phi_*(z) = 0.75 - 0.5 = 0.25$ (შესაბამისად, $\Phi_*(z) = 0.5 - 0.25 = 0.25$). სტანდარტული ნორმალური განაწილების ცხრილში მოძებნოთ 0.25-თან უახლოესი მნიშვნელობები (ალბათობები), ესენია: 0.2486 და 0.2517. მოძებნოთ ამ ალბათობების შესაბამისი არგუმენტები, ესენია: 0.67 და 0.68, ანუ $\Phi_*(0.67) = 0.2486$ და $\Phi_*(0.68) = 0.2517$. ამიტომ, საძიებელი z -ის მიახლოებითი მნიშვნელობად გამოდგება როგორც 0.67, ისე 0.68. უფრო ზუსტი

მიახლოება იქნება ამ მნიშვნელობების საშუალო არითმეტიკული, კერძოდ $(0.67 + 0.68) / 2 = 0.675$.

მაგალითი 8.4. ოტიტი (ოტიტის მედია) – ყურის გავრცელებული დაავადებაა ბავშვებში. ეს დაავადება ხშირად მეორდება და იწვევს სმენის დაქვეითებას. დადგენილი იყო, რომ ასაკი, როცა ეს დაავადება პირველად ფიქსირდება, ნორმალურად განაწილებული შემთხვევითი სიდიდეა საშუალოთი – $\mu = 3.4$ წელი და სტანდარტული გადახრით – $\sigma = 1.2$ წელი.

- იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული ბავშვისთვის ეს ასაკი ნაკლებია ან ტოლი 4 წელზე;
- დაავადების პირველად დაფიქსირების ასაკი ბავშვების რამდენ პროცენტშია 5 წელზე მეტი ან ტოლი?
- რომელ ასაკში ემართებათ ბავშვების 70%-ს პირველად ოტიტი?
- როგორ გამოვიყენოთ მიღებული შედეგები?

ამოხსნა: X – ასაკი, როცა ბავშვი პირველად გახდა ავად, განაწილებულია ნორმალურად $X \equiv N(3.4; 1.2^2)$. საძიებელი ალბათობებია:

- $P(X \leq 4) = P\left(\frac{X - 3.4}{1.2} \leq \frac{4 - 3.4}{1.2}\right) = P(Z \leq 0.5) = 0.5 + \Phi_0(0.5) = 0.5 + 0.1915 = 0.6915$;
- $P(X \geq 5) = P\left(\frac{X - 3.4}{1.2} \geq \frac{5 - 3.4}{1.2}\right) = P(Z \geq 1.33) = 0.5 - \Phi_0(1.33) = 0.5 - 0.4082 = 0.0918$;
- დავუშვათ, რომ საძიებელი ასაკია – x . მაშინ გვაქვს: $P(X \leq x) = \frac{70\%}{100\%} = 0.7$.

შესაბამისად,

$$0.7 = P(X \leq x) = P\left(\frac{X - 3.4}{1.2} \leq \frac{x - 3.4}{1.2}\right) = P\left(Z \leq \frac{x - 3.4}{1.2}\right) = 0.5 + \Phi_0\left(\frac{x - 3.4}{1.2}\right).$$

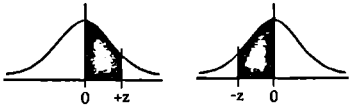
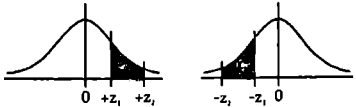
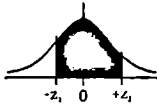
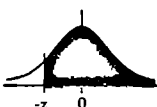
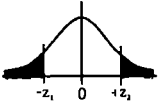
აქედან $\Phi_0\left(\frac{x - 3.4}{1.2}\right) = 0.7 - 0.5 = 0.2$. ამიტომ, 8.3 მაგალითის მიხედვით, ნორმა-

ლური განაწილების ცხრილიდან დავასკენით, რომ $\frac{x - 3.4}{1.2} = \frac{0.52 + 0.53}{2} = 0.525$, საიდანაც ვღებულობთ:

$$x = 1.2 \cdot 0.525 + 3.4 = 4.03.$$

- ვინაიდან ბ. პუნქტის მიხედვით $P(X \geq 5) = 0.0918$, ამიტომ $P(X < 5) = 1 - 0.0918 = 0.9082$, ანუ ბავშვების $0.9082 \cdot 100\% \approx 90\%$ -ს ოტიტი პირველად ემართება 5 წელზე ნაკლებ ასაკში. შესაბამისად, პროფილაქტიკური ღონისძიებები უნდა ჩატარდეს ბავშვის სკოლაში შესვლამდე.

სტანდარტული ნორმალური განაწილების წირის ქვეშ მდებარე
არის ფართობის მოსაძებნი პროცედურული ცხრილი

1. 0-სა და ნებისმიერ Z-ს შორის: ნორმალური განაწილების ცხრილიდან ვპოულობთ $\Phi_0(z)$-ს. 	2. ნებისმიერი კუდის შემთხვევა: ა. ვნახოთ $\Phi_0(z)$-ის მნიშვნელობა ცხრილში. ბ. 0.5000-ს გამოვაკლოთ $\Phi_0(z)$. 
3. Z-ის ორ მნიშვნელობას შორის საშუალოს ერთ მხარეს: ა. ვიპოვოთ $\Phi_0(z_1)$ და $\Phi_0(z_2)$. ბ. დიდს გამოვაკლოთ პატარა. 	4. Z-ის ორ მნიშვნელობას შორის საშუალოს სხვადასხვა მხარეს: ა. ვიპოვოთ $\Phi_0(z_1)$ და $\Phi_0(z_2)$. ბ. შევკრიბოთ ისინი. 
5. საშუალოზე მეტი Z-ის მარცხნივ: ა. ცხრილში ვპოულობთ $\Phi_0(z)$-ს. ბ. ვუმატებთ მას 0.5000-ს. 	6. საშუალოზე ნაკლები $-Z$-ის მარჯვნივ: ა. ცხრილში ვპოულობთ $\Phi_0(z)$-ს. ბ. ვუმატებთ მას 0.5000-ს. 
7. ორი კუდის შემთხვევა: ა. ვიპოვოთ $\Phi_0(z_1)$ და $\Phi_0(z_2)$. ბ. 0.5000-ს გამოვაკლოთ თითოეული. გ. შევკრიბოთ სხვაობები. 	პროცედურა - დავხაზოთ ნახაზი. - მოვნიშოთ სასურველი არე. - ვიპოვოთ შესაბამისი ბლოკი. - ვიმოკმედოთ ბლოკის სქემით. შენიშვნა: $\Phi(z) = \Phi_0(z) + 0.5000$, როცა $z > 0$.

ცენტრალური ზღვარი თეორემა (ცზთ)

რეალური პოპულაციის მონაცემები არასდროს არის ისე ზუსტად განაწილებული, რომ განაწილება ზუსტად სიმეტრიული, ნორმალური იყოს. მაგრამ, რატომ არის ნორმალური განაწილება ასეთი მნიშვნელოვანი? ამას ხსნის ცენტრალური ზღვარი თეორემა (ცზთ): *ბევრი დამოუკიდებელი და ერთნაირად განაწილებული შემთხვევითი სიდიდეების ჯამი ან საშუალო დაახლოებით ნორმალურადაა განაწილებული.*

იმისათვის, რომ ცზთ სრულდებოდეს, არ არის ლაპარაკი შემთხვევითი სიდიდეების კონკრეტულ განაწილებებზე. საჭიროა მხოლოდ ის, რომ შემთხვევითი სიდიდეები დამოუკიდებელი, ერთნაირად განაწილებული და საკმაოდ ბევრი იყოს. თუ ჩვენ გვაქვს ბევრი მონაცემი, მათი საშუალო ნორმალურად იქნება განაწილებული.

მაგალითი 8.5. ყუთში 100 ერთნაირი ბურთულაა, გადანომრილი 100-დან 199-მდე. ყუთში ჩაუხედავად ვიღებთ ბურთულას, ვინერთ მის ნომერს და უკან ვაბრუნებთ. შერჩევის მოცულობა 30-ის ტოლია. ვთქვათ, ასეთი წესით მივიღეთ 1000 შერჩევა. პირველი 4 შერჩევა წარმოდგენილია 4, ფოთლებიანი ლეროების მსგავსი დიაგრამით:

შერჩევა 1		შერჩევა 2		შერჩევა 3		შერჩევა 4	
ლერო	ფოთოლი	ლერო	ფოთოლი	ლერო	ფოთოლი	ლერო	ფოთოლი
10	0000599	10	0189	10	688	10	12467
11	0199	11	3567	11	178	11	56
12	46	12	148	12	26889	12	22369
13	125	13	8	13	28	13	224
14		14	4	14	09	14	224
15	1568	15	18	15	78	15	34
16	47789	16	0029	16	002239	16	06788
17	1	17	0779	17	18	17	01
18	27	18	06	18	67	18	4
19	02	19	55899	19	778	19	35

ნახ. 8.4. ოთხი შერჩევის ფოთლებიანი ლეროების მსგავსი დიაგრამა. შერჩევებში 30 შემთხვევითი ამოღებული ბურთულების ნომრებია 100-დან 199-მდე.

პირველი შერჩევის 30 ნომრის (მონაცემის) საშუალოა – $\bar{x} = 140.23$;

მეორე შერჩევის საშუალოა – $\bar{x} = 151.63$;

მესამე შერჩევის საშუალოა – $\bar{x} = 148.87$;

მეოთხე შერჩევის საშუალოა – $\bar{x} = 141.6$.

ყველა მიღებული საშუალო ბევრად არ განსხვავდება პოპულაციის საშუალოსგან, რომელიც $(100 + 101 + 102 + \dots + 198 + 199) / 100 = 149.5$ -ის ტოლია. 1000

შერჩევის საშუალოებით აგებული პოლიგონი ძალიან ახლოს არის ნორმალური განაწილების მრუდთან, $\mu = 149.5$ -ის ტოლი საშუალოთი. ცხთ თანახმად, უნდა ყოფილიყო ზუსტად ასეთი სურათი.

ცხთ – მუშაობს! პოპულაცია არ იყო ნორმალური, მაგრამ საშუალოების განაწილება დაემსგავსა ნორმალურს.

რამდენად დიდი უნდა იყოს შერჩევის მოცულობა n , რომ სრულდებოდეს ცხთ?

ეს დამოკიდებულია განაწილების ფორმაზე. თუ განაწილება სიმეტრიულია, საჭიროა ნაკლები n . საკმაოდ ბევრი განაწილებისათვის საკმარისია $n = 30$ იმისათვის, რომ დამოუკიდებელი, ერთნაირად განაწილებული შემთხვევითი სიდიდეების ჯამის განაწილება ახლოს იყოს ნორმალურ განაწილებასთან.

შერჩევითი განაწილებები

ვიგულისხმობთ, რომ შერჩევის მოცულობა მეტია ან ტოლი 30-ზე. მაშინ, თუ პოპულაციიდან, რომლის საშუალოა μ და სტანდარტული გადახრა σ , ავიღებთ ბევრ შერჩევას, გამოვთვლით ამ შერჩევების საშუალოებს და ავაგებთ ამ საშუალოების განაწილებას. მიღებული განაწილების ფორმა ახლოს იქნება ნორმალური განაწილების ფორმასთან.

შერჩევების საშუალოების საშუალო არის პოპულაციის საშუალო μ და სტანდარტული გადახრა $\sigma/\sqrt{n}$. სტატისტიკების (მაგალითად, შერჩევის საშუალოების) განაწილებებს ეწოდებათ *შერჩევითი განაწილებები*.

საშუალოების შერჩევითი განაწილების სტანდარტულ გადახრას ($\sigma/\sqrt{n}$ -ს) ეწოდება *საშუალოს სტანდარტული შეცდომა*.

$\sigma/\sqrt{n}$ – საშუალოს სტანდარტული შეცდომაა.

თუ დამოუკიდებელი შემთხვევითი სიდიდეები $X_1, X_2, \dots, X_n$ ნორმალურადაა განაწილებული საშუალოთი μ და სტანდარტული გადახრით σ , მაშინ მათი საშუალო $\bar{X}$ ნორმალურადაა განაწილებული საშუალოთი μ და სტანდარტული გადახრით $\sigma/\sqrt{n}$ (n ნებისმიერი ნატურალური რიცხვია).

თუ დამოუკიდებელი შემთხვევითი სიდიდეები $X_1, X_2, \dots, X_n$ ერთნაირად არიან განაწილებული საშუალოთი μ და სტანდარტული გადახრით σ და $n \geq 30$ -ზე, მაშინ მათი საშუალო $\bar{X}$ ნორმალურადაა განაწილებული საშუალოთი μ და სტანდარტული გადახრით $\sigma/\sqrt{n}$.

თუ $X_1, X_2, \dots, X_n$ არიან ნორმალურები, ან $n \geq 30$, მაშინ $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ შემთხვევით

სიდიდეს აქვს სტანდარტული ნორმალური განაწილება $Z \equiv N(0,1)$.

მაგალითი 8.6. ვთქვათ, პოპულაციის საშუალო $\mu = 10$ და $\sigma = 5$. პოპულაციიდან აღებული შერჩევის მოცულობა $n = 100$. გამოვთვალოთ $P(9.5 < \bar{X} < 10.5)$.

ამოხსნა: ცხთ თანახმად $\bar{X} \equiv N(10; \frac{5}{\sqrt{100}})$. ამიტომ გვაქვს:

$$P(9.5 < \bar{X} < 10.5) = P(\frac{9.5 - 10}{5/\sqrt{100}} < \frac{\bar{X} - 10}{5/\sqrt{100}} < \frac{10.5 - 10}{5/\sqrt{100}}) =$$

$$= P(-1 < Z < 1) = P(-1 < Z < 0) + P(0 \leq Z < 1) = 2\Phi_e(1) = 2 \cdot 0.3413 = 0.6826.$$

ბინომური განაწილების ნორმალური აპროქსიმაცია

ბინომურად განაწილებული X შემთხვევითი სიდიდე არის წარმატებათა რაოდენობა n დამოუკიდებელ ცდაში. წარმატებას მიუწერენ (შეუსაბამებენ) ქულა 1-ს, მარცხს კი ქულა 0-ს. ყველა ამ ქულის (ნულების და ერთების) შეკრებით მიღებული ჯამი ზუსტად X -ის ტოლი იქნება. წარმატებათა რაოდენობა X – ერთნაირად განაწილებული დამოუკიდებელი შემთხვევითი სიდიდეების ჯამია და ამიტომ შეიძლება ცენტრალური ზღვარითი თეორემის (ცხთ) გამოყენება.

ბინომურად განაწილებული შემთხვევითი სიდიდე $X \equiv \text{Bin}(n, p)$, როცა $n \geq 30$, მიახლოებით ნორმალურადაა განაწილებული საშუალოთი np და დისპერსიით $np(1-p)$.

წარმატებათა რაოდენობა არ შეიძლება იყოს ნულზე ნაკლები ან ცდათა რაოდენობაზე მეტი და თუ გავიხსენებთ, რომ ნორმალურად განაწილებული შემთხვევითი სიდიდე საშუალოდან არ უნდა იყოს დაშორებული სამ სტანდარტულ გადახრაზე მეტი მანძილით, მივიღებთ, რომ უნდა სრულდებოდეს უტოლობა:

$$0 \leq np \pm 3\sqrt{np(1-p)} \leq n.$$

ეს მოთხოვნა კი სრულდება, თუ

$$np \geq 5 \quad \text{და} \quad n(1-p) \geq 5.$$

ვინაიდან, $X \equiv \text{Bin}(n, p)$ – დისკრეტული შემთხვევითი სიდიდეა, ხოლო

$$Y \equiv N(np; np(1-p))$$

– უწყვეტი შემთხვევითი სიდიდე, ამიტომ, იმისათვის, რომ გამოვიყენოთ ნორმალური აპროქსიმაცია, უნდა მოვახდინოთ ე.წ. *კორექცია უწყვეტობაზე*, დისკრეტული სიდიდის მნიშვნელობას გამოვაკლოთ და დაეუმატოთ 0.5 და მიღებული რიცხვები გამოვაცხადოთ უწყვეტი სიდიდის მნიშვნელობათა არის საზღვრებად: $P(X = x) \approx P(x - 0.5 < Y < x + 0.5)$, სადაც $X \equiv \text{Bin}(n, p)$ – დისკრეტულია და $Y \equiv N(np; np(1-p))$ – უწყვეტი.

მაგალითი 8.7. სტატისტიკის ლექციის სტუდენტების 15% აცდენს. კურსზე 200 სტუდენტი. იპოვეთ ალბათობა იმისა, რომ:

- ლექციას დაესწრება 160 ან მეტი სტუდენტი;
- ლექციას დაესწრება 180-ზე ნაკლები სტუდენტი;
- ლექციას დაესწრება 165-ზე მეტი და 185-ზე ნაკლები სტუდენტი.

ამოხსნა: შევამოწმოთ, შეიძლება თუ არა ნორმალური აპროქსიმაციის გამოყენება. კონკრეტული სტუდენტი ლექციას აცდენს ალბათობით $p = 15\% / 100\% = 0.15$. შესაბამისად:

$$np = 200 \cdot 0.15 = 30 \geq 5 \quad \text{და} \quad n(1-p) = 200 \cdot (1-0.15) = 170 \geq 5.$$

მაშასადამე, ნორმალური (საშუალოთი $np = 30$ და სტანდარტული გადახრით $\sqrt{np(1-p)} = \sqrt{30 \cdot 0.85} = 5.05$) აპროქსიმაციის გამოყენება შესაძლებელია. ამიტომ საძიებელი ალბათობებია:

- იპოვეთ ალბათობა იმისა, რომ ლექციას დაესწრება 160 ან მეტი სტუდენტი ემთხვევა ალბათობას იმისა, რომ ლექციას გააცდენს 40-ზე ნაკლები სტუდენტი. მოვახდინოთ კორექცია უწყვეტობაზე, გვექნება:

$$\begin{aligned} P(X \geq 160) &= P(X < 40) \approx P(Y < 39.5) = P\left(\frac{Y-30}{5.05} < \frac{39.5-30}{5.05}\right) = \\ &= P\left(Z < \frac{39.5-30}{5.05}\right) = P(Z < 1.88) = 0.5 + \Phi_e(1.88) = 0.5 + 0.4699 = 0.9699. \end{aligned}$$

- ანალოგიურად, აქ უნდა გამოთვალოთ ალბათობა იმისა, რომ 20 ან მეტი სტუდენტი გააცდენს ლექციას. გვაქვს:

$$\begin{aligned} P(X < 180) &= P(X \geq 20) \approx P(Y \geq 19.5) = P\left(\frac{Y-30}{5.05} \geq \frac{19.5-30}{5.05}\right) = \\ &= P\left(Z \geq \frac{19.5-30}{5.05}\right) = P(Z \geq -2.08) = P(Z \leq 2.08) = 0.5 + \Phi_e(2.08) = 0.9812. \end{aligned}$$

- აქ უნდა გამოთვალოთ ალბათობა იმისა, რომ ლექციას არ დაესწრება 15-ზე მეტი ან ტოლი და 35 ნაკლები ან ტოლი სტუდენტი. შესაბამისად, გვექნება:

$$\begin{aligned} P(165 < X < 185) &= P(15 \leq X \leq 35) \approx P(14.5 \leq Y \leq 35.5) = P\left(\frac{14.5-30}{5.05} \leq \frac{Y-30}{5.05} \leq \frac{35.5-30}{5.05}\right) = \\ &= P(-3.07 \leq Z \leq 1.09) = P(-3.07 \leq Z \leq 0) + P(0 < Z \leq 1.09) = \\ &= \Phi_e(3.07) + \Phi_e(1.09) = 0.4989 + 0.3621 = 0.861. \end{aligned}$$

მაგალითი 8.8. სასტუმროში 250 ერთნაირი ნომერია. ალბათობა იმისა, რომ ნომერი დაკავებულია, არის 0.75. იგულისხმება, რომ ნომრებს იკავებენ ერთმანეთისგან დამოუკიდებლად. ვთქვათ, X დაკავებული ნომრების რაოდენობაა. ვიპოვოთ:

- ა. $P(X \leq 175)$;
 ბ. $P(175 < X \leq 200)$;
 გ. $P(X > 200)$.

ამოხსნა: X – ბინომურად განაწილებული შემთხვევითი სიდიდეა. $n = 250$, $p = 0.75$. ბინომური განაწილების ცხრილში არ არის $n = 250$ (ამ ცხრილებში, როგორც წესი, $n \leq 20$). შევამოწმოთ, შესაძლებელია თუ არა ნორმალური აპროქსიმაციის გამოყენება? ამ შემთხვევაში:

$$n \geq 30, \quad np = 250 \cdot 0.75 = 187.5, \quad 3\sqrt{np(1-p)} = 3 \cdot \sqrt{187.5 \cdot 0.25} = 20.54, \\ np \pm 3\sqrt{np(1-p)} = 187.5 \pm 20.54 = (166.96; 208.04)$$

და ცხადია, რომ $0 \leq np \pm 3\sqrt{np(1-p)} \leq n$.

უფრო მარტივად გამოსათვლელი პირობაა: $np = 250 \cdot 0.75 = 187.5 \geq 5$ და $n(1-p) = 250 \cdot 0.25 = 62.5 \geq 5$, ამიტომ Y ნორმალურად განაწილებული სიდიდეა პარამეტრებით $\mu = 187.5$ და $\sigma = 46.875$. შესაბამისად, გვაქვს:

- ა. $P(X \leq 175) = P(X = 0, 1, \dots, 175) \approx P(Y \leq 175.5) = P\left(\frac{Y - 187.5}{\sqrt{46.875}} \leq \frac{175.5 - 187.5}{\sqrt{46.875}}\right) =$
 $= P(Z \leq -1.75) = P(Z \geq 1.75) = 0.5 - \Phi_0(1.75) = 0.5 - 0.4599 = 0.0401$.
- ბ. $P(175 < X \leq 200) = P(X = 176, 177, \dots, 200) \approx P(175.5 < Y \leq 200.5) =$
 $= P\left(\frac{175.5 - 187.5}{\sqrt{46.875}} < \frac{Y - 187.5}{\sqrt{46.875}} \leq \frac{200.5 - 187.5}{\sqrt{46.875}}\right) = P(-1.75 < Z \leq 1.9) =$
 $= P(-1.75 < Z \leq 0) + P(0 < Z \leq 1.9) = \Phi_0(1.75) + \Phi_0(1.9) = 0.4599 + 0.4713 = 0.9312$.
- გ. $P(X > 200) = P(X = 201, 202, \dots, 250) \approx P(Y > 200.5) = P\left(\frac{Y - 187.5}{\sqrt{46.875}} > \frac{200.5 - 187.5}{\sqrt{46.875}}\right) =$
 $= P(Z > 1.9) = 0.5 - P(0 \leq Z \leq 1.9) = 0.5 - \Phi_0(1.9) = 0.5 - 0.4713 = 0.0287$.

ცენტრალური ზღვართი თეორემა და სიხშირეები

შერჩევაში რაიმე მონაცემის ფარდობითი სიხშირე გამოითვლება ფორმულით: $\hat{P} = X/n$, სადაც X – გვიჩვენებს შერჩევაში ამ მონაცემის რაოდენობას, რომელიც შერჩევიდან შერჩევამდე იცვლება. მათემატიკურ მოდელში X წარმოადგენს ბინომურ შემთხვევით სიდიდეს, რომელიც წარმოადგენს ბერნულის შემთხვევითი სიდიდეების ჯამს. ამიტომ ფარდობითი სიხშირე $\hat{P}$ შეგვიძლია განვიხილოთ როგორც საშუალო და ცხტ თანახმად, თუ n დიდია, $\hat{P}$ მიახლოებით ნორმალურად იქნება განაწილებული.

თუ $n \geq 30$ და $0 \leq p \pm 3\sqrt{p(1-p)/n} \leq 1$ შერჩევის სიხშირე $\hat{P}$ – მიახლოებით ნორმალურად განაწილებული შემთხვევითი სიდიდეა საშუალოთი p და დისპერსიით $p(1-p)/n$. p – პოპულაციაში ფარდობითი სიხშირე ანუ პროპორციაა (აღბათობა).

მაგალითი 8.9. 50 მოსწავლიდან 26% ცისფერთვალეა. თუ პოპულაციის 20% ცისფერთვალეა, რას უდრის ალბათობა, რომ 50 მოსწავლეში 26% ან მეტი იქნება ცისფერთვალეა.

ამოხსნა: ვინაიდან $n = 50 > 30$ და

$$p \pm 3\sqrt{p(1-p)/n} = 0.2 - 3\sqrt{0.2 \times 0.8 / 50} = 0.2 - 0.1697 = 0.03,$$

ამიტომ $\hat{P}$ მიახლოებით ნორმალურად იქნება განაწილებული საშუალოთი $p = 0.2$ და დისპერსიით $p(1-p)/n = 0.2 \cdot 0.8 / 50 = 0.0032$. შესაბამისად:

$$\begin{aligned} P(\hat{P} \geq \frac{26}{100}) &\approx P(N(0.2, 0.0032) \geq 0.26) = P\left(\frac{N(0.2, 0.0032) - 0.2}{\sqrt{0.0032}} \geq \frac{0.26 - 0.2}{\sqrt{0.0032}}\right) = \\ &= P(Z \geq 1.05) = 0.5 - \Phi_0(1.05) = 0.5 - 0.3531 = 0.1469. \end{aligned}$$

p – სიდიდეები

p – სიდიდე არის ალბათობა, რომელიც გვიჩვენებს რამდენად მოსალოდნელია კონკრეტული პოპულაციიდან ამა თუ იმ შერჩევის მიღება. თუ ეს ალბათობა (p – სიდიდე) დიდია, ჩვენ გვჯერა, რომ შერჩევა ამ პოპულაციიდანაა, რადგან ამ შერჩევის მიღების შანსი დიდია. თუ p – სიდიდე პატარაა, ჩვენი შერჩევის მიღება ამ პოპულაციიდან არ არის მოსალოდნელი.

მაგალითი 8.10. 1 ტესტის ქულები ნორმალურადაა განაწილებული საშუალოთი $\mu = 100$ და სტანდარტული გადახრით $\sigma = 15$. ვთქვათ, ეს ტესტი დაანერინეს 100 სტუდენტს და ქულების საშუალო მიიღეს 105-ის ტოლი (ან 101-ის ტოლი). რამდენად მოსალოდნელია ამ შედეგების მიღება?

ამოხსნა: ვინაიდან $n = 100 > 30$, ამიტომ უფლება გვაქვს ვისარგებლოთ ცენტრით, ანუ ჩავთვალოთ, რომ საშუალო ქულა $\bar{X} \cong N(100, 15^2 / 100)$. შესაბამისად, გვაქვს:

$$\begin{aligned} p = P(\bar{X} \geq 105) &= P\left(\frac{\bar{X} - 100}{15 / 10} \geq \frac{105 - 100}{15 / 10}\right) = \\ &= P(Z \geq 3.33) = 0.5 - \Phi_0(3.33) = 0.5 - 0.4999 = 0.0001. \end{aligned}$$

რადგანაც p – სიდიდე არის ძალიან პატარა, ამიტომ ასეთი საშუალოს მიღების შანსი პოპულაციიდან, რომლის საშუალო 100-ის ტოლია, ძალიან პატარაა. შე-

საბამისად, უნდა ჩავთვალოთ, რომ ეს შერჩევა მიღებულია სხვა პოპულაციიდან, რომლის საშუალო მეტია 100-ზე. თუკი საშუალო ქულა 101-ის ტოლია, მაშინ:

$$\begin{aligned} p &= P(\bar{X} \geq 105) = P\left(\frac{\bar{X} - 100}{15/10} \geq \frac{101 - 100}{15/10}\right) = \\ &= P(Z \geq 0.67) = 0.5 - \Phi_0(0.67) = 0.5 - 0.2486 = 0.2514. \end{aligned}$$

ამ შემთხვევაში p - სიდიდე უკვე არის არის პატარა. შესაბამისად, ასეთი შერჩევა შესაძლებელია მიღებულ იქნეს აღნიშნული პოპულაციიდან.

შეჯამება

ნორმალური განაწილება წარმოადგენს ერთ-ერთ მნიშვნელოვან განაწილებას. ნორმალური განაწილება შეიძლება სრულიად აღინეროს, თუ ცნობილია განაწილების საშუალო μ და სტანდარტული გადახრა σ . რეალური პოპულაციის მონაცემები არასოდესაა ისე განაწილებული, რომ განაწილება იყოს ზუსტად ნორმალური. მაგრამ ხშირად შეიძლება დავუშვათ, რომ პოპულაციის მონაცემებს აქვთ მიახლოებითი ნორმალური განაწილება. ნორმალური განაწილების მონაცემები სტანდარტიზაციით გარდაიქმნება z ქულებში. ამ გარდაქმნის ფორმულაა:

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

ეს ფორმულა მონაცემს გარდაქმნის რიცხვად, რომელიც გვიჩვენებს, თუ რამდენი სტანდარტული გადახრითაა ის დაშორებული საშუალოსგან.

Z ქულები ნორმალურადაა განაწილებული ნულის ტოლი საშუალოთი და ერთის ტოლი სტანდარტული გადახრით. ეს ფაქტი აღინიშნება $N(0,1)$ სიმბოლოთი. თუ ჩვენ საქმე გვაქვს ნორმალური პოპულაციის ინდივიდუალურ მონაცემებთან და გვჭირდება ალბათობების გამოთვლა, უნდა ვისარგებლოთ z - ქულით

$$Z = \frac{X - \mu}{\sigma} \equiv N(0,1).$$

თუ ალბათობის გამოთვლა საჭიროა შერჩევის საშუალოსათვის და $n \geq 30$, მაშინ

$$Z = \frac{\bar{X} - \mu}{\sigma} \equiv N(0,1).$$

თუ p - შერჩევის ფარდობითი სიხშირეა და, ამასთანავე, თუ $n \geq 30$ და $0 \leq p \leq 1$, მაშინ

$$Z = \frac{P - p}{\sqrt{p(1-p)/n}} \equiv N(0,1).$$

კითხვები

1. იპოვეთ შემდეგი ალბათობები:
 - ა. $P(0 < Z < 1.5)$; ბ. $P(-1.76 < Z \leq 0)$; გ. $P(-2 \leq Z < 1.6)$; დ. $P(1 \leq Z \leq 2.5)$;
 - ე. $P(Z > 1.96)$; ვ. $P(Z \leq 1.65)$; ზ. $P(Z > -1)$.
2. იპოვეთ z თუ
 - ა. $P(Z \geq z) = 0.025$; ბ. $P(Z > z) = 0.6$; გ. $P(Z < z) = 0.25$;
 - დ. $P(Z < z) = 0.9$; ე. $P(-z < Z < z) = 0.95$.
3. $X \equiv N(10, 0.25)$. იპოვეთ:
 - ა. $P(X > 10.9)$; ბ. $P(9 \leq X \leq 11)$; გ. $P(X \leq 8.7)$; დ. $P(\bar{X} > 10.25)$, სადაც $n = 10$;
 - ე. იპოვეთ ისეთი x , რომ $P(\bar{X} < x) = 0.85$, სადაც $n = 25$.
4. თუ ჩათვლით, რომ 20 წლის ახალგაზრდების საპატიმრო სასჯელების ვადები ნორმალურად განაწილებული სიდიდეა - $\mu = 4.2$, $\sigma = 2.8$.
 - ა. იპოვეთ ალბათობა იმისა, რომ 20 წლის შემთხვევით შერჩეული პატიმრის სასჯელის ვადა ერთ წელიწადზე ნაკლებია;
 - ბ. 20 წლის პატიმრებიდან რამდენი პროცენტის სასჯელის ვადა აღემატება 6 წელს?
 - გ. დაადგინეთ საშუალოს მიმართ სასჯელის ვადების ისეთი სიმეტრიული ინტერვალი, რომელშიც ხვდება სასჯელის ვადების 80%;
 - დ. შეაფასეთ, რამდენად კარგად აღწერს სასჯელის ვადების განაწილებას ნორმალური განაწილება.
5. 18 წლის ვაჟების სიმაღლეები ნორმალურადაა განაწილებული - $\mu = 170$, $\sigma = 15$. შემთხვევით შერჩეის 18 წლის 100 ახალგაზრდა. გამოთვალეთ ალბათობა იმისა, რომ შერჩევის საშუალო სიმაღლე აღემატება 175 სმ, $P(\bar{X} > 175) = ?$.
6. 50 წლის მალაროელების 30%-ს აქვს სუნთქვითი პრობლემები. შეაფასეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული 80 მალაროელიდან 35-ზე მეტს აქვს სუნთქვითი პრობლემები.
7. გამოკვლევამ აჩვენა, რომ დედების პირველი მშობიარობის ასაკი ნორმალურად არის განაწილებული საშუალოთი - 26.4 წელი და სტანდარტული გადახრით - 3.2 წელი.
 - ა. დედების რამდენი პროცენტის ასაკია 20 წელზე ნაკლები?
 - ბ. დედების 75%-ის ასაკი რომელ წლოვანებაზე ნაკლებია?
8. 1999 წლის ანგარიშის მიხედვით, აშშ-ის მოქალაქეები კვირაში საშუალოდ 45 საათს მუშაობენ 15 საათი სტანდარტული გადახრით. ჩათვალეთ, რომ ეს სიდიდე ნორმალურადაა განაწილებული. ამერიკელების რამდენი პროცენტი მუშაობდა კვირაში 40 საათზე მეტს?
9. ჩვილების ფსიქიკური განვითარების ინდექსი (ჩფგი) გამოიყენება მაღალი რისკის მქონე ჩვილების გამოკვლევისას. ჩათვალეთ, რომ ინდექსი ნორმალურადაა განაწილებული 100-ის ტოლი საშუალოთი და 16-ის ტოლი სტანდარტული გადახრით.
 - ა. იპოვეთ ჩვილების პროპორცია, რომლის ჩფგი ნაკლებია 80-ზე;
 - ბ. იპოვეთ 98%-იანი პროცენტილი;

- გ. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული ჩვილის ჩფგი 103-ზე მეტია;
 - დ. იპოვეთ ალბათობა იმისა, რომ შემთხვევით შერჩეული ჩვილის ჩფგი 97-ზე მეტია და 103-ზე ნაკლები;
 - ე. იპოვეთ 90-ის ტოლი ინდექსის შესაბამისი z -ქულა. ასეთი ქულის მიღების შანსი დიდია თუ პატარა?
 - ვ. ეთქვათ ყველა ინდექსი გადაიყვანეს z -ქულებში. რას უდრის მიღებული ქულების საშუალო და სტანდარტული გადახრა?
10. ჩვილების ფსიქიკური განვითარების ინდექსი (ჩფგი) გამოიყენება მაღალი რისკის მქონე ჩვილების გამოკვლევისას. ჩათვალეთ, რომ ინდექსი ნორმალურადაა განაწილებული 100-ის ტოლი საშუალოთი და 16-ის ტოლი სტანდარტული გადახრით. გამოიკვლიეს 225 ჩვილი.
- ა. იპოვეთ ალბათობა იმისა, რომ შერჩევის საშუალო 80-ზე ნაკლებია;
 - ბ. იპოვეთ საშუალოების 98% პროცენტილი;
 - გ. იპოვეთ ალბათობა იმისა, რომ საშუალო 103-ზე მეტია;
 - დ. იპოვეთ ალბათობა, რომ საშუალო 97-სა და 103-ის შორისაა;
 - ე. ეთქვათ, 225 ჩვილის საშუალო ინდექსი 90-ის ტოლია. იპოვეთ ამ სიდიდის z -ქულა. ასეთი ქულის მიღების შანსი დიდია თუ არა?
 - ვ. მიღებული პასუხები შეადარეთ წინა ამოცანის პასუხებს.
11. იპოვეთ სტანდარტული ნორმალური განაწილების პირველი, მეორე და მესამე კვარტილები.
12. იპოვეთ სტანდარტული ნორმალური განაწილების პირველი და მე-9 დეცილები.
13. 1992 წლის მონაცემებით, აშშ-ში 74 დიდი ქალაქის მკვლელობის ინდექსის საშუალოა 20.8, სტანდარტული გადახრა – 14.6.
- ა. თქვენი აზრით, არის თუ არა ნორმალური ამ ინდექსების განაწილება?
 - ბ. იპოვეთ ვაშინგტონის ინდექსის z -ქულა, თუ მისი ინდექსი უდრის 75.2-ს. თუ ჩავთვლით, რომ ინდექსებს აქვთ ნორმალური განაწილება, მოსალოდნელია თუ არა ასეთი სიდიდის მიღება?
 - გ. რას უდრის ქალაქის მკვლელობათა ინდექსი, თუ მისი z -ქულა ნულის ტოლია?
14. გადასახადები უძრავ ქონებაზე ფლორიდაში მიახლოებით ნორმალურადაა განაწილებული საშუალოთი 1400 და სტანდარტული გადახრით 600. ერთი კონკრეტული ობიექტის გადასახადი შეადგენს 1700.
- ა. იპოვეთ ამ სიდიდის z -ქულა;
 - ბ. გადასახადების რამდენი პროცენტია მეტი 1700-ზე?
 - გ. გადასახადების განაწილება უნიმოდალურია, მაგრამ ასიმეტრიული. როგორ ფიქრობთ, საით არის გადახრილი ეს განაწილება?
15. ფლორიდაში ოჯახი საშუალოდ ხარჯავს 800 კილოვატ/საათ ელექტროენერგიას 500 კილოვატ/საათი სტანდარტული გადახრით. მონაცემები გადაყვანილი იყო z -ქულებში. ოჯახებს ერთზე მეტი z -ქულით (ელექტროენერგიის მოხმარება საშუალოზე ერთი სტანდარტული გადახრით მეტი) მოუვიდათ შეტყობინება.
- ა. თუ ჩავთვლით, რომ განაწილება ნორმალურია, ოჯახების რამდენ პროცენტს მიუვიდა შეტყობინება?
 - ბ. როგორ ფიქრობთ, საით არის გადახრილი ეს განაწილება?

16. ააგეთ გერბის მოსვლის ფარდობითი სიხშირის განაწილება, როცა:
 - ა. აგდებთ ერთ მონეტას;
 - ბ. ორ მონეტას;
 - გ. სამ მონეტას;
 - დ. ოთხ მონეტას;
 - ე. აღწერეთ როგორ იცვლება განაწილების ფორმა მონეტების რაოდენობის ზრდასთან ერთად.
17. კალიფორნიაში ემიგრანტი მუშების საშუალო თვიური ანაზღაურება შეადგენს 1000 დოლარს 75-ის ტოლი სტანდარტული გადახრით.
 - ა. ჩათვალეთ, რომ ეს განაწილება ნორმალურია და იპოვეთ ალბათობა იმისა, როცა მუშა იღებს 1200 დოლარზე მეტს;
 - ბ. იპოვეთ ალბათობა, იმისა რომ 9 მუშის საშუალო ანაზღაურება 1200 დოლარზე მეტია.
18. აგვისტოში საშუალოდ ტყის 430 000 აკრი იწვის 75 000 აკრი სტანდარტული გადახრით. სხვადასხვა წელს დამწვარი ფართობი ნორმალურადაა განაწილებული.
 - ა. რას უდრის ალბათობა იმისა, რომ 500 000 აკრზე მეტი ფართობი დაიწვება?
 - ბ. რას უდრის ალბათობა იმისა, რომ 400 000 აკრზე ნაკლები ფართობი დაიწვება?
 - გ. თუ იწვის 550 000 აკრზე მეტი ფართობი, საჭიროა დამატებითი ტექნიკის შემოყვანა. რას უდრის ალბათობა იმისა, რომ საჭიროა დამატებითი ტექნიკის შემოყვანა?
19. ალბათობა იმისა, რომ ქარხნის პროდუქციის ნაწარმი უვარგისია, უდრის 0.076. შესამოწმებლად აირჩიეს 150 ნაწარმი.

იპოვეთ ალბათობა იმისა, რომ მათ შორის 21-ზე ნაკლები იქნება უვარგისი.
20. ფოტოლაბორატორიის გაძალებული სურათების 98% კარგია. შეარჩიეს 500 სურათი. იგულისხმება, რომ განაწილება ბინომურია.
 - ა. იპოვეთ უვარგისი სურათების მოსალოდნელი რაოდენობა;
 - ბ. იპოვეთ განაწილების სტანდარტული გადახრა;
 - გ. იპოვეთ ალბათობა იმისა, რომ დეფექტური სურათების რაოდენობა შეიძლება იქნება 13-ზე ნაკლები.
21. ტურისტული სააგენტო აწყობს 150 კაცისთვის გათვალისწინებულ სამდღიან ტურებს სამი ავტობუსით. პრაქტიკამ აჩვენა, რომ შეკვეთების 20 % უქმდება. თუ ტურში მონაწილეობას იღებს 115 მგზავრზე ნაკლები, სააგენტო ზარალდება. იპოვეთ ალბათობა იმისა, რომ ტური იქნება მომგებიანი.
22. მოიყვანეთ ნორმალურად განაწილებული სიდიდის მაგალითი.
23. მსხვილფეხა რქოსან პირუტყვს ასუქებენ დაკელის წინ. ხბო დღეში საშუალოდ 1.5 ფუნტს იმატებს, 0.25 ფუნტი სტანდარტული გადახრით. ვიგულისხმობთ, რომ წონაში მატება ნორმალურადაა განაწილებული. რას უდრის ალბათობა იმისა, რომ
 - ა. წონაში მატება მეტი ან ტოლი იქნება 2 ფუნტზე;
 - ბ. წონაში მატება იქნება ერთიდან ორ ფუნტამდე?
 - გ. ორი დამოუკიდებლად შერჩეული ხბო დღეში მოიმატებს 1.5 ფუნტზე ნაკლებს?
 - დ. გამოთვალეთ ეს ალბათობები, თუ სტანდარტული გადახრა შეადგენს 0.2 ფუნტს.

24. კალიფორნიაში მოხმარებული წყლის რაოდენობა ნორმალურად არის განაწილებული 18 გალონი საშუალოთი და 6 გალონი სტანდარტული გადახრით. წყლის მუდმივი დეფიციტის გამო, მთავრობა აპირებს ფასდაკლების შემოღებას მომხმარებლების 20%-თვის, ვინც ნაკლებ წყალს ხარჯავს. იპოვეთ დღეში მაქსიმუმ რამდენ გალონს უნდა ხარჯავდეს მომხმარებელი, რომ მიიღოს ფასდაკლება.
25. შემთხვევითი სიდიდე ნორმალურადაა განაწილებული $\mu = 22$ -ის ტოლი საშუალოთი. მონაცემთა 75% მეტია 17-ზე. რას უდრის ამ განაწილების სტანდარტული გადახრა?
26. შემთხვევითი სიდიდე ნორმალურადაა განაწილებული $\sigma = 6.8$ -ის ტოლი სტანდარტული გადახრით. თუ განაწილების 18% 17.5-ზე ნაკლებია, რას უდრის ამ განაწილების საშუალო?
27. უხარისხო დეტალის აღმოჩენის ალბათობაა – 0.076. იგულისხმება, რომ საქმე გვაქვს ბინომურ განაწილებასთან და შესამოწმებელია 150 დეტალი. იპოვეთ ალბათობა იმისა, რომ მათ შორის 21-ზე ნაკლები იქნება უხარისხო.
28. დეტალების 98% – ხარისხიანია. 500 დეტალიდან რამდენი შეიძლება აღმოჩნდეს უხარისხო. ჩათვალეთ, რომ განაწილება ბინომურია.
29. აეროპორტის ცენტრალურ ოფისში კვირის მანძილზე შემოსული ზარების რაოდენობა ნორმალურადაა განაწილებული საშუალოთი – 12 000 ზარი და სტანდარტული გადახრით – 2500 ზარი. იპოვეთ ალბათობა იმისა, რომ კვირის მანძილზე ზარების რაოდენობა იქნება
- ა. 11 000-ზე მეტი;
 - ბ. 10 800-დან 13 400-მდე;
 - გ. 12 800-ზე მეტი ან 11 100-ზე ნაკლები.
30. მენეჯერმა გამოიკვლია, რომ პერსონალის მიერ წელიწადში გაცემილი საათების რაოდენობა ნორმალურადაა განაწილებული საშუალოთი – 58 საათი და სტანდარტული გადახრით – 14 საათი.
- ა. მენეჯერმა ჩათვალა, რომ ორმა თანამშრომელმა სხვებთან შედარებით მეტი საათი გააცდინა, ერთმა – 74, მეორემ კი – 90 საათი. როგორ შეაფასებთ მენეჯერის პრეტენზიას?
 - ბ. კომპანია ანაზღაურებს მხოლოდ 40-საათიან გაცდენებს. შეაფასეთ კომპანიის პოლიტიკა.
31. ფირმა მარკეტში ნათურებს აგზავნის ყუთებით. თითოეულ ყუთში 1000 ნათურაა. გზაში ნათურების 10% ზიანდება.
- ა. იპოვეთ ალბათობა იმისა, რომ დაზიანდება 110-დან 130-მდე ნათურა;
 - ბ. დაზიანდება 130 ნათურაზე მეტი;
 - გ. ვთქვათ, თქვენ აღმოაჩინეთ ყუთი 200 დაზიანებული ნათურით. რამდენად მოსალოდნელი იყო ეს?

სტატისტიკური დასკვნების თეორია

ამ თავში განვიხილავთ ორ განსხვავებულ, მაგრამ ერთმანეთთან დაკავშირებულ თემას:

- შერჩევაზე დაყრდნობით, განაწილების პარამეტრების შეფასება;
- ჰიპოთეზათა შემოწმება პოპულაციის პარამეტრების შესახებ.

მაგალითი 9.1. ვთქვათ, გვინდა შევაფასოთ ქალაქში ბინის ქირის საშუალო ფასი. შევადგოვით მონაცემები რამდენიმე ბინის ქირის ფასების შესახებ და საშუალოდ მივიღეთ $x = \$ 328$. მიღებული რიცხვი არის პოპულაციის საშუალოს (ქალაქის ყველა ბინის) ნერტილოვანი შეფასება.

ისმის კითხვა: რამდენად კარგია ეს შეფასება, რამდენად ახლოს არის ის ნამდვილ საშუალოსთან? თუ არ გაგვაჩნია სხვა რაიმე ინფორმაცია პოპულაციის საშუალოზე, უმჯობესია განვიხილოთ ინტერვალური შეფასება, რომელსაც ხშირად მოიხსენიებენ როგორც ნდობის ინტერვალს.

თუ გაგვაჩნია რაიმე წინასწარი ინფორმაცია პოპულაციის შესახებ და ვარაუდი პოპულაციის პარამეტრის შესახებ, ჩვენ შეგვიძლია დავსვათ კითხვა: რამდენად ახლოს არის გამოთვლილი ნერტილოვანი შეფასება ამ პოპულაციის პარამეტრთან, ანუ შევამოწმოთ შესაბამისი ჰიპოთეზა.

ნდობის ინტერვალის პოპულაციის საშუალოსთვის

თუ შერჩევის მოცულობა დიდია, ანუ $n \geq 30$, ძალაშია ცენტრალური ზღვარითი თეორემა). ეს ნიშნავს, რომ სტანდარტიზებული შერჩევის საშუალოს აქვს $N(0,1)$ სტანდარტული ნორმალური განაწილება:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \cong N(0,1).$$

ხშირ შემთხვევაში σ არ არის ცნობილი, მაგრამ, როცა n დიდია, σ -ს ვანაცვლებთ s -ით:

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \cong N(0,1).$$

ნდობის ინტერვალს აგებენ შერჩევის განაწილების გამოყენებით (ამ შემთხვევაში სტანდარტული ნორმალური განაწილებით). ვიცით, რომ Z -ის 95%

კონცენტრირებული $(-1.96; +1.96)$ ინტერვალში და გარდაქმნით ვიღებთ ინტერვალს μ -თვის:

$$\begin{aligned} 0.95 &= P(-1.96 < Z < 1.96) = P\left(-1.96 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right) = \\ &= P\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

ბოლო უტოლობას შეგვიძლია მივცეთ შემდეგი ინტერპრეტაცია: 0.95-ის ტოლია ალბათობა იმისა, რომ μ -ს უცნობი მნიშვნელობა მოთავსებული იქნება

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ ინტერვალში. მაგრამ, ამავე დროს, არსებობს რისკის ალბათობა 0.05,

რომ $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ ინტერვალი არ მოიცავდეს μ -ს.

რა ხდება, როცა ჩვენ ვიღებთ n მოცულობის მქონე შერჩევას და ვითვლით საშუალოს ნერტილოვან შეფასებას? რადგან $\bar{X}$ შემთხვევითი სიდიდეა, სხვადასხვა შერჩევისათვის მივიღებთ განსხვავებულ მნიშვნელობას, მაგრამ არსებობს 0.95 ალბათობა ანუ საიმედოობა, რომ $\bar{X}$ არ იქნება μ -გან განსხვავებული

$\pm 1.96 \frac{\sigma}{\sqrt{n}}$ -ზე მეტი სიდიდით.

ნდობის ინტერვალისათვის ტერმინი *ნდობა*, *საიმედოობა* აღნიშნავს ალბათობას, რომ μ -ს უცნობი მნიშვნელობა იქნება ნდობის ინტერვალში.

ნდობის ინტერვალისთვის ტერმინი – *რისკი* აღნიშნავს ალბათობას, რომ μ -ს უცნობი მნიშვნელობა ნდობის ინტერვალში არ იქნება.

მაგალითი 9.2. ვთქვათ, გვსურს შევაფასოთ რა დრო სჭირდება სტუდენტს დავალების მოსამზადებლად. შემთხვევით შერჩეულ 49 სტუდენტს საშუალოდ სჭირდება 3.3 საათი, სტანდარტული გადახრით – 1.1 საათი. ააგეთ 95%-იანი ნდობის ინტერვალი.

ამოხსნა: $n = 49 > 30$, ამიტომ პოპულაციის σ სტანდარტული გადახრა შეგვიძლია ჩავანაცვლოთ შერჩევის სტანდარტული გადახრით. Z -ის მნიშვნელობების 95% არის ± 1.96 რიცხვებს შორის, $x = 3.3$, $s = 1.1$, $n > 30$, ამიტომ უფლება გვაქვს გამოვიყენოთ ცხტ და ჩავწეროთ μ -ს ნდობის ინტერვალი –

$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} = 3.3 \pm 1.96 \frac{1.1}{\sqrt{49}}$. μ -ს 95%-იანი ნდობის ინტერვალია $(2.992; 3.608)$. რადგან დარწმუნებული ვართ, რომ μ არის ამ ინტერვალში, აქედან გამომდინარე, ასეთი ინტერვალის 95% მოიცავს μ -ს.

ნდობის ინტერვალი გვაძლევს μ -ს საყარაუდო მოსალოდნელ დიაპაზონს, ასევე მიგვითითებს, რომელი მნიშვნელობა არ არის მოსალოდნელი. ასე, მაგალითად, არ არის მოსალოდნელი, რომ საშუალო დრო იყოს 2.992-ზე ნაკლები და 3.608-ზე მეტი. თუ საჭიროა ვუპასუხოთ კითხვას, უდრის თუ არა საშუალო 3.8-ს, სტატისტიკაში სწორედ ამ ფაქტს იყენებენ. ამ შემთხვევაში პასუხი იქნება უარყოფითი, რადგან ნდობის ინტერვალი არ მოიცავს 3.8-ს. ალტერნატიულ პროცედურას იყენებენ ჰიპოთეზათა შემოწმების ტესტებში.

ჰიპოთეზათა შემოწმება საშუალოსთვის

ძირითადი იდეა ისაა, რომ ზომავენ სხვაობას შერჩევის გამოთვლილ და პოპულაციის საყარაუდო სიდიდეებს შორის და არკვევენ რამდენად დიდია ეს სხვაობა (მაგალითად, შერჩევის საშუალოსა და პოპულაციის საშუალოს შორის).

ჰიპოთეზათა შემოწმების ეტაპებია:

1. ჰიპოთეზების ფორმულირება;
2. კრიტერიუმის სტატისტიკის მნიშვნელობის გამოთვლა;
3. უარყოფის არის (კრიტიკული არის) დადგენა;
4. გადაწყვეტილების მიღება.

განვიხილოთ ცალ-ცალკე ოთხივე ეტაპი.

1. ჰიპოთეზების ფორმულირება

ჰიპოთეზათა შემოწმების ტესტში ყოველთვის ფიგურირებს ორი ჰიპოთეზა: ნულოვანი ჰიპოთეზა H_0 და ალტერნატიული ჰიპოთეზა H_A . ჰიპოთეზათა ფორმულირება ყოველთვის ხდება მონაცემთა შეგროვების წინ. ნულოვანი ჰიპოთეზა ხშირად ასეთი სახისაა $\mu = \mu_0$. ალტერნატიული ჰიპოთეზა შეიძლება იყოს სამი სახის:

- ორმხრივი ალტერნატიული ჰიპოთეზა $H_A: \mu \neq \mu_0$;
- მარჯვენა ცალმხრივი $H_A: \mu > \mu_0$;
- მარცხენა ცალმხრივი $H_A: \mu < \mu_0$.

ძალიან მნიშვნელოვანია ის, რომ ჰიპოთეზათა ფორმულირება ხდება შერჩევიდან მიღებული ინფორმაციის წინ. ითვალისწინებენ სიტუაციის ცოდნას და არა იმას რა მივიღეთ შერჩევაში.

შენიშვნა 9.1. ნულოვანი ჰიპოთეზა უნდა იყოს ტოლობის მომცველი გამოსახულება ($=$, $\leq$ და $\geq$), ალტერნატივა კი – ტოლობის გამომრიცხავი გამოსახულება (შესაბამისად, $\neq$, $>$ და $<$).

მაგალითი 9.3. ჩამოაყალიბეთ ნულოვანი და ალტერნატიული ჰიპოთეზები:

1. ჯანმრთელ 70 წლის ადამიანს საშუალოდ სჭირდება 30 წუთი სპეციალური ტესტის შესასრულებად. მოსალოდნელია, რომ პაციენტებს დაქვეითებული მეხსიერებით დასჭირდებათ მეტი დრო.
2. მოსალოდნელია, რომ ახალი არტერიული წნევის პრეპარატი შეამცირებს პაციენტის წნევას.

3. სარეცხი მანქანის მუშაობის ვადა მეტია საგარანტიო 2 წლის ვადაზე.
 4. დასავლეთის ბაყაყების სიცოცხლის ხანგრძლივობა მეტია აღმოსავლეთის ბაყაყების სიცოცხლის ხანგრძლივობაზე, რომელიც შეადგენს 30 წელს.
 5. ქალაქში ქურდობის ფაქტების 82% იმოქმედებს ტურისტების რაოდენობაზე.
- ამოხსნა:** ჩვენ უნდა განვსაზღვროთ μ_0 -ს საყარაუდო მნიშვნელობა და ჩამოვაყალიბოთ ჰიპოთეზები:

1. $H_0: \mu = 30$, $H_A: \mu > 30$. შერჩევაში შევლენ დაქვეითებული მეხსიერების მქონე 70 წლის პაციენტები.
2. $H_0: \mu = \mu_0$, $H_A: \mu > \mu_0$. μ_0 არის პაციენტების არსებული საშუალო წნევა. შერჩევაში შევლენ პაციენტები, რომლებიც ღებულობენ ახალ პრეპარატს.
3. $H_0: \mu = 2$, $H_A: \mu > 2$. შერჩევაში შევა ორწლიანი გარანტიით გაცემული მანქანების მუშაობის ვადები.
4. $H_0: \mu = 30$, $H_A: \mu > 30$. საჭიროა ინფორმაცია დასავლეთის ბაყაყებზე.
5. $H_0: \mu = \mu_0$, $H_A: \mu \neq \mu_0$. რადგან არ ვიცით როგორ იმოქმედებს ეს ფაქტი ტურისტებზე, შევაგროვებთ მონაცემებს ჩამოსული ტურისტების რაოდენობაზე და შევადარებთ წინა წლებში ჩამოსული ტურისტების საშუალო რაოდენობას. ჰიპოთეზათა შემოწმების მეორე ეტაპია კრიტერიუმის სტატისტიკის მნიშვნელობის გამოთვლა.

2. კრიტერიუმის სტატისტიკის მნიშვნელობის გამოთვლა

ჰიპოთეზათა შემოწმების მეორე ეტაპი ეფუძნება შერჩევის მონაცემებს, ამიტომ საჭიროა ამ მონაცემების შეგროვება და ისეთი სტატისტიკის შერჩევა, რომელიც მოგვცემს საშუალებას უარვყოთ ან არ უარვყოთ ჰიპოთეზა. ასეთ სტატისტიკას კრიტერიუმის სტატისტიკა ეწოდება. თუ იგულისხმება, რომ შერჩევის მოცულობა დიდია, $n \geq 30$, კრიტერიუმის სტატისტიკის ასაგებად უნდა მოვახდინოთ შერჩევის საშუალოს სტანდარტიზაცია. ეს კი გვაძლევს საშუალებას გამოვიყენოთ ტაბულირებული $N(0,1)$ ალბათობები. კრიტერიუმის სტატისტიკა ასეთია:

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

აქ σ -ს მაგივრად ჩასმულია s , რადგან, როცა დიდია n , s ძალიან ახლოსაა σ -სთან.

ამ ფორმულაში შედის μ . ნდობის ინტერვალის აგების დროს ეს სიდიდე უცნობი იყო და ჩვენ ვადგენდით მის საყარაუდო დიაპაზონს. ამ ამოცანაში მოცემულია H_0 -ით შემოთავაზებული ჰიპოთეტური სიდიდე.

მაგალითი 9.4. გვაქვს $n = 49$ დაკვირვება, $\bar{x} = 3.3$ -ის ტოლი საშუალოთი და $s = 1.1$ სტანდარტული გადახრით. ნულოვანი ჰიპოთეზა — $H_0: \mu = 3.8$. თუ ნულოვანი ჰიპოთეზა ჭეშმარიტია, შერჩევის საშუალო ახლოს უნდა იყოს 3.8-თან. ეს სიახ-

ლოვე იზომება სდანდარტული გადახრების ერთეულებში. კრიტერიუმის სტატისტიკის მნიშვნელობა ასეთია:

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{3.3 - 3.8}{\frac{1.1}{\sqrt{49}}} = -3.18.$$

რადგან $n = 49 > 30$ -ზე, Z -ს აქვს ნორმალური განაწილება $N(0,1)$.

3. უარყოფის არის დადგენა

უარყოფის არის დასადგენად, უნდა გავარკვიოთ Z -ის რომელი მნიშვნელობები არ არის მოსალოდნელი, თუ ნულოვანი ჰიპოთეზა სამართლიანია.

უარყოფის არე დამოკიდებულია:

1. კრიტერიუმის სტატისტიკის განაწილებაზე;
2. განაწილების რა კუდები იქნება გამოყენებული;
3. ამ კუდებში მოხვედრის ალბათობებზე.

9.4 მაგალითში Z -ს აქვს $N(0,1)$ განაწილება. განაწილების რომელ კუდებს გამოვიყენებთ, ეს დამოკიდებულია ალტერნატიულ ჰიპოთეზაზე. არსებობს სამი შესაძლებლობა: $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$.

ვთქვათ, ალტერნატივაა $\mu > \mu_0$ და თუ ის სამართლიანია, მაშინ მოსალოდნელია, რომ სხვაობა $\bar{x} - \mu$ დიდი და ნულოვანი ჰიპოთეზის უარყოფის არე იქნება განაწილების მარჯვენა კუდი, ანუ ინტერვალი $Z > z'$.

თუ ალტერნატივაა $\mu < \mu_0$ -ზე, ნულოვანი ჰიპოთეზის უარყოფის არედ ავიღებთ განაწილების მარცხენა კუდს, ანუ ინტერვალს $Z < -z'$.

თუ $\mu \neq \mu_0$, მაშინ $\bar{x}$ შეიძლება იყოს μ_0 -ზე მეტი, ან ნაკლები და უარყოფის არედ ავიღებთ განაწილების ორივე კუდს, ანუ ორ ინტერვალს $Z < -z'$, $Z > z'$, სადაც z' -ს ვირჩევთ ისე, რომ, თუ სამართლიანია H_0 , მაშინ ძალიან მცირე უნდა იყოს უარყოფის არეში მოხვედრის ალბათობა. მაგალითად, 0.01 ან 0.05 ან 0.1-ის ტოლი. ამ ალბათობას კრიტერიუმის მნიშვნელოვნების დონე ეწოდება და მას აღნიშნავენ α ასოთი.

z' სიდიდეებს კრიტიკული მნიშვნელობები ეწოდება. მარჯვენა ცალმხრივი კრიტერიუმის შემთხვევაში $-z'$ სტანდარტული ნორმალური განაწილების ზედა α -კრიტიკული წერტილია $z' = z_\alpha$, მარცხენა ცალმხრივი კრიტერიუმის შემთხვევაში $-z'$ სტანდარტული ნორმალური განაწილების ზედა α -კრიტიკული წერტილის მოპირდაპირე რიცხვია $z' = -z_\alpha$, ხოლო ორმხრივი კრიტერიუმის შემთხვევაში z' სტანდარტული ნორმალური განაწილების ზედა $\alpha/2$ -კრიტიკული წერტილია $z' = z_{\alpha/2}$.

შენიშვნა 9.2. თუ $\alpha \leq 0.5$, მაშინ $\Phi(z_\alpha) = 0.5 - \alpha$ (თუ $\alpha \geq 0.5$, მაშინ $\Phi(z_\alpha) = \alpha - 0.5$).

მაგალითი 9.5. ავავთ კრიტერიუმების უარყოფის არეები შემდეგი ჰიპოთეზებისათვის:

1. $H_0: \mu = 3.8$ და $H_A: \mu > 3.8$, $\alpha = 0.05$.

H_0 იქნება უარყოფილი, თუ $Z > 1.645$.

2. $H_0: \mu = 3.8$ და $H_A: \mu < 3.8$, $\alpha = 0.05$.

H_0 იქნება უარყოფილი, თუ $Z < -1.645$.

3. $H_0: \mu = 3.8$ და $H_A: \mu \neq 3.8$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$.

H_0 იქნება უარყოფილი, თუ $Z > 1.96$ ან $Z < -1.96$.

კრიტიკული მნიშვნელობები ვიპოვეთ $\Phi(z)$ ფუნქციის ცხრილის გამოყენებით. თუ $\alpha = 0.05$, მაშინ მარჯვენა ცალმხრივი კრიტერიუმისათვის - $z' = z_{0.05} = 1.645$, მარცხენა ცალმხრივი კრიტერიუმისათვის - $-z' = -z_{0.05} = -1.645$ და ორმხრივი კრიტერიუმისათვის - $\mp z' = \mp z_{\alpha/2} = \pm 1.96$ ($\alpha/2 = 0.025$).

4. გადაწყვეტილების მიღება

ჰიპოთეზათა შემოწმების ბოლო ეტაპია გადაწყვეტილების მიღება.

- თუ კრიტერიუმის სტატისტიკის მნიშვნელობა მოხვდება უარყოფის არეში, ჩვენ უარყოფთ ნულოვან ჰიპოთეზას α მნიშვნელოვნების დონეზე.
- თუ კრიტერიუმის სტატისტიკის მნიშვნელობა არ მოხვდება უარყოფის არეში, ჩვენ ნულოვან ჰიპოთეზას ვერ უარყოფთ α მნიშვნელოვნების დონეზე.

მაგალითი 9.6. $n = 49$ სტუდენტის მიერ დავალებაზე დახარჯული საშუალო დროა $\bar{x} = 3.3$ საათი, სტანდარტული გადახრით - $s = 1.1$ საათი, ხოლო $\mu = 3.8$ საათი. მაშინ:

$$Z = \frac{3.3 - 3.8}{\frac{1.1}{\sqrt{49}}} = -3.18.$$

ალტერნატივაა - $H_A: \mu < 3.8$. გვაქვს მარცხენა ცალმხრივი კრიტერიუმი, $\alpha = 0.05$, შესაბამისი კრიტიკული მნიშვნელობა $z' = -1.645$ -ს. შევადაროთ სტატისტიკის გამოთვლილი მნიშვნელობა - -3.18 და $\Phi(z)$ ფუნქციის ცხრილიდან მიღებული კრიტიკული მნიშვნელობა -1.645 . კრიტერიუმის სტატისტიკის მნიშვნელობა მოხვდა უარყოფის არეში და ამიტომ უარყოფთ ნულოვან ჰიპოთეზას. გადაწყვეტილება: საშუალო დრო 3.8-ზე ნაკლებია. სტუდენტებს ნაკლები დრო სჭირდებათ დავალების შესასრულებლად.

სტატისტიკური ტესტი ისეთნაირადაა ჩამოყალიბებული, რომ, როცა ნულოვანი ჰიპოთეზა სამართლიანია, უფრო მეტად მოსალოდნელია, რომ მას ვერ უარყოფთ. ცნობილია გამოთქმა - „უდანაშაულობის პრეზუმფცია“. სამართლიან სასამართლოში დამნაშავედ არ იქნება გამოცხადებული პიროვნება საფუძვლიანი ეჭვის გარეშე. ასევე სტატისტიკაში მოითხოვება დამაჯერებელი ფაქტის განხორ-

ციელება, ანუ ისეთი ხდომილების მოხდენა, რომლის მოხდენის ალბათობა არ არის პატარა. ამიტომ გადაწყვეტილება „უარეყოფთ H_0 -ს“, განიხილება როგორც ძლიერი გადაწყვეტილება და გადაწყვეტილება „ვერ უარეყოფთ H_0 -ს“ – როგორც სუსტი.

***p*-სიდიდეების გამოყენება ჰიპოთეზათა შემოწმებაში**

p-სიდიდეები შეგვიძლია განვიხილოთ როგორც დაკვირვებული მნიშვნელოვნების დონეები. ვთქვათ, შერჩევაზე დაყრდნობით გამოთვლილია სტატისტიკის მნიშვნელობა – z . ვიპოვოთ ამ მნიშვნელობაზე უფრო ექსტრემალური მნიშვნელობის მიღების ალბათობა და ვუნოდოთ მას *p*-სიდიდე. თუ ეს ალბათობა დიდაა, მაშინ მოსალოდნელი იყო კიდეც ასეთი შერჩევის მიღება. თუ *p*-სიდიდე ძალიან პატარაა, H_0 -ის სამართლიანობის დროს არაა მოსალოდნელი ასეთი შერჩევის მიღება.

შენიშვნა 9.3. მარჯვენა ცალმხრივი კრიტერიუმის შემთხვევაში – $p = P(Z > z) = 1 - \Phi(z)$ მარცხენა ცალმხრივი კრიტერიუმის შემთხვევაში – $p = P(Z < z) = \Phi(z)$, ხოლო ორმხრივი კრიტერიუმის შემთხვევაში – $p = P(|Z| > |z|) = 2(1 - \Phi(|z|))$.

გადაწყვეტილების მიღება: თუ *p*-სიდიდე α -ზე ნაკლებია, H_0 -ს უარეყოფთ. თუ *p*-სიდიდე α -ზე მეტია, H_0 -ს ვერ უარეყოფთ.

მაგალითი 9.7. წინა მაგალითის პირობებში ვთქვათ, $H_0: \mu = 3.8$ და $H_A: \mu < 3.8$.

ამოხსნა: $\bar{x} = 3.3$, $n = 49$, $s = 1.1$, $Z = \frac{3.3 - 3.8}{1.1/\sqrt{49}} = -3.18$. სტანდარტული ნორმა-

ლური განაწილების ცხრილიდან -3.18 -ზე ნაკლები ან ტოლი მნიშვნელობის მიღების ალბათობა უდრის 0.0007 -ს. შესაბამისად, *p*-სიდიდე $= 0.0007$, რაც გაცილებით ნაკლებია $\alpha = 0.05$ -ზე, ამიტომ ნულოვანი ჰიპოთეზა – $\mu = 3.8$ უნდა უარეყოთ ალტერნატივის სასარგებლოდ.

პირველი და მეორე გვარის შეცდომები სტატისტიკური კრიტერიუმების გამოყენებისას

სტატისტიკური კრიტერიუმი მკვლევარს აძლევს საშუალებას უარყოს ან ვერ უარყოს ნულოვანი ჰიპოთეზა. გადაწყვეტილების სისწორე დამოკიდებულია იმაზე, არის თუ არა ნულოვანი ჰიპოთეზა ჭეშმარიტი. გადაწყვეტილების მიღების დროს შესაძლებელია დაშვებული იყოს შეცდომები: **პირველი გვარის შეცდომა**, როცა უარყოფენ ნულოვან ჰიპოთეზას მაშინ, როცა ის ჭეშმარიტია და **მეორე გვარის შეცდომა**, როცა არ უარყოფენ მცდარ ნულოვან ჰიპოთეზას. ცხრილი 9.1 გვიჩვენებს ამ შეცდომებს და შესაძლო სწორ გადაწყვეტილებებს.

შეცდომები ხდება იმის გამო, რომ გადაწყვეტილების მიღება ალბათური ხასიათისაა. პირველი გვარის დაშვების ალბათობა დამოკიდებულია არჩეული მნიშვნელოვნების დონეზე. როცა $\alpha = 0.05$, არსებობს ხუთი შანსი ასიდან (ანუ ერთი შანსი ოციდან), რომ დაშვებული იყოს პირველი გვარის შეცდომა ნულოვანი ჰიპოთეზის უარყოფით, როცა ის ნამდვილად ჭეშმარიტია. ანალოგიურად, როცა $\alpha = 0.01$ -ს, არსებობს ერთი შანსი 100-დან, რომ დაშვებული იყოს პირველი გვარის შეცდომა. პირველი გვარის შეცდომის შემცირებით იზრდება მეორე გვარის შეცდომის დაშვების შანსი. მეორე გვარის შეცდომას აღნიშნავენ β -თი. H_A -ს მიღების ალბათობა, როცა ის ჭეშმარიტია, არის $1 - \beta$. $1 - \beta$ -ს ეწოდება სტატისტიკური კრიტერიუმის სიმძლავრე.

ცხრილი 9.1. შეცდომები და სწორი გადაწყვეტილებები სტატისტიკურ გადაწყვეტილებებში. თითოეულ შედეგთან დაკავშირებული ალბათობა

ჭეშმარიტი სიტუაცია	გადაწყვეტილება	
H_0 არ არის უარყოფილი	H_0 უარყოფილი	
H_0 ჭეშმარიტია	სწორი გადაწყვეტილება $p = 1 - \alpha$	პირველი გვარის შეცდომა $p = \alpha$
H_A ჭეშმარიტია	მეორე გვარის შეცდომა $p = \beta$	სწორი გადაწყვეტილება $p = 1 - \beta$

ჰიპოთეზათა შემოწმება ნორმალური პოპულაციის საშუალოსათვის ($n < 30$)

თუ შერჩევის მოცულობა $n < 30$, შერჩევის საშუალოს არა აქვს ნორმალური განაწილება.

როცა შერჩევა არის მიღებული ნორმალური პოპულაციიდან, $n < 30$ და პოპულაციის დისპერსია უცნობია, იყენებენ t -განაწილებას:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \cong t(n-1),$$

სადაც $t(n-1)$ არის t -განაწილება (სტიუდენტის განაწილება) თავისუფლების ხარისხით $n-1$. თავისუფლების ხარისხს აღნიშნავენ df -ით. t -განაწილებები არის სიმეტრიული და უნიმოდალური. საშუალო, მოდა და მედიანა უდრის ნულს. დისპერსია დამოკიდებულია თავისუფლების ხარისხზე. განაწილების ფორმა დამოკიდებულია თავისუფლების ხარისხზე, df იღებს მხოლოდ მთელ მნიშვნელობებს $1, 2, \dots$. თავისუფლების ხარისხის ზრდასთან ერთად დისპერსიის მნიშვნელობები უახლოვდებიან ერთს და t -განაწილება ემსგავსება სტანდარტულ ნორმალურ გა-

ნაწილებას. როცა $n \geq 30$ -ზე, იმდენად უმნიშვნელოა განსხვავება $t(n-1)$ განაწილებასა და $N(0,1)$ განაწილებას შორის, რომ კრიტერიუმის სტატისტიკის განაწილებად იღებენ სტანდარტულ ნორმალურ განაწილებას.

$100(1-\alpha)\%$ -იანი ნდობის ინტერვალი μ -თვის, როცა $n < 30$, არის:

$$(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}),$$

სადაც $t_{n-1, \alpha/2}$ - თავისუფლების $n-1$ ხარისხის მქონე სტიუდენტის განაწილების ზედა α -კრიტიკული წერტილია, რომელსაც ვპოულობთ სტიუდენტის ზედა α -კრიტიკული წერტილების ცხრილიდან. ამ ცხრილში სტრიქონი შეესაბამება თავისუფლების ხარისხს, სვეტი კი $\alpha/2$ -ს.

მაგალითი 9.8. გამოკითხეს 12 სტუდენტი, თუ რამდენი საათი სჭირდებათ მათ დავალების შესასრულებლად და მიიღეს შემდეგი მონაცემები: 2.4, 3.4, 1.9, 5.5, 2.4, 1.2, 2.0, 1.7, 1.8, 2.8, 2.7, 2.3. ააგეთ 95%-იანი ნდობის ინტერვალი პოპულაციის საშუალოსათვის.

ამოხსნა: ამ მონაცემების საშუალოა $\bar{x} = 2.51$, სტანდარტული გადახრა $s = 1.1$. მოცულობა $n = 12$ ნაკლებია 30-ზე. უნდა ვივარაუდოთ, რომ შერჩევა მიღებულია ნორმალური პოპულაციიდან (ამისათვის შერჩევის პოლიგონი უნდა იყოს სიმეტრიული, უნიმოდალური და არ უნდა იყოს ექსტრემალური მონაცემები). მაშინ შეგვიძლია ვისარგებლოთ თავისუფლების $n-1$ ხარისხის მქონე t -განაწილებით. t -განაწილების ცხრილიდან, როცა $\alpha/2 = 0.025$ და $df = 12 - 1 = 11$, გვაქვს $t_{11, 0.025} = 2.201$. შესაბამისად, 95%-იანი ნდობის ინტერვალი იქნება:

$$(2.51 - 2.201 \cdot \frac{1.1}{\sqrt{12}}, 2.51 + 2.201 \cdot \frac{1.1}{\sqrt{12}}) = (2.51 - 0.70, 2.51 + 0.70) = (1.81, 3.21).$$

ნდობის ინტერვალი დისპერსიისათვის

α მნიშვნელოვნების დონის მქონე ნდობის ინტერვალი ნორმალური $N(a, \sigma^2)$ პოპულაციის დისპერსიისათვის ცნობილი საშუალოს შემთხვევაში არის:

$$\left(\frac{\sum_{i=1}^n (x_i - a)^2}{\chi_{n-1, \alpha/2}^2}, \frac{\sum_{i=1}^n (x_i - a)^2}{\chi_{n-1, 1-\alpha/2}^2} \right),$$

ხოლო უცნობი საშუალოს შემთხვევაში კი:

$$\left(\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2} \right),$$

სადაც $\chi^2_{m,\alpha}$ – თავისუფლების m ხარისხის მქონე ხი-კვადრატ განაწილების ($\chi^2(m)$) ზედა α -კრიტიკული წერტილია.

ზოგჯერ გამოიყენება აღნიშვნები: $\chi^2_{\alpha/2, n-1} := \chi^2_{n-1, \alpha/2}$ და $\chi^2_{1-\alpha/2, n-1} := \chi^2_{n-1, 1-\alpha/2}$.

α მნიშვნელობების დონის მქონე ნდობის ინტერვალის ნორმალური პოპულაციის სტანდარტული გადახრისათვის შერჩევის დიდი მოცულობის შემთხვევაში დაახლოებით ასეთია:

$$(s' - z_{\alpha/2} \cdot \frac{s'}{\sqrt{2n}}, s' + z_{\alpha/2} \cdot \frac{s'}{\sqrt{2n}}).$$

შერჩევითი დისპერსია ეწოდება სიდიდეს:

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i).$$

შესწორებული შერჩევითი დისპერსია ეწოდება სიდიდეს:

$$s^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

შერჩევითი სტანდარტული გადახრა (შესაბამისად, შესწორებული შერჩევითი სტანდარტული გადახრა) ეწოდება არითმეტიკულ კვადრატულ ფესვს შერჩევითი დისპერსიიდან (შესაბამისად, შესწორებული შერჩევითი დისპერსიიდან).

მაგალითი 9.9. იპოვეთ თამბაქოში ნიკოტინის შემცველობის დისპერსიისა და სტანდარტული გადახრისათვის 95%-იანი ნდობის ინტერვალის, თუ ცნობილია, რომ 20 სიგარეტისაგან შედგენილ შერჩევაში ნიკოტინის შემცველობის შესწორებული სტანდარტული გადახრა 1.6 მილიგრამის ტოლია.

ამოხსნა. ვინაიდან ამ შემთხვევაში $\alpha = 1 - 0.95 = 0.05$, ამიტომ ხი-კვადრატ განაწილების ორი კრიტიკული მნიშვნელობა, რომელიც შეესაბამება $\alpha/2 = 0.025$ -სა და $1 - \alpha/2 = 0.975$ -ს და თავისუფლების ხარისხის რიცხვს, რომელიც ტოლია $n-1 = 20-1 = 19$, შესაბამისად იქნება: $\chi^2_{19, 0.025} = 32.852$ და $\chi^2_{19, 0.975} = 8.907$. ამიტომ საძიებელი ნდობის ინტერვალის იქნება:

$$\left(\frac{(20-1) \cdot (1.6)^2}{32.852}, \frac{(20-1) \cdot (1.6)^2}{8.907} \right), \text{ ანუ } (1.5, 5.5).$$

ამრიგად, ჩვენ შეგვიძლია 95%-იანი საიმედოობით ვიგულისხმოთ, რომ სიგარეტში ნიკოტინის შემცველობის რეალური დისპერსია მოთავსებულია 1.5 მილიგრამ²-სა და 5.5 მილიგრამ²-ს შორის.

ნდობის ინტერვალის სტანდარტული გადახრისათვის იქნება:

$$(\sqrt{1.5}, \sqrt{5.5}), \text{ ანუ } (1.2, 2.3).$$

მაშასადამე, სიგარეტში ნიკოტინის შემცველობის რეალური სტანდარტული გადახრის 95%-იანი ნდობის ინტერვალია (1.2 მგ, 2.3 მგ).

ჰიპოთეზათა შემოწმება (პოპულაცია ნორმალურია, $n < 30$)

კრიტერიუმის სტატისტიკას

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

აქვს t -განაწილება თავისუფლების ხარისხით $n-1$.

მაგალითი 9.10. წინა ინფორმაციაზე დაყრდნობით, სტუდენტებს დავალების შესასრულებლად საშუალოდ სჭირდებათ 3.3 საათი. 0.05 მნიშვნელოვნების დონეზე შეამოწმეთ, მნიშვნელოვნად განსხვავდება თუ არა ეს დრო 12 სტუდენტის საშუალო დროისაგან, რომელიც 2.51-ის ტოლია.

ამოხსნა: ჩამოვაყალიბოთ ნულოვანი და ალტერნატიული ჰიპოთეზა:

$$H_0: \mu = 3.3, \quad H_A: \mu \neq 3.3.$$

გამოვთვალოთ კრიტერიუმის მნიშვნელობა:

$$t = \frac{2.51 - 3.3}{\frac{1.1}{\sqrt{12}}} = -2.49.$$

t -განაწილების ცხრილიდან, როცა $\alpha/2 = 0.025$ და $df = n-1 = 11$, გვაქვს $t_{11,0.025} = 2.201$. შესაბამისად, უარყოფის არეა: $t \leq -2.201$ ან $t \geq 2.201$. კრიტერიუმის მნიშვნელობა -2.49 ხდება H_0 -ის უარყოფის არეში. ნულოვანი ჰიპოთეზა უნდა უარვეყოთ ალტერნატივის სასარგებლოდ. დავალების შესასრულებლად სტუდენტებს, საშუალოდ, არ სჭირდებათ 3.3 საათი, არამედ სჭირდებათ უფრო ნაკლები დრო.

მაგალითი 9.11. ექიმების აზრით, მორბენალი ადამიანი უფრო მეტ ჟანგბადს მოიხმარს, ვიდრე ის, რომელიც ფიზიკურად ნაკლებადაა დატვირთული და ეს მოხმარება შეადგენს 36.7 მლ/კგ. გვაქვს თუ არა საკმარისი საფუძველი დავუჯეროთ ექიმებს $\alpha = 0.05$ მნიშვნელოვნების დონით? შემთხვევით შერჩეული 15 მორბენალისათვის ჟანგბადის მოხმარების საშუალომ შეადგინა 40.6 მლ/კგ, ხოლო სტანდარტული გადახრაა 6 მლ/კგ.

ამოხსნა: ჩამოვაყალიბოთ ნულოვანი და ალტერნატიული ჰიპოთეზა:

$$H_0: \mu = 36.7, \quad H_A: \mu > 36.7$$

გამოვთვალოთ კრიტერიუმის მნიშვნელობა:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{40.6 - 36.7}{\frac{6}{\sqrt{15}}} = 2.517.$$

$df = 15 - 1 = 14$, გვაქვს ცალმხრივი მარჯვენა ალტერნატივა. t -განაწილების ცხრილიდან, როცა $\alpha = 0.05$ და $df = 15 - 1 = 14$, გვაქვს $t_{14,0.05} = 1.761$. შესაბამისად, უარყოფის არეა: $t \geq 1.761$. რადგან $2.517 > 1.761$, ნულოვან ჰიპოთეზას უარყოფთ. გვაქვს საფუძველი, გავიზიაროთ ექიმების მტკიცებულება.

ჰიპოთეზათა შემოწმება პოპულაციის პროპორციისათვის ($n \geq 30$)

p ალბათობის შესაფასებლად, ბუნებრივია, განვიხილოთ შერჩევის ფარდობითი სიხშირე $\hat{p}$. $\hat{P} = \frac{X}{n}$, სადაც X წარმატებათა რაოდენობაა შერჩევაში, ხოლო n - დაკვირვებათა რაოდენობა.

ბინომურად განაწილებული შემთხვევითი სიდიდე $X \equiv \text{Bin}(n, p)$, როცა $n \geq 30$ ნორმალურად არის განაწილებული საშუალოთი np და დისპერსიით $np(1-p)$. მოითხოვება, რომ სრულდებოდეს პირობა $0 \leq np \pm 3\sqrt{np(1-p)} \leq n$ ან უნდა სრულდებოდეს უტოლობები: $np \geq 5$ და $n(1-p) \geq 5$.

$\hat{P}$ -ს აქვს ნორმალური განაწილებასაშუალოთი p და სტანდარტული გადახრით $\sqrt{\frac{p(1-p)}{n}}$. $\hat{P}$ -ის სტანდარტიზაციით მივიღებთ:

$$\frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{a}{\approx} N(0, 1).$$

ავაგოთ ნდობის ინტერვალი უცნობი p ალბათობისათვის. შევნიშნოთ, რომ $0 < p < 1$ და $0 < \hat{p} < 1$, ამიტომ ნდობის ინტერვალის ასაგებად უნდა სრულდებოდეს პირობა

$$0 \leq p \pm 3\sqrt{\frac{p(1-p)}{n}} \leq 1.$$

95%-იანი ნდობის ინტერვალი უცნობი ალბათობისათვის (პოპულაციის პროპორციისთვის) არის:

$$\hat{p} \mp 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

შევამოწმოთ ჰიპოთეზა:

$$H_0: p = p_0, \quad H_A: p < p_0 \quad \text{ან} \quad H_A: p > p_0 \quad \text{ან} \quad H_A: p \neq p_0.$$

კრიტერიუმის სტატისტიკაა:

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}.$$

Z -ს აქვს სტანდარტული ნორმალური განაწილება.

მაგალითი 9.12. სატელეფონო კომპანიის შეფასებით, მათი მომხმარებლების 40%-ს გააჩნია ლოდინის რეჟიმის მქონე მონყობილობა. ამ პიპოთეზის შესამოწმებლად შეირჩა 100 მომხმარებელი და გაირკვა, რომ მათ შორის 37%-ს აქვს ლოდინის რეჟიმის მქონე მონყობილობა. $\alpha = 0.05$ მნიშვნელობების დონით გვაქვს თუ არა საკმარისი საფუძველი უკუვაგდოთ კომპანიის მოსაზრება?

ამოხსნა: ჩამოვაყალიბოთ პიპოთეზები:

$$H_0: p = 40\%, \quad H_A: p \neq 40\%.$$

გამოვთვალოთ კრიტერიუმის სტატისტიკის მნიშვნელობა:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.37 - 0.40}{\sqrt{\frac{0.40(1-0.40)}{100}}} = -0.612.$$

H_0 პიპოთეზას ვერ უარვყოფთ.

აევაგოთ 95%-იანი ნდობის ინტერვალი

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.37 \pm 1.96 \cdot \sqrt{\frac{0.37 \cdot (1-0.37)}{100}} = (0.275, 0.465).$$

როგორც ვხედავთ, ეს ინტერვალი მოიცავს $p = 40\%$ -ს.

პიპოთეზათა შემოწმება პოპულაციის პროპორციისათვის, როცა შერჩევის მოცულობა პატარაა

როცა შერჩევის მოცულობა პატარაა, არა გვაქვს ცზთ-ის გამოყენების უფლება. ამ შემთხვევაში უნდა ვისარგებლოთ იმ ფაქტით, რომ წარმატებათა რაოდენობას შერჩევაში აქვს ბინომური განაწილება.

განვიხილოთ $H_0: p = p_0$ ნულოვანი პიპოთეზა ალტერნატივის სანინალმდეგოდ $H_A: p < p_0$ ან $H_A: p > p_0$ ან $H_A: p \neq p_0$.

ცალმხრივი კრიტერიუმის სტატისტიკა არის შერჩევაში წარმატებათა რაოდენობა X . ორმხრივი ალტერნატივისათვის კრიტერიუმის სტატისტიკაა მინიმალური მნიშვნელობა წარმატებათა რაოდენობასა და მარცხის რაოდენობას შორის.

X -ს აქვს ბინომური განაწილება პარამეტრებით n და p_0 .

p -სიდიდე = $P(\text{Bin}(n, p_0) \geq X)$, თუ $H_A: p > p_0$,

p -სიდიდე = $P(\text{Bin}(n, p_0) \leq X)$, თუ $H_A: p < p_0$,

p -სიდიდე = $P(\text{Bin}(n, p_0) \geq X) + P(\text{Bin}(n, p_0) \leq n - X)$, თუ $H_A: p \neq p_0$.

მაგალითი 9.13. როცა დაიბადა გიორგის ქალიშვილი, ის იყო ერთადერთი გოგონა იმ დღეს სამშობიაროში დაბადებულ 8 ბავშვს შორის. არღვევს თუ არა ეს ფაქტი მოსაზრებას, რომ გოგონას დაბადების ალბათობა უდრის 0.5-ს.

ამოხსნა: შერჩევის მოცულობა პატარაა, ამიტომ უნდა გამოვიყენოთ ბინომური განაწილება და შევამოწმოთ ჰიპოთეზა: $H_0: p = 0.5$ ალტერნატივის – $H_A: p \neq 0.5$ სანინალმდეგოდ.

ალტერნატივა ორმხრივია, ამიტომ P -სიდიდე უდრის

$$P(\text{Bin}(8, 0.5) \leq 1) + P(\text{Bin}(8, 0.5) \geq 7) = \\ = 0.0039 + 0.0313 + 0.0313 + 0.0039 = 0.0704.$$

P -სიდიდე მეტია, ვიდრე $\alpha = 0.05$, ამიტომ ნულოვან ჰიპოთეზას ვერ უარყოფთ 0.05 მნიშვნელოვნების დონეზე. მიუხედავად იმისა, რომ ფაქტი შეიძლება უცნაურად მოგვეჩვენოს, შერჩევის პატარა მოცულობა არ გვაძლევს საკმარის მტკიცებულებას ნულოვანი ჰიპოთეზის უარყოფისათვის.

სასრული პოპულაციები

ნდობის ინტერვალის აგებისას და ჰიპოთეზათა შემოწმებისას გვულისხმობდით, რომ პოპულაციის მოცულობა იყო ძალიან დიდი (უსასრულო). თუ პოპულაციის მოცულობა არ არის დიდი, ყველა მიღებულ ფორმულაში უნდა შევიტანოთ მამრავლი (finite population correction factor).

ეს მამრავლი სასრული N მოცულობის მქონე პოპულაციისთვის უდრის

$$\sqrt{\frac{N-n}{N-1}} \text{ -ს. ის ყოველთვის ერთზე ნაკლებია დაშერჩევის მოცულობასთან შედარე-}$$

ბით პოპულაციის მოცულობის გაზრდისას უახლოვდება ერთს.

სასრულო პოპულაციისათვის შერჩევის საშუალოს სტანდარტული შეცდომა

$$\text{იქნება } \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}. \text{ 95\%-იანი ნდობის ინტერვალი } \mu \text{-თვის არის } \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

$$\text{კრიტერიუმის სტატისტიკაა } \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}.$$

p -ს შეფასების სტანდარტული შეცდომაა:

$$\sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}.$$

95%-იანი ნდობის ინტერვალია:

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N-1}}$$

და კრიტერიუმის სტატისტიკაა:

$$\frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}}.$$

მენიშენა 9.4. თუ $n < 5\%N$, საკორექციო მამრავლი არ არის საჭირო.

მაგალითი 9.14. ფირმაში მომუშავე 912 თანამშრომლიდან შეარჩიეს 50 პროფესიონალი, რომელთა საშუალო ასაკი იყო 43.7 წელი. ფირმის ყველა თანამშრომლის საშუალო ასაკია 31.6 წელი, სტანდარტული გადახრა 9.2 წელი. მნიშვნელოვნად განსხვავდებიან თუ არა საშუალოები ერთმანეთისგან?

ამოხსნა: შესამოწმებელია ჰიპოთეზა $H_0: \mu = 31.6$ და $H_A: \mu > 31.6$.

$N = 912$, $n = 50$, $\bar{x} = 43.7$, $\sigma = 9.2$. კრიტერიუმის სტატისტიკის მნიშვნელობაა:

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}} = \frac{43.7 - 31.6}{9.2 \sqrt{\frac{912-50}{912-1}}} = 9.56.$$

მიღებული სიდიდე გაცილებით მეტია 1%-იან კრიტიკულ სიდიდეზე $Z = 2.3263$. შესაბამისად, პროფესიონალების საშუალო ასაკი გაცილებით მეტია ფირმის თანამშრომლების საშუალო ასაკზე.

მაგალითი 9.15. ფირმაში თანამშრომლების ნახევარი ქალბატონებია. 50 პროფესიონალიდან (მაგალითი 9.14) 31 ქალბატონია. შეამოწმეთ, ჰიპოთეზა ალემატება თუ არა პროფესიონალი ქალბატონების წილის 0.5-ს.

ამოხსნა: $H_0: p = 0.5$, $H_A: p > 0.5$.

გვაქვს $\hat{p} = \frac{X}{n} = \frac{31}{50} = 0.62$. კრიტერიუმის სტატისტიკის მნიშვნელობაა:

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}} = \frac{0.62 - 0.5}{\sqrt{\frac{0.5 \cdot (1-0.5)}{50}} \sqrt{\frac{912-50}{912-1}}} = \frac{0.12}{0.071 \cdot 0.973} = 1.74.$$

P -სიდიდე უდრის $P(Z > 1.74) = 0.5 - 0.4591 = 0.0409$ და ნულოვან ჰიპოთეზას 5% დონით ვერ უარეყოფთ.

ცდომილება, სიზუსტე და შერჩევის მოცულობა

როცა შერჩევის მოცულობა დიდია, $1 - \alpha$ საიმედოობის ნდობის ინტერვალს აქვს შემდეგი ფორმა:

- ნერტილოვანი შეფასება $\pm z_{\alpha/2}$ (z - $\alpha/2$ კრიტიკული ნერტილი) $\times$ (გამრავლებული) ნერტილოვანი შეფასების სტანდარტულ შეცდომაზე,
- ნერტილოვანი შეფასება $\mp$ ცდომილება, ან
- ნერტილოვანი შეფასება $\mp$ სიზუსტე.

ცდომილება დამოკიდებულია Z -ზე (რომელიც დამოკიდებულია α -ზე) და სტანდარტულ შეცდომაზე:

$\frac{\sigma}{\sqrt{n}}$ – სტანდარტული შეცდომა საშუალოსათვის,

$\sqrt{\frac{p(1-p)}{n}}$ – სტანდარტული შეცდომა ალბათობისათვის.

ნდობის ინტერვალის სიგრძე – $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ მით უფრო დიდია, რაც უფრო დიდია

$z_{\alpha/2}$, σ მით უფრო დიდია, რაც უფრო დიდია საიმედოობა – $1-\alpha$, ან რაც უფრო პატარაა α . რაც უფრო განიერია ნდობის ინტერვალი, მით უფრო საიმედოა შეფასება, მაგრამ, ამავე დროს, უფრო დიდია ცდომილება.

90%-იანი ($1-\alpha=90/100=0.9$, $\alpha=0.1$, $\alpha/2=0.05$, $z_{0.05}=1.65$) ნდობის ინტერვალის სიგრძეა $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}=2 \cdot 1.65 \frac{\sigma}{\sqrt{n}}$;

95%-იანი ($\alpha=0.05$, $\alpha/2=0.025$, $z_{0.025}=1.96$) ნდობის ინტერვალის სიგრძეა $2 \cdot 1.96 \frac{\sigma}{\sqrt{n}}$;

99%-იანი ($\alpha=0.01$, $\alpha/2=0.005$, $z_{0.005}=2.58$) ნდობის ინტერვალის სიგრძეა $2 \cdot 2.58 \frac{\sigma}{\sqrt{n}}$.

ნდობის ინტერვალის სიგრძის შემცირება (სიზუსტის გაზრდა) შესაძლებელია შერჩევის n მოცულობის გაზრდით.

ბუნებრივია, ისმის კითხვა: შესაძლებელია თუ არა ნდობის ფიქსირებული დონისათვის მოიძებნოს ის მინიმალური n' მოცულობა, რომელიც უზრუნველყოფს შეფასების წინასწარ ფიქსირებულ m სიზუსტეს. n' მოსაძებნია შემდეგი განტოლებიდან: $z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = m$, საიდანაც მივიღებთ, რომ $n' = \left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2$ (თუ $\left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2$

არაა მთელი, მაშინ უნდა ავიღოთ – $n' = \left[\left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2\right] + 1$). სასრული პოპულაციისათ-

ვის $n' = \frac{n}{1 + n/N}$, სადაც $n' = \left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2$ (თუ $\frac{n}{1 + n/N}$ არაა მთელი, მაშინ უნდა ავიღოთ – $n' = \left[\frac{n}{1 + n/N}\right] + 1$).

მაგალითი 9.16. რისი ტოლი უნდა იყოს შერჩევის მოცულობა იმისათვის, რომ შევაფასოთ ქალების პროპორცია 2%-იანი სიზუსტით და 5%-იანი მნიშვნელოვნების დონით 9.13 მაგალითში, სადაც $N=912$?

ამოხსნა: ჯერ ვიგულისხმოთ, რომ პოპულაცია არ არის სასრული. $m=2\%=0.02$, ავიღოთ $p=0.5$, $\alpha=0.05$, $z_{0.025}=1.96$. $\sigma^2=p(1-p)=0.5 \cdot 0.5=0.25$,

$$n' = \frac{1.96^2}{0.02^2} \cdot 0.25 = 2401.$$

ვინაიდან ეს რაოდენობა გაცილებით მეტია პოპულაციის მოცულობაზე, ამიტომ მოვახდინოთ კორექცია:

$$n = \frac{n'}{1 + \frac{n'}{N}} = \frac{2401}{1 + \frac{2401}{912}} = 660.95 \approx 661.$$

იმისათვის, რომ შევაფასოთ ქალების პროპორცია 2% -იანი სიზუსტით და 95% -იანი საიმედოობით, 912 თანამშრომლიდან უნდა გამოვკითხოთ 661.

მაგალითი 9.17. კავშირი შერჩევის სტატისტიკებს და პოპულაციის პარამეტრებს შორის.

ერთი და იმავე პოპულაციიდან აღებული ორი შერჩევის სტატისტიკა, როგორც წესი, განსხვავდება ერთმანეთისაგან და ამან შეიძლება მიგვიყვანოს სხვადასხვა გადანუვებითილებამდე.

ვთქვათ, საპენსიო ფონდის თანხა ჩადებულია სხვადასხვა აქციაში. ცხრილში მოცემულია შარშანდელი აქციიდან თანხის დაბრუნება %-ში:

აქცია	1	2	3	4	5
დაბრუნება	7%	12%	-3%	21%	3%

პოპულაციის საშუალო

$$= \frac{7\% + 12\% - 3\% + 21\% + 3\%}{5} = 8\%.$$

პოპულაციის სტანდარტული გადახრა

$$= \sqrt{\frac{(7-8)^2 + (12-8)^2 + (-3-8)^2 + (21-8)^2 + (3-8)^2}{5}} = 8.15\%.$$

შევარჩიოთ სამი აქცია და გამოვთვალოთ შერჩევის საშუალო.

შერჩეული აქციები	დაბრუნება			$\bar{X}$
1, 2, 3	7%	12%	-3%	5.33
1, 2, 4	7%	12%	21%	13.33
1, 2, 5	7%	12%	3%	7.33
1, 3, 4	7%	-3%	21%	8.33
1, 3, 5	7%	-3%	3%	2.33
1, 4, 5	7%	21%	3%	10.33
2, 3, 4	12%	-3%	21%	10.00
2, 3, 5	12%	-3%	3%	4.00
2, 4, 5	12%	21%	3%	12.00
3, 4, 5	-3%	21%	3%	7.00

სულ გვაქვს 10 შესაძლო შერჩევა. გამოვთვალოთ ამ 10 შერჩევის საშუალოების საშუალო. ის ტოლია:

$$\frac{5.33 + 13.33 + 7.33 + 8.33 + 2.33 + 10.33 + 10.00 + 4.00 + 12.00 + 7.00}{10} = 0.08 = 8\%.$$

ჩვენ ვხედავთ, რომ საშუალოების საშუალო უდრის პოპულაციის საშუალოს. ახლა გამოვთვალოთ საშუალოების სტანდარტული გადახრა:

$$\sqrt{\frac{(5.33-8)^2 + (13.33-8)^2 + \dots + (7-8)^2}{10}} = \sqrt{11.07} = 3.327.$$

მიაქციეთ ყურადღება, რომ მიღებული საშუალოების სტანდარტული გადახრა 3.326. მას უნოდებენ საშუალოს სტანდარტულ შეცდომას. ის ნაკლებია პოპულაციის სტანდარტულ შეცდომაზე - 8.15.

თუ გავითვალისწინებთ, რომ პოპულაცია არის სასრული და გამოვიყენებთ შესწორებას $-\sqrt{\frac{N-n}{N-1}}$, საშუალოს სტანდარტული შეცდომა იქნება:

$$\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{8.15}{\sqrt{3}} \sqrt{\frac{5-3}{5-1}} = 3.327.$$

ჰიპოთეზათა შემოწმება დისპერსიისათვის

სტატისტიკური ჰიპოთეზის შემოწმება ნორმალური $N(a, \sigma^2)$ განაწილების დისპერსიის შესახებ:

$$\text{ჰიპოთეზა: } H_0: \sigma^2 = \sigma_0^2$$

მნიშვნელოვნების დონე: α

$$\text{კრიტერიუმის სტატისტიკა: } \begin{cases} \frac{\sum_{i=1}^n (X_i - a)^2}{\sigma_0^2} \equiv \chi^2(n), \text{ თუ } a \text{ ცნობილია,} \\ \frac{(n-1)S^2}{\sigma_0^2} \equiv \chi^2(n-1), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

$$\text{კრიტერიუმის მნიშვნელობა T.V.} = \begin{cases} \frac{\sum_{i=1}^n (x_i - a)^2}{\sigma_0^2}, \text{ თუ } a \text{ ცნობილია,} \\ \frac{(n-1)s^2}{\sigma_0^2}, \text{ თუ } a \text{ უცნობია.} \end{cases}$$

ალტერნატივა კრიტიკული არე C.R. (H_0 -ის უარყოფის არე)

$$H_A: \sigma^2 > \sigma_0^2 \begin{cases} [\chi_{n-1, \alpha}^2, +\infty), \text{ თუ } a \text{ ცნობილია,} \\ [\chi_{n-1, \alpha}^2, +\infty), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

$$H_A: \sigma^2 < \sigma_0^2 \begin{cases} (0, \chi_{n-1, \alpha}^2], \text{ თუ } a \text{ ცნობილია,} \\ (0, \chi_{n-1, 1-\alpha}^2], \text{ თუ } a \text{ უცნობია.} \end{cases}$$

$$H_A: \sigma^2 \neq \sigma_0^2 \begin{cases} (0, \chi_{n-1, \alpha/2}^2] \cup [\chi_{n-1, 1-\alpha/2}^2, +\infty), \text{ თუ } a \text{ ცნობილია,} \\ (0, \chi_{n-1, \alpha/2}^2] \cup [\chi_{n-1, 1-\alpha/2}^2, +\infty), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

(სადაც $\chi_{n, \alpha}^2$ არის $\chi^2(n)$ განაწილების ზედა α -კრიტიკული წერტილი ანუ კრიტიკული მნიშვნელობა C.V.).

ალტერნატივა P - მნიშვნელობა:

$$H_1: \sigma^2 > \sigma_0^2 \quad P = \begin{cases} P\{\chi^2(n) > T.V.\}, \text{ თუ } a \text{ ცნობილია,} \\ P\{\chi^2(n-1) > T.V.\}, \text{ თუ } a \text{ უცნობია.} \end{cases}$$

$$H_1: \sigma^2 < \sigma_0^2 \quad P = \begin{cases} P\{\chi^2(n) < T.V.\}, \text{ თუ } a \text{ ცნობილია,} \\ P\{\chi^2(n-1) < T.V.\}, \text{ თუ } a \text{ უცნობია.} \end{cases}$$

$$H_1: \sigma^2 \neq \sigma_0^2 \quad P = \begin{cases} 2 \cdot \min\{P\{\chi^2(n) > T.V.\}, P\{\chi^2(n) < T.V.\}\}, \text{ თუ } a \text{ ცნობილია,} \\ 2 \cdot \min\{P\{\chi^2(n-1) > T.V.\}, P\{\chi^2(n-1) < T.V.\}\}, \text{ თუ } a \text{ უცნობია.} \end{cases}$$

გადაწყვეტილება: თუ $T.V. \in C.R.$, მაშინ H_0 ჰიპოთეზას უკუვაგდებთ, წინააღმდეგ შემთხვევაში ამის საფუძველი არა გვაქვს.

P - მნიშვნელობის მეთოდი: თუ $P \leq \alpha$, მაშინ H_0 ჰიპოთეზას უკუვაგდებთ, წინააღმდეგ შემთხვევაში ამის საფუძველი არა გვაქვს.

მაგალითი 9.18. პროფესორის სურს თავის ჯგუფში, სადაც 23 სტუდენტია, გარკვეოს ქულების ცვალებადობა არის თუ არა ნაკლები, ვიდრე პოპულაციის დისპერსია 225. პროფესორის ჯგუფში ქულების შესწორებული დისპერსიაა 198. $\alpha = 0.05$ მნიშვნელოვნების დონით გვაქვს თუ არა საკმარისი საფუძველი ვამტკიცოთ, რომ პროფესორის ჯგუფში ქულების ცვალებადობა ნაკლებია პოპულაციის დისპერსიაზე? ჩათვალით, რომ ქულები ნორმალურადაა განაწილებული.

ამოხსნა: ჩამოვყალიბოთ ძირითადი და ალტერნატიული ჰიპოთეზები:

$H_0: D\xi \geq 225$, $H_1: D\xi < 225$. ვიპოვოთ კრიტიკული მნიშვნელობა: ვინაიდან კრიტერიუმი მარცხენა ცალმხრივია, $1 - \alpha = 0.95$ და თავისუფლების ხარისხი $n - 1 = 22$, ამიტომ ხი-კვადრატ განაწილების ზედა კრიტიკული წერტილების ცხრილიდან ეპოულობთ, რომ $\chi_{n-1, 1-\alpha}^2 = \chi_{22, 0.95}^2 = 12.338$. მაშასადამე, კრიტიკული არეა $C.R. = (0, 12.338]$. გამოვთვალოთ კრიტერიუმის მნიშვნელობა:

$$T.V. = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(23-1) \cdot 198}{225} = 19.36.$$

ვინაიდან კრიტერიუმის მნიშვნელობა არ ვარდება კრიტიკულ არეში, ამიტომ ნულოვან ჰიპოთეზას ვღებულობთ, ანუ არ არსებობს საკმარისი საფუძველი იმი-

სათვის, რომ პროფესორის ჯგუფში ქულების ცვალებადობა ნაკლებია პოპულაციის დისპერსიაზე.

მაგალითი 9.19. საავადმყოფოს ადმინისტრატორი დარწმუნებულია, რომ სტანდარტული გადახრა იმ პაციენტების რიცხვის, რომლებიც ამბულატორიულად მკურნალობენ, დღეში 8-ზე მეტია. ჩავთვალოთ, რომ პაციენტების რიცხვი განაწილებულია ნორმალურად. $\alpha = 0.1$ მნიშვნელობების დონით შეიძლება თუ არა გავიზიაროთ ადმინისტრატორის მოსაზრება, თუ 15 შემთხვევით შერჩეული დღის მონაცემებია:

25	30	5	15	18
42	16	9	10	12
12	38	8	14	27

ამოხსნა: ჩამოვაცალიბოთ ძირითადი და ალტერნატიული ჰიპოთეზები: $H_0: D\bar{x} \leq 64$, $H_1: D\bar{x} > 64$. რამდენადაც კრიტერიუმის მარჯვენა ცალმხრივია, $\alpha = 0.1$ და თავისუფლების ხარისხი $n-1=14$, ამიტომ კრიტიკული მნიშვნელობა იქნება: $\chi^2_{n-1,\alpha} = 21.0641$. გამოვთვალოთ შესწორებული სტანდარტული გადახრა, გვაქვს:

$$s' = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \frac{1}{n} \sum_{j=1}^n x_j)^2} = 11.2.$$

ამიტომ კრიტერიუმის მნიშვნელობა იქნება:

$$T.V. = \frac{(n-1)s'^2}{\sigma_0^2} = \frac{(15-1) \cdot (11.2)^2}{8^2} = 27.44.$$

ვინაიდან, კრიტიკული მნიშვნელობა ვარდება კრიტიკულ არეში ($27.44 \in [21.0641, +\infty)$), ამიტომ H_0 უნდა უკუვაგდოთ, ანუ ჩვენ გვაქვს საკმარისი საფუძველი, რათა ჩავთვალოთ, რომ პოპულაციის სტანდარტული გადახრა მეტია 8-ზე.

შეჯამება

შერჩევის ნებისმიერ ფუნქციას **სტატისტიკა** ეწოდება. სტატისტიკას, რომელიც შეიძლება ჩაითვალოს განაწილების პარამეტრის მიახლოებად და გამოყენებულ იქნეს უცნობი პარამეტრის ნაცვლად, **ნერტილოვანი შეფასება** ეწოდება.

ასე, მაგალითად, შერჩევის საშუალო $\bar{x}$ არის პოპულაციის μ საშუალოს ნერტილოვანი შეფასება; შერჩევის პროპორცია $\hat{p}$ არის პოპულაციის p პროპორციის (ალბათობის) ნერტილოვანი შეფასება; შერჩევის სტანდარტული გადახრა s არის პოპულაციის σ სტანდარტული გადახრის შეფასება.

95%-იანი ნდობის ინტერვალი ფარავს შესაფასებელ უცნობ პარამეტრს 0.95-ზე არანაკლები ალბათობით.

95%-იანი ნდობის ინტერვალი ნორმალური პოპულაციის μ საშუალოსათვის, როცა შერჩევის მოცულობა დიდია, არის:

$$\left(\bar{x} - 1.96 \frac{s}{\sqrt{n}}, \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right)$$

(საზოგადოდ, $1-\alpha$ საიმედოობის ნდობის ინტერვალი:

$$\left(\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right).$$

$$s - \text{შერჩევის სტანდარტული გადახრა } s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}.$$

$1-\alpha$ საიმედოობის ნდობის ინტერვალი ნორმალური პოპულაციის საშუალოსათვის, როცა შერჩევის მოცულობა პატარაა ($n < 30$), არის:

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right),$$

$$\left(\bar{x} - t_{n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1} \frac{s}{\sqrt{n}} \right).$$

$t_{n-1, \alpha/2}$ -ის მნიშვნელობა უნდა მოიძებნოს სტიუდენტის განაწილების ზედა კრიტიკული წერტილების ცხრილიდან.

შერჩევის საჭირო მინიმალური მოცულობა შეფასების წინასწარ დაფიქსირებული m სიზუსტის და 95%-იანი საიმედოობის უზრუნველსაყოფად არის:

$$n' = \left(\frac{\sigma}{m} \cdot 1.96 \right)^2,$$

(საზოგადოდ, $n' = \left(\frac{z_{\alpha/2}}{m} \right)^2 \sigma^2$).

რაიმე მოსაზრებას პოპულაციის განაწილების პარამეტრის შესახებ სტატისტიკური ჰიპოთეზა ეწოდება. ჰიპოთეზათა შემოწმების ამოცანა მდგომარეობს ისეთი წესის – სტატისტიკური კრიტერიუმის ჩამოყალიბებაში, რომელიც საშუალებას მოგვცემს ჩატარებულ დაკვირვებებზე დაყრდნობით შევამოწმოთ გამოთქმული მოსაზრება.

H_0 ჰიპოთეზა ამტკიცებს, რომ არ არსებობს განსხვავება პარამეტრსა და მის კონკრეტულ მნიშვნელობას შორის. ალტერნატიული H_A ჰიპოთეზა ამტკიცებს, რომ არსებობს განსხვავება პარამეტრსა და მის კონკრეტულ მნიშვნელობას შორის.

სამართლიანი H_0 ჰიპოთეზის უკუგდებას პირველი გვარის შეცდომა ეწოდება. პირველი გვარის შეცდომის დაშვების ალბათობას მნიშვნელოვნების დონე ეწოდება და აღინიშნება α ასოთი. α -ს ირჩევენ იმდენად მცირეს, რომ α -ს ტოლი

ალბათობის მქონე ხდომილების მოხდენა, თუ H_0 სამართლიანია, პრაქტიკულად შეუძლებლად ითვლება. არასამართლიანი H_0 -ის მიღებას მეორე გვარის შეცდომა ეწოდება. მისი დაშვების ალბათობა β ასოთი აღინიშნება. $1 - \beta$ არის ალბათობა იმისა, რომ არ იქნება დაშვებული მეორე გვარის შეცდომა. მასკრიტიერიუმის სიმძლავრე ეწოდება.

სტატისტიკური ჰიპოთეზის შემოწმება ნორმალური პოპულაციის საშუალოს შესახებ, როცა $n \geq 30$: $H_0: \mu = \mu_0$,

$$\text{კრიტერიუმის სტატისტიკა: } Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}.$$

ალტერნატივა	H_0 -ის უარყოფის არე
ორმხრივი ალტერნატიული ჰიპოთეზა $H_A: \mu \neq \mu_0$	$z \leq -z_{\alpha/2}$ ან $z \geq z_{\alpha/2}$
მარჯვენა ცალმხრივი $H_A: \mu > \mu_0$	$z \geq z_{\alpha}$
მარცხენა ცალმხრივი $H_A: \mu < \mu_0$	$z \leq -z_{\alpha}$

$z_{\alpha/2}$ და z_{α} სიდიდეებს პოულობენ სტანდარტული ნორმალური განაწილების ცხრილების საშუალებით ($\Phi(z_{\alpha}) = 1 - \alpha$, $\Phi_0(z_{\alpha}) = 0.5 - \alpha$).

სტატისტიკური ჰიპოთეზის შემოწმება ნორმალური პოპულაციის საშუალოს შესახებ, როცა $n < 30$: $H_0: \mu = \mu_0$,

$$\text{კრიტერიუმის სტატისტიკა: } t(n-1) = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

ალტერნატივა	H_0 -ის უარყოფის არე
ორმხრივი ალტერნატიული ჰიპოთეზა $H_A: \mu \neq \mu_0$	$t \leq -t_{n-1, \alpha/2}$ ან $t \geq t_{n-1, \alpha/2}$
მარჯვენა ცალმხრივი $H_A: \mu > \mu_0$	$t \geq t_{n-1, \alpha}$
მარცხენა ცალმხრივი $H_A: \mu < \mu_0$	$t \leq -t_{n-1, \alpha}$

$t_{n-1, \alpha/2}$ და $t_{n-1, \alpha}$ სიდიდეებს პოულობენ სტიუდენტის განაწილების ზედა კრიტიკული წერტილების ცხრილიდან.

P მნიშვნელობის მეთოდი: თუ $p \leq \alpha$, მაშინ H_0 ჰიპოთეზას უკუევადებთ α მნიშვნელოვნების დონით, წინააღმდეგ შემთხვევაში ამის საუძველი არა გვაქვს.

95%-იანი ნდობის ინტერვალი უცნობი p -ალბათობისთვის, როცა $n \geq 30$

არის $\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, სადაც $\hat{p}$ შერჩევის პროპორციაა.

p -ალბათობისთვის m სიზუსტისა და 95%-იანი საიმედოობის ნდობის ინტერვალის ასაგებად საჭირო შერჩევის მინიმალური მოცულობაა:

$$n' = \hat{p}(1-\hat{p}) \frac{1.96^2}{m^2}$$

(საზოგადოდ, $n' = \hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{m}\right)^2$).

ჰიპოთეზათა შემოწმება წარმატების ალბათობისათვის, როცა $n \geq 30$:

$$H_0: p = p_0. \text{ კრიტერიუმის სტატისტიკა: } Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

ორმხრივი ალტერნატიული ჰიპოთეზა $H_A: p \neq p_0$	$z \leq -z_{\alpha/2}$ ან $z \geq z_{\alpha/2}$
მარჯვენა ცალმხრივი $H_A: p > p_0$	$z \geq z_\alpha$
მარცხენა ცალმხრივი $H_A: p < p_0$	$z \leq -z_\alpha$

$z_{\alpha/2}$ და z_α სიდიდეებს პოულობენ სტანდარტული ნორმალური განაწილების ცხრილების საშუალებით ($\Phi(z_\alpha) = 1 - \alpha$, $\Phi_0(z_\alpha) = 0.5 - \alpha$).

ჰიპოთეზათა შემოწმება წარმატების ალბათობისათვის, როცა $n < 30$. კრიტერიუმის სტატისტიკა:

ცალმხრივი ალტერნატივისათვის – წარმატებათა რაოდენობა შერჩევაში.

ორმხრივი ალტერნატივისათვის – მინიმალური წარმატებათა რაოდენობასა და მარცხის რაოდენობას შორის.

წარმატებათა რაოდენობას აქვს ბინომიალური განაწილება პარამეტრებით p_0, n .

სასრული პოპულაციისათვის ყველა სტატისტიკა მრავლდება მამრავლზე

$$\sqrt{\frac{N-n}{N-1}}.$$

α მნიშვნელოვნების დონის მქონე ნდობის ინტერვალი ნორმალური $N(\alpha, \sigma^2)$ პოპულაციის დისპერსიისათვის უცნობი საშუალოს შემთხვევაში:

$$\left(\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2} \right).$$

სტატისტიკური ჰიპოთეზის შემოწმება ნორმალური $N(a, \sigma^2)$ განაწილების დისპერსიის შესახებ:

$$\text{ჰიპოთეზა: } H_0 : \sigma^2 = \sigma_0^2$$

მნიშვნელოვნების დონე: α

$$\text{კრიტერიუმის სტატისტიკა: } \begin{cases} \frac{\sum_{i=1}^n (X_i - a)^2}{\sigma_0^2} \equiv \chi^2(n), \text{ თუ } a \text{ ცნობილია,} \\ \frac{(n-1)S^2}{\sigma_0^2} \equiv \chi^2(n-1), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

ალტერნატივა კრიტიკული არე C.R. (H_0 -ის უარყოფის არე)

$$H_A : \sigma^2 \neq \sigma_0^2 \quad \begin{cases} (0, \chi_{n-1, 1-\alpha/2}^2] \cup [\chi_{n-1, \alpha/2}^2, +\infty), \text{ თუ } a \text{ ცნობილია.} \\ (0, \chi_{n-1, 1-\alpha/2}^2] \cup [\chi_{n-1, \alpha/2}^2, +\infty), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

(სადაც $\chi_{n, \alpha}^2$ არის $\chi^2(n)$ განაწილების ზედა α -კრიტიკული წერტილი).

კითხვები

- პოპულაცია ნორმალურად არის განაწილებული: $\mu = 58$, $\sigma = 18$.
 - თუ პოპულაციიდან აღებულია ცხრის ტოლი მოცულობის შერჩევები, რას უდრის ყველა შესაძლო შერჩევების საშუალო და სტანდარტული გადახრა?
 - რისი ტოლია საშუალოს შეფასების სტანდარტული შეცდომა?
 - როგორ შეიცვლება გამოთვლილი სიდიდეები, თუ შერჩევების მოცულობა 36-მდე გაიზარდება?
 - რისი ტოლი უნდა იყოს შერჩევის მოცულობა იმისათვის, რომ ბ) პუნქტში გამოთვლილი საშუალოს სტანდარტული შეცდომა ორჯერ შემცირდეს?
- ქარხანაში 3000 ადამიანია დასაქმებული. დასაქმებულების ნლოვანება ნორმალურად არის განაწილებული – საშუალოთი 31 წელი და სტანდარტული გადახრით 4.3 წელი. სადაზღვევო კომპანიას სურს ახალი სადაზღვევო პაკეტის შემოღება და ფასის დასადგენად კომპანიამ შემთხვევით შეარჩია 25 მუშა.
 - რას უდრის ალბათობა იმისა, რომ 25 მუშის საშუალო ასაკი იქნება 31-ზე ნაკლები?
 - რას უდრისამ შერჩევასთან დაკავშირებული სტანდარტული შეცდომა?
 - რა უნდა გაკეთდეს შეცდომის შესამცირებლად?
- პოპულაციის მოცულობა უდრის 100-ს, საშუალო 17.5-ია, სტანდარტული შეცდომა კი – 2.2. პოპულაციიდან აღებულია 35მოცულობის მქონე შერჩევა.
 - რას უდრის ყველა 35 მოცულობის მქონე შერჩევების საშუალოების საშუალო?
 - რას უდრის შერჩევების საშუალოების განაწილების სტანდარტული გადახრა?
 - სტანდარტული გადახრის გამოთვლისას საჭიროა თუ არა იმის გათვალისწინება, რომ პოპულაცია არის სასრული?

4. ტესტების კრებულის რეკლამაში წერია, რომ ამ ტესტების დამუშავების შედეგად სტუდენტის ქულა მათემატიკაში საშუალოდ 40-ით გაიზრდება, 15-ის ტოლი სტანდარტული გადახრით. ერთი სტუდენტი დარჩა უქმყოფილო, რადგანაც მისი ქულა საერთოდ არ შეცვლილა.
 - ა) რას უდრის ალბათობა იმისა რომ ერთი სტუდენტის ქულა არ შეიცვლება?
 - ბ) რას უდრის ალბათობა იმისა, რომ 30 სტუდენტის საშუალო ქულა 32-ით ან მეტით გაიზრდება?
5. სტუდენტებს დაავალეს ჩაეტარებინათ გამოკითხვა ქალაქში, სადაც ცნობილი იყო, რომ ამომრჩეველთა 60% „დემოკრატია“. სტუდენტების ანგარიშში ეწერა, რომ 760 რესპონდენტიდან 395 „დემოკრატია“. თუ სტუდენტების შერჩევა შემთხვევითია, რას უდრის შერჩევაში 395 ან ნაკლები „დემოკრატის“ მიღების ალბათობა.
6. ფირმის 240 თანამშრომლის საშუალო წლიური შემოსავალი 21000 დოლარია, 5000 დოლარი სტანდარტული გადახრით.
 - ა) რას უდრის ალბათობა იმისა, რომ 100 თანამშრომლის საშუალო წლიური შემოსავალი 21000 დოლარს გადააჭარბებს?
 - ბ) რას უდრის ალბათობა იმისა, რომ 100 თანამშრომლის საშუალო წლიური შემოსავალი 22000 დოლარზე ნაკლები იქნება?
 - გ) საჭიროა თუ არა და როგორ იმის გათვალისწინება, რომ პოპულაცია არის სასრული?
7. ჟურნალისტმა შეაგროვა 40 საგაზეთო წერილი უცხოპლანეტელების შესახებ. თუ წერილების სტრიქონების რაოდენობის სტანდარტული გადახრა უდრის 8-ს, რას უდრის წერილის სტრიქონების რაოდენობის საშუალოს სტანდარტული შეცდომა?
8. მსხვილფეხა რქოსანი პირუტყვის ფერმაში 200 ხბო გადაყვანილი იყო კვების ახალ რეჟიმზე. წონის დღიურმა მატებამ შეადგინა 1.2 ფუნტი, 0.50 ფუნტი სტანდარტული გადახრით. ააგეთ 95%-იანი ნდობის ინტერვალი წონის დღიური მატებისთვის.
9. კომპანიაში 23000 ავტომობილია. დარეგისტრირებული იყო 200 მანქანის მიერ დღეში გავლილი მანძილები და მიიღეს $\bar{x} = 54.5$ მილი, $s = 14.0$ მილი. ააგეთ 23000 მანქანის მიერ გავლილი მანძილების საშუალოს 90%-იანი ნდობის ინტერვალი.
10. ფირმას, რომელიც უშვებს ახალი ტიპის სახურავებს, უნდა ფილიალის გახსნა ერთ-ერთ ქალაქში. ფილიალის გახსნამდე შეფასებული იყო ქალაქში სახლების პროპორცია ახალი ტიპის სახურავებით. 200 სახლის დათვალიერების შედეგად მიიღეს, რომ ახალი სახურავებით სახლების პროპორცია $\hat{p} = 0.36$.
 - ა) ააგეთ 95%-იანი ნდობის ინტერვალი სახლების რეალური პროპორციისათვის.
 - ბ) მიღებულ ინფორმაციაზე დაყრდნობით გახსნიდით თუ არა ფილიალს ამ ქალაქში?
 - გ) დამატებით რამდენი სახლი უნდა იყოს დათვალიერებული, რომ მიიღოთ შეფასება 95%-იანი სანდოობით და 4%-იანი სიზუსტით?
11. სამშენებლო ფირმამ ააშენა 50 სახლი. სახლის აშენებას საშუალოდ დასჭირდა 35 დღე, სტანდარტული გადახრით 7 დღე. გამოკითხეს ამ სახლების 49 მკვატრონე.

- ერთ-ერთი კითხვა იყო – რამდენ დღეში აშენდა თქვენი სახლი? და მიიღეს $\bar{x} = 46$. რა კომენტარს გააკეთებთ ამ პასუხებზე?
12. სტატისტიკის დეპარტამენტის მონაცემებით, დედაქალაქში სასტუმროს ერთი ნომრის საშუალო ღირებულება შეადგენს 69.2 ლარს. ამ ჰიპოთეზის შესამოწმებლად შემთხვევით შეარჩიეს 36 ნომერი და დაადგინეს, რომ მათი საშუალო ღირებულება შეადგენს 68.4 ლარს, სტანდარტული გადახრილობა 3.7 ლარი. $\alpha = 0.05$ მნიშვნელოვნების დონით გვაქვს თუ არა საფუძველი უარყოთ ჰიპოთეზა?
13. ინდუსტრიულ რაიონებში ხშირად მოდის მჟავა წვიმა. წვიმის დროს ქალაქის 81 ადგილას აიღეს წყლის სინჯი და გაზომეს pH . აღმოჩნდა, რომ: საშუალო – $\bar{x} = 5.538$, სტანდარტული გადახრა – $s = 0.488$. pH ზომავს მჟავიანობას, სადაც 0 = მჟავიანობა, 14 = ტუტეიანობა, სუფთა წვიმის $pH = 5.7$. შეამოწმეთ ჰიპოთეზა $H_0: \mu = 5.7$ ალტერნატივის საწინააღმდეგოდ $H_A: \mu < 5.7$. გამოიყენეთ 0.05 მნიშვნელოვნების დონე.
14. 2003 წლის სექტემბერში, ერთი კვირის განმავლობაში, ერთი ქალაქის 215 სკოლის მონაცემებით, დღეში საშუალოდ დაფიქსირდა 118 გაცდენა, სტანდარტული გადახრა იყო $s = 79$.
- ა) ააგეთ 90%-იანი ნდობის ინტერვალის ქალაქის სკოლების ერთი დღის გაცდენებისათვის.
- ბ) შეამოწმეთ ჰიპოთეზა $\mu = 125$ გაცდენა დღეში, მნიშვნელოვნების დონით $\alpha = 0.01$.
- გ) რამდენი სკოლა უნდა შეამოწმოს, რომ საშუალოს შეფასებისას უზრუნველყოთ 90%-იანი საიმედოობა და 0.04 სიზუსტე?
15. 64 ღორს გაუკეთეს ანტიბიოტიკის ინექცია გათვლილი 10 მგ/კგ წონაზე. 90 წთ-ის შემდეგ გაზომილი იყო სისხლში ანტიბიოტიკის კონცენტრაცია. საშუალო აღმოჩნდა 24.86 მგ/მლ, ხოლო სტანდარტული გადახრა კი 3.92 მგ/მლ.
- ა) ააგეთ 95%-იანი ნდობის ინტერვალის პოპულაციის საშუალო კონცენტრაციისათვის.
- ბ) რამდენი ღორი უნდა იყოს შერჩეული, რომ უზრუნველყოთ 95%-იანი საიმედოობა და 0.5 სიზუსტე μ -ს შეფასებისას?
16. გდანსკის მახლობლად ბალტიის ზღვაში აიღეს წყლის 8 სინჯი და გაზომეს სტრონციუმის შემცველობა (მთლიანი შერჩევის წონის %). მიღებული მონაცემებია: 4.0, 5.5, 4.5, 4.25, 6.0, 5.75, 4.5, 4.0.
- ა) იპოვეთ 95%-იანი ნდობის ინტერვალის ბალტიის ზღვაში გდანსკის მახლობლად წყალში სტრონციუმის შემცველობისთვის.
- ბ) ააგეთ მიღებული მონაცემების ბოქსპლოტი. რა ვარაუდი შეგიძლიათ გამოთქვათ თქვენ მიერ აგებულ ნდობის ინტერვალზე?
17. გდანსკის მახლობლად ბალტიის ზღვაში ტყვიის საშუალო შემცველობა შეადგენდა 18%-ს. ჩატარდა ტოქსიკური მძიმე მეტალებისგან გამწმენდ ღონისძიებათა სერია. ამის შემდეგ 9 სინჯის ტყვიის შემცველობის საშუალო გამოვიდა 17.0-ის ტოლი, სტანდარტული გადახრა კი 2.6. $\alpha = 0.05$ მნიშვნელოვნების დონით შეამოწმეთ ჰიპოთეზა. გაუმჯობესდა, თუ არა სიტუაცია.

18. 45-50 წლის მამაკაცებში ტესტოსტერონის დონე შეადგენს 9.4 ნგ/დ. შემთხვევით შერჩეულ ამ ასაკის 16 მამაკაცში ტესტოსტერონის საშუალო დონე აღმოჩნდა 10.2, სტანდარტული გადახრა კი 5.29. ადასტურებს თუ არა ეს ინფორმაცია იმას, რომ ამ ასაკის მამაკაცებში ტესტოსტერონის დონეა 9.4? გამოიყენეთ 5%-იანი მნიშვნელოვნების დონე.
19. უნივერსიტეტის 50 შემთხვევით შერჩეული სტუდენტიდან 14 რეგულარულად ვარჯიშობს. ააგეთ 95%-იანი ნდობის ინტერვალი ამ უნივერსიტეტის ყველა სტუდენტის პროპორციისთვის, ვინც რეგულარულად ვარჯიშობს.
20. 2003 წელს რეგისტრირებული 69224 ახალშობილიდან 33937 გოგონა იყო.
- რას უდრის გოგონების დაბადების ალბათობის ნერტილოვანი შეფასება?
 - ააგეთ ამ ალბათობის 95%-იანი ნდობის ინტერვალი.
 - რა მოცულობის შერჩევის აღება არის საჭირო, რომ უზრუნველყოთ შეფასების 95%-იანი საიმედოობა და 0.002 სიზუსტე?
21. ერთ ზოოპარკშიმყოფი 50 ოპუსუმიდან 23-ს აღმოაჩნდა ტუბერკულოზი.
- შესაძლებელია თუ არა ქვეყნის ზოოპარკებში ტუბერკულოზით დაავადებული ყველა ოპუსუმის პროპორციისთვის ნდობის ინტერვალის ასაგებად ცოტ-ს გამოყენება?
 - ააგეთ 99%-იანი ნდობის ინტერვალი დაავადებული ოპუსუმების პროპორციისათვის.
 - რას უდრის შეფასების ცდომილება თქვენ მიერ ბ) პუნქტში აგებულ ნდობის ინტერვალში? როგორ შეიძლება ამ ცდომილების შემცირება?
22. ცნობილია, რომ 530 ხავსის სახეობიდან 3 გადაშენების პირასაა. რას უდრის გადაშენების პირას მისული ხავსის სახეობის აღმოჩენის ალბათობის ნერტილოვანი შეფასება? შესაძლებელია თუ არა 95%-იანი ნდობის ინტერვალის აგება ამ უცნობი ალბათობისათვის? თუ ჩათვლით, რომ შესაძლებელია, ააგეთ ეს ინტერვალი. თუ ჩათვლით, რომ არ არის შესაძლებელი, ახსენით რატომ.
23. 627 რესპონდენტს სთხოვეს აღენიშნათ მათი პოლიტიკური შეხედულებანი შეიძლება სკალაზე: 1 – ექსტრემალური ლიბერალი, 2 – ლიბერალი, 3 – ოდნავ ლიბერალი, 4 – ზომიერი, 5 – ოდნავ კონსერვატორი, 6 – კონსერვატორი, 7 – ექსტრემალური კონსერვატორი. პოლიტიკური კუთვნილება იზომება რიგის სკალაზე, კატეგორიების გადანომვრით ცვლადი განიხილება როგორც რაოდენობრივი და მკვლევარს ეძლევა საშუალება გამოთვალოს საშუალო, სტანდარტული გადახრა და დაინახოს სკალის რომელ მხარეს (ლიბერალურ, თუ კონსერვატულ) არის გადახრილი რესპონდენტების შეხედულებანი. თუ საშუალო 4-ზე ნაკლებია, სახეზეა მიდრეკილება ლიბერალიზმისაკენ, თუ საშუალო მეტია 4-ზე – კონსერვატიზმისკენ.

ქვემოთ მოყვანილია ამონაბეჭდი კომპიუტერის პროგრამიდან:

N	Mean	Std Dev	Std Err
627	4.032	1.257	0.050

- ა) ააგეთ საშუალო პოლიტიკური კუთვნილების 95%-იანი და 99%-იანი ნდობის ინტერვალები და შეადარეთ ერთმანეთს;

- ბ) შეამოწმეთ ჰიპოთეზა: არსებობს თუ არა ტენდენცია $H_0: \mu = 4$ ლიბერალური ან კონსერვატიული მიმართულებით $H_A: \mu \neq 4$ ალტერნატივის სანინალმდეგოდ, $\alpha = 0.05$.
- გ) მოიცავენ თუ არა თქვენ მიერ ა) პუნქტში აგებული ნდობის ინტერვალები $\mu = 4$ ჰიპოთეტურ მნიშვნელობას?
24. უნდა გამოეკვლიათ ანორექსიით დაავადებული ახალგაზრდა გოგონები. ფიქსირდებოდა თითოეული გოგონას წონა მკურნალობამდე და მკურნალობის შემდეგ. წონათა სხვაობა უარყოფითია, თუ გოგონა იკლებდა წონაში და დადებითია, თუ იმატებდა წონაში. μ არის პოპულაციის წონების საშუალო მატება. მკურნალობა (კოგნიტიური ქცევის მეთოდი) ითვლება წარმატებულად, თუ $\mu = 0$. SPSS პროგრამიდან ამონაბეჭდი:

Number of Cases	Mean	SD	SE	of Mean
Variable CHANGE	39	3.01	7.31	1.357

- ა) შეამოწმეთ $H_0: \mu = 0$ ჰიპოთეზა $H_A: \mu > 0$ ალტერნატივის სანინალმდეგოდ.
- ბ) შეამოწმეთ $H_0: \mu \leq 0$ ჰიპოთეზა $H_A: \mu > 0$ ალტერნატივის სანინალმდეგოდ.
25. ამერიკელ რესპონდენტებს დაუსვეს კითხვა: აიღებს თუ არა მთავრობა თავის თავზე პასუხისმგებლობას შეამციროს შემოსავალთა სხვაობა მდიდრებსა და ღარიბებს შორის? 1227 ადამიანიდან 591 დადებითად უპასუხა. ვთქვათ, p აღნიშნავს ამერიკელების პროპორციას, ვინც ამ კითხვაზე დადებითად უპასუხებს. თუ $p < 0.5$, ეს არის პოპულაციის უმცირესობა, თუ $p > 0.5$ – პოპულაციის უმრავლესობა. შეამოწმეთ $H_0: p = 0.5$ ჰიპოთეზა $H_A: p \neq 0.5$ ალტერნატივის სანინალმდეგოდ, $\alpha = 0.05$.
26. საავადმყოფოს ადმინისტრატორს დაავალეს შეეფასებინა, საშუალოდ რამდენ დღეს რჩება პაციენტი საავადმყოფოში. 100 პაციენტი საავადმყოფოში საშუალოდ რჩება 5.3 დღის განმავლობაში. 95%-იანი ნდობის ინტერვალის ცდომილება გამოვიდა 1.0-ის ტოლი. ადმინისტრატორმა ჩათვალა, რომ ეს ცდომილება საკმაოდ დიდია. შერჩევის რა მოცულობის ალება საჭირო იმისათვის, რომ ცდომილება განახევრდეს?
27. 1964 რესპონდენტს დაუსვეს კითხვა: დღეში საშუალოდ რამდენ საათს უყურებდა ტელევიზორს. საშუალო დრო გამოვიდა 2.8-ის ტოლი, საშუალოს სტანდარტული შეცდომა კი – 0.05.
- ა) ააგეთ 98%-იანი ნდობის ინტერვალი ტელევიზორის ყურების საშუალო დროსათვის.
- ბ) როგორ ფიქრობთ, არის თუ არა განაწილება ზარისებური?
28. 1958 რესპონდენტს დაუსვეს კითხვა: გჯერათ, რომ არსებობს ცხოვრება სიკვდილის შემდეგ? 1425 რესპონდენტმა დადებითად უპასუხა. ანგარიში ეწერა, რომ ამერიკელების 73%-ს სჯერა, რომ არსებობს ცხოვრება სიკვდილის შემდეგ, ცდომილება შეადგენს 2% -ს. ახსენით, როგორ იყო მიღებული ეს შედეგი.

29. არჩევნებამდე გამოკითხული იყო 1400 შემთხვევით შერჩეული ამომრჩეველი. 742 დემოკრატ კანდიდატს აძლევდა ხმას, 658 – რესპუბლიკელ კანდიდატს.
- საკმარისია თუ არა ეს ინფორმაცია არჩევნების შედეგის პროგნოზისათვის? (ააგეთ 99%-იანი ნდობის ინტერვალი);
 - ააგეთ 99%-იანი ნდობის ინტერვალი და ახსენით, რატომ არის საჭირო მეტი სანდოობის მოთხოვნა პროგნოზის გასაკეთებლად.
30. მკვლევარს სურდა დაედგინა იმ სტუდენტების პროპორცია, ვინც სწავლობს იმავე შტატში, სადაც ცხოვრობენ მათი მშობლები. შერჩევის რა მოცულობა უნდა იყოს აღებული, რომ მიიღოს შეფასება:
- 0.1 სიზუსტით და ალბათობით (საიმედოობით) 0.95.
 - 0.05 სიზუსტით და ალბათობით 0.95.
 - 0.05 სიზუსტით და ალბათობით 0.99.
 - 0.01 სიზუსტით და ალბათობით 0.99.
31. ცხრილში მოცემულია SPSS-ის ამონაბეჭდი. მონაცემები – გულის ცემის ცვლილება. როგორ უნდა შეამოწმოთ ჰიპოთეზა: პოპულაციის გულის ცემის ცვლილება ნულის ტოლია.

Variable	Number of Cases		Mean	SD	SE of Mean
Change	15		10.70	17.70	4.570
Mean	95% CI		t - value	df	2-Tail sig
	Lower	Upper			
10.70	.898	20.501	2.341	14	.0346

32. ნაფიც მსაჯულთა სიაში 53% ქალია. თუ სიიდან შემთხვევით იქნება ამორჩეული 12 მსაჯული, იპოვეთ ალბათობა იმისა, რომ მათ შორის:
- არცერთი ქალი არ იქნება.
 - ზუსტად ერთი ქალი იქნება.
 - იპოვეთ ქალების მოსალოდნელი რაოდენობა და სტანდარტული გადახრა.
 - შეამოწმეთ ჰიპოთეზა, რომ შერჩევა შემთხვევითია, თუ 12 ამორჩეულში არცერთი ქალი არ არის.
33. თვითმკვლელობის ალბათობა ამერიკელ კაცებს შორის უდრის 0.02-ს.
- იპოვეთ 100 გარდაცვლილ მამაკაცში თვითმკვლელობის მოსალოდნელი სიდიდე და სტანდარტული გადახრა.
 - გაგიკვირდებათ თუ არა, რომ მათ შორის არცერთი თვითმკვლელი არ იქნება. ახსენით რატომ.
34. საფეხბურთო გუნდის თამაშის მოგების ალბათობა უდრის 0.5-ს.
- იპოვეთ ალბათობა იმისა, რომ ექვსი თამაშიდან გუნდი ყველა თამაშს მოიგებს, ყველა თამაშს წააგებს.
 - თუ გუნდი ყველა თამაშს მოიგებს, სარწმუნოა, რომ ამ გუნდის მოგების ალბათობა 0.5-ის ტოლია?

ნინა თავებში ჩვენ ვაგებდით ნდობის ინტერვალებს და ვამონმებდით პიპოთეზებს, როცა მონაცემები აღებული იყო ერთი ფაქტორისათვის. ახლა კი შევადარებთ ერთმანეთს ორ ფაქტორს, მათი საშუალოების დახმარებით.

საშუალოთა შორის განსხვავების პიპოთეზის შემონმება დიდი მოცულობის მქონე შერჩევებისას

μ_1 და μ_2 საშუალოები ტოლია, თუ მათი $\mu_1 - \mu_2$ სხვაობა ნულის ტოლია. თუ პირველი შერჩევის საშუალოა $\bar{X}_1$, ხოლო მეორე შერჩევის საშუალოა $\bar{X}_2$, მაშინ $\mu_1 - \mu_2$ -ის წერტილოვან შეფასებად უნდა ჩავთვალოთ $\bar{X}_1 - \bar{X}_2$ სხვაობა.

ვთქვათ, ორივე შერჩევის მოცულობა დიდია: ეს იმას ნიშნავს, რომ $n_1 \geq 30$ და $n_2 \geq 30$ (ან ორივე პოპულაცია ნორმალურადაა განაწილებული და ცნობილია საშუალო კვადრატული გადახრები σ_1 და σ_2). მაშინ, ცენტრალური ზღვართი თეორემის თანახმად, $\bar{X}_1 - \bar{X}_2$ სხვაობა მიახლოებით ნორმალურადაა განაწილებული:

$$\bar{X}_1 - \bar{X}_2 \equiv N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

და, შესაბამისად:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \equiv N(0, 1).$$

ამიტომ პოპულაციის საშუალოთა შორის სხვაობის $(1 - \alpha) \cdot 100\%$ -იანი ნდობის ინტერვალი მოიცემა ფორმულით:

$$\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

სადაც $z_{\alpha/2}$ - სტანდარტული ნორმალური განაწილების ზედა $\alpha/2$ -კრიტიკული წერტილია, რომელსაც ვპოულობთ სტანდარტული ნორმალური განაწილების ცხრილებიდან ($\Phi(z_{\alpha/2}) = 1 - \alpha/2$ ან $\Phi_0(z_{\alpha/2}) = 0.5 - \alpha/2$). ამ ინტერვალის აგებისას,

ხშირ შემთხვევაში შესაფასებელი გვექნება სტანდარტული შეცდომები: როცა $n_1 \geq 30$ და $n_2 \geq 30$, მაშინ $\sigma_1 \approx s_1$ და $\sigma_2 \approx s_2$. 95%-იანი ნდობის ინტერვალი $\mu_1 - \mu_2$ -თვის არის:

$$\bar{x}_1 - \bar{x}_2 \pm 1.96 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

თუ ეს ინტერვალი მოიცავს ნულს, შეგვიძლია დავასკვნათ, რომ პოპულაციის საშუალოები არ განსხვავდებიან ერთმანეთისგან 0.05 მნიშვნელოვნების დონით.

მაგალითი 10.1. არსებობს აზრი იმის შესახებ, რომ მამაკაცები უფრო მკაცრად ისჯებიან ერთნაირი სახის ნაკლებად მნიშვნელოვანი დანაშაულისათვის, ვიდრე ქალები. შეგროვილი იყო მასალა 40 ქალის და 49 კაცის პატიმრობის ვადებზე (დღეები).

	შერჩევის საშუალო	შერჩევის სტანდარტული შეცდომა	შერჩევის მოცულობა
ქალები	16.3	14.6	40
კაცები	29.5	17.2	49

შეაფასეთ პოპულაციების საშუალოთა შორის სხვაობა.

ამოხსნა: μ_1 იყოს კაცების პატიმრობის საშუალო ვადა. μ_2 კი ქალების პატიმრობის საშუალო ვადა. 95%-იანი ინტერვალი $\mu_1 - \mu_2$ -თვის არის:

$$29.5 - 16.3 \pm \sqrt{\frac{17.2^2}{49} + \frac{14.6^2}{40}} = 13.2 \pm 6.61 = (6.59, 19.8).$$

კაცების პატიმრობის საშუალო ვადა მოსალოდნელია იყოს 6.6-დან 19.8-მდე დღით მეტი, ვიდრე ქალების პატიმრობის საშუალო ვადა.

ჰიპოთეზათა შემოწმება $\mu_1 - \mu_2$ -ის შესახებ

ნულოვანი ჰიპოთეზა: $H_0: \mu_1 = \mu_2$ ერთ-ერთი ალტერნატივის წინააღმდეგ:

$$H_A: \mu_1 \neq \mu_2 \text{ ან } H_A: \mu_1 < \mu_2 \text{ ან } H_A: \mu_1 > \mu_2.$$

თუ ორი დამოუკიდებელი პოპულაცია ნორმალურია ან ორივე შერჩევის მოცულობა მეტია ან ტოლი 30-ზე, მაშინ კრიტერიუმის სტატისტიკა:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \stackrel{w}{\sim} N(0, 1).$$

რადგან კრიტერიუმის სტატისტიკას აქვს მიახლოებით ნორმალური განაწილება, უარყოფის არის აგება ხდება სტანდარტული ნორმალური განაწილების ცხრილის დახმარებით ალტერნატიული ჰიპოთეზის შესაბამისად.

მაგალითი 10.2. გამოვიყენოთ მაგალითი 10.1-ის მონაცემები და შევამოწმოთ ჰიპოთეზა, რომ კაცებს პატიმრობის მეტი ვადა აქვთ მისჯილი, ვიდრე ქალებს ერთი და იმავე სახის უმნიშვნელო დანაშაულისათვის.

გვინდა შევამოწმოთ $H_0: \mu_1 = \mu_2$ ჰიპოთეზა $H_1: \mu_1 > \mu_2$ ალტერნატივის წინააღმდეგ, $\alpha = 0.05$.

$$z = \frac{29.5 - 16.3}{\sqrt{\frac{14.6^2}{40} + \frac{17.2^2}{49}}} = \frac{13.2}{2.371} = 3.92.$$

სტანდარტული ნორმალური განაწილების ცხრილებიდან, როცა $\alpha = 0.05$, ვპოულობთ, რომ $z_{0.05} = 1.65$. კრიტიკული არეა ინტერვალი $-[1.65, +\infty)$. H_0 ჰიპოთეზა უნდა უარეყოთ, რადგან $3.92 > 1.65$ -ზე. დასკვნა: 0.05 მნიშვნელოვნების დონით კაცების პატიმრობის ვადა მეტია ქალების პატიმრობის ვადაზე.

ორი დამოუკიდებელი პოპულაციის საშუალოთა შორის განსხვავების ჰიპოთეზათა შემოწმება

(ორივე ან ერთ-ერთი შერჩევის მოცულობა ნაკლებია 30-ზე)

როცა ორივე ან ერთ-ერთი შერჩევის მოცულობა < 30 , შეუძლებელია ვივარაუდოთ, რომ საშუალოთა სხვაობას ექნება ნორმალური განაწილება. თუ ორივე შერჩევის მონაცემები დამოუკიდებელი ნორმალური პოპულაციებიდან არის აღებული და პოპულაციების უცნობი დისპერსიები ტოლია, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, მაშინ T სტატისტიკას აქვს t -განაწილება $n_1 + n_2 - 2$ თავისუფლების ხარისხით:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \equiv t(n_1 + n_2 - 2),$$

სადაც

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

წარმოადგენს საერთო უცნობი σ^2 დისპერსიის შეფასებას ორივე შერჩევის საფუძველზე. ეს სტატისტიკა შეგვიძლია გამოვიყენოთ ორი დამოუკიდებელი პოპულაციის საშუალოთა შორის განსხვავების ჰიპოთეზათა შემოწმებისთვის. თავისუფლების ხარისხი უდრის $n_1 + n_2 - 2$ -ს, რადგან შესაფასებელია ორი შერჩევის საშუალო ($n_1 + n_2 - 2 = n_1 - 1 + n_2 - 1$).

$\mu_1 - \mu_2$ -ის ნდობის ინტერვალის მცირე მოცულობის შერჩევებისათვის

საშუალოთა $\mu_1 - \mu_2$ სხვაობის $(1-\alpha) \cdot 100\%$ -იანი ნდობის ინტერვალის მოიცემა ფორმულით:

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2, \alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

სადაც $t_{n_1+n_2-2, \alpha/2}$ თავისუფლების $n_1 + n_2 - 2$ ხარისხის მქონე სტიუდენტის განაწილების ზედა $\alpha/2$ -კრიტიკული წერტილია.

მაგალითი 10.3. ტყის ჩრდილოეთ ნაწილში, შვიდ უბანზე, აღებული ხავსის საშუალო სისქეა $\bar{x}_1 = 0.84$ სმ, სტანდარტული გადახრა კი $s_1 = 0.41$. სამხრეთ ნაწილში, ათ უბანზე, აღებული ხავსის საშუალო სისქეა $\bar{x}_2 = 0.73$, სტანდარტული გადახრა – $s_2 = 0.23$. შემოწმებამ აჩვენა, რომ ორივე პოპულაცია ნორმალურადაა განაწილებული და პოპულაციების დისპერსიები ტოლია. 95%-იანი სანდობით შეაფასეთ ჩრდილოეთ და სამხრეთ ნაწილებში ხავსის სისქის სხვაობა.

ამოხსნა: შევაფასოთ საერთო დისპერსია:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{6 \cdot 0.23^2 + 9 \cdot 0.41^2}{7 + 10 - 2} = 0.122,$$

$$s_p = \sqrt{0.122} = 0.349.$$

საქმე გვაქვს $t(n_1 + n_2 - 2)$ განაწილებასთან $n_1 + n_2 - 2 = 15$ თავისუფლების ხარისხით. სტიუდენტის განაწილების ზედა კრიტიკული წერტილების ცხრილიდან ვპოულობთ, რომ $t_{15, \alpha/2} = t_{15, 0.025} = 2.131$. შესაბამისად, $\mu_1 - \mu_2$ -ის 95%-იანი ნდობის ინტერვალია:

$$0.84 - 0.73 \pm 2.131 \cdot 0.349 \cdot \sqrt{\frac{1}{10} + \frac{1}{7}} = 0.11 \pm 0.37 = (-0.48, 0.26).$$

ეს ინტერვალის მოიცავს ნულს, ამიტომ ამ შერჩევებზე დაყრდნობით ვერ ვიტყვი, რომ ხავსის სისქე ტყის სხვადასხვა მხარეს მნიშვნელოვნად განსხვავდება ერთმანეთისაგან.

მაგალითი 10.4. ჩავთვალოთ, რომ პოპულაციების დისპერსიები მნიშვნელოვნად არ განსხვავდებიან ერთმანეთისგან და შევამოწმოთ ჰიპოთეზა, რომ ხავსის პოპულაციების სისქეები არ განსხვავდებიან ერთმანეთისაგან. გამოვიყენოთ $\alpha = 0.05$.

$$H_0: \mu_1 - \mu_2 = 0,$$

$$H_A: \mu_1 - \mu_2 > 0.$$

კრიტერიუმის სტატისტიკის მნიშვნელობაა:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.84 - 0.73 - 0}{\sqrt{0.122 \cdot \left(\frac{1}{10} + \frac{1}{7} \right)}} = \frac{0.11}{0.172} = 0.64.$$

ამ სტატისტიკას აქვს t -განაწილება $n_1 + n_2 - 2$ თავისუფლების ხარისხით. კრიტიკული მნიშვნელობა იქნება: $t_{0.05} = 1.753$. შესაბამისად, უარყოფის არეა ინტერვალი $- [1.753, +\infty)$. კრიტერიუმის სტატისტიკის მნიშვნელობა 0.64 ნაკლებია კრიტიკულ მნიშვნელობაზე 1.753 და ამიტომ 5% -იანი მნიშვნელოვნების დონით ვერ უარყოფთ ნულოვან ჰიპოთეზას, ვერ ვიტყვით, რომ ხავსების სისქეები მნიშვნელოვნად განსხვავდება.

ტექნიკური შენიშვნა. თუ პოპულაციის დისპერსიები მნიშვნელოვნად განსხვავდებიან ერთმანეთისაგან და შეუძლებელია ვივარაუდოთ, რომ $\sigma_1^2 = \sigma_2^2 = \sigma^2$, უნდა გამოვიყენოთ შემდეგი სტატისტიკა:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

T სტატისტიკას აქვს t -განაწილება. დასადგენია მისი თავისუფლების ხარისხი. არსებობს ორი მიდგომა: ე.წ. გამარტივებული პროცედურა (რომელიც შედარებით არაზუსტია) და სატერტვაიტის მეთოდი. გამარტივებული პროცედურის მიხედვით t -განაწილების თავისუფლების ხარისხია $- df = \min(n_1 - 1, n_2 - 1)$, ხოლო სატერტვაიტის მეთოდის მიხედვით კი t -განაწილების თავისუფლების ხარისხი გამოითვლება როგორც შემდეგი c რიცხვის მთელი ნაწილი ($df = [c]$):

$$c = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{(s_1^2 / n_1)^2 / (n_1 - 1) + (s_2^2 / n_2)^2 / (n_2 - 1)}.$$

ორივე შემთხვევაში გადაწყვეტილების მიღების წესი ასე ყალიბდება: თუ მოცემული α მნიშვნელოვნების დონისათვის $-t_{\alpha/2, df} \leq t \leq t_{\alpha/2, df}$ (სადაც t კრიტერიუმის სტატისტიკის მნიშვნელობაა), მაშინ α მნიშვნელოვნების დონით $H_0: \mu_1 - \mu_2 = 0$ ჰიპოთეზის უარყოფის საფუძველი არა გვაქვს. წინააღმდეგ შემთხვევაში მას უარყოფთ.

საშუალოთა $\mu_1 - \mu_2$ სხვაობის $(1 - \alpha) \cdot 100\%$ -იანი ნდობის ინტერვალი მოიცემა ფორმულით:

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2, df} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

ყველა ზემოთ განხილული კრიტერიუმით მიწმდებოდა საშუალოთა შორის განსხვავების არსებობა ორი დამოუკიდებელი ჯგუფისთვის. ამ სტატისტიკას ეწოდება t -დამოუკიდებელი და სტატისტიკურ მეთოდს – ცდისპირთა შორის სქემა.

ჰიპოთეზათა შემოწმება ორი დამოუკიდებელი ნორმალური პოპულაციის დისპერსიების შესახებ

საშუალოთა შორის განსხვავების ჰიპოთეზათა შემოწმების დროს დაშვებული იყო ვარაუდი, რომ ნორმალური პოპულაციების დისპერსიები ტოლია. ამ ვარაუდის შემოწმება ხდება ფიშერის F -განაწილების დახმარებით. ფიშერის განაწილებას ახასიათებს ორი პარამეტრი, მრიცხველის თავისუფლების ხარისხი და მნიშვნელის თავისუფლების ხარისხი. $\frac{s_1^2}{s_2^2}$ შეფარდებას აქვს F -განაწილება $n_1 - 1$ და $n_2 - 1$ თავისუფლების ხარისხებით.

შევამოწმოთ $H_0: \sigma_1^2 = \sigma_2^2$ ჰიპოთეზა $H_0: \sigma_1^2 \neq \sigma_2^2$ ალტერნატივის წინააღმდეგ.

თუ H_0 სამართლიანია, მაშინ s_1^2 და s_2^2 მიახლოებით ერთმანეთის ტოლია და შეფარდება $\frac{s_1^2}{s_2^2} \approx 1$ მიახლოებით ერთის ტოლია.

თუ პოპულაციები ნორმალურია და $s_1^2 > s_2^2$, მაშინ კრიტერიუმის სტატისტიკას

$$F = \frac{s_1^2}{s_2^2}$$

აქვს F -განაწილება $n_1 - 1$ და $n_2 - 1$ თავისუფლების ხარისხებით.

თუ პოპულაციები ნორმალურია და $s_1^2 < s_2^2$, მაშინ კრიტერიუმის სტატისტიკად განიხილება

$F = \frac{s_2^2}{s_1^2}$, რომელსაც აქვს F -განაწილება $n_2 - 1$ და $n_1 - 1$ თავისუფლების ხარისხებით.

უარყოფის არე

დაშვებული იყო, რომ $s_1^2 > s_2^2$, ანუ უარყოფის არეში გასათვალისწინებელია განაწილების მარჯვენა ბოლო, მაგრამ, რადგან კრიტერიუმი ორმხრივია, უნდა გამოვიყენოთ $\frac{\alpha}{2}$ ალბათობა. კრიტიკული $F_{crit} = F_{n_1-1, n_2-1, \alpha/2}$ სიდიდე უნდა ავიღოთ ფიშერის განაწილების ზედა კრიტიკული წერტილების $\alpha/2$ ალბათობის შესაბამისი ცხრილიდან და უარყოფით ნულოვანი ჰიპოთეზა α მნიშვნელოვნების დონით, თუ $F \geq F_{crit}$.

დისპერსიათა σ_1^2 / σ_2^2 შეფარდების $(1-\alpha) \cdot 100\%$ -იანი ნდობის ინტერვალის მოიცემა ფორმულით:

$$\left(\frac{1}{F_{n_2-1, n_1-1, \alpha/2}} \cdot \frac{s_1^2}{s_2^2}, F_{n_1-1, n_2-1, \alpha/2} \cdot \frac{s_1^2}{s_2^2} \right).$$

მაგალითი 10.5. მაგალითებში 10.3 და 10.4 დაშვებული იყო, რომ პოპულაციების დისპერსიები მნიშვნელოვნად არ განსხვავდებიან ერთმანეთისგან. შევამოწმოთ ეს ჰიპოთეზა 5% დონეზე.

მონმდება $H_0: \sigma_1^2 = \sigma_2^2$ ჰიპოთეზა $H_0: \sigma_1^2 \neq \sigma_2^2$ ალტერნატივის წინააღმდეგ. რადგან $s_2^2 > s_1^2$, ამიტომ კრიტერიუმის სტატისტიკა:

$$F = \frac{s_2^2}{s_1^2}$$

და მას აქვს F -განაწილება $n_2 - 1$ და $n_1 - 1$ თავისუფლების ხარისხებით. კრიტერიუმის მნიშვნელობა:

$$F = \frac{s_2^2}{s_1^2} = \frac{0.41^2}{0.23^2} = 3.178,$$

ხოლო კრიტიკული მნიშვნელობა ფიშერის განაწილების ზედა კრიტიკული წერტილების 0.025 ალბათობის შესაბამისი ცხრილიდან იქნება $F_{crit} = F_{10-1, 7-1, 0.025} = F_{9, 6, 0.025} = 5.5234$.

რადგან კრიტერიუმის სტატისტიკის მნიშვნელობა $3.178 < 5.5234$, ამ შემთხვევაში ნულოვან ჰიპოთეზას ვერ უარყოფთ. დისპერსიების ტოლობით ჩვენ უკვე ვისარგებლეთ 10.4 მაგალითში.

ჰიპოთეზათა შემოწმება დანყვილებულ მონაცემთა

საშუალოების სხვაობებისათვის

ჩვენ ვამოწმებდით ჰიპოთეზებს ორი პოპულაციის საშუალოების სხვაობისათვის ორი დამოუკიდებელი შერჩევის საფუძველზე: μ_1 მოცულობის შერჩევა აღებული იყო ერთი პოპულაციიდან, μ_2 მოცულობის შერჩევა სრულიად დამოუკიდებლად მეორე პოპულაციიდან, რომელიც, იგულისხმებოდა, რომ დამოუკიდებელი იყო პირველი პოპულაციიდან.

ამის საპირისპირო სიტუაციაა, როდესაც საქმე გვაქვს ე.წ. დანყვილებულ მონაცემებთან. მაგალითად, თუ გვაინტერესებს პიროვნების ქცევის შეცვლა ალკოჰოლის მიღების შემდეგ, ან წონის შეცვლა დიეტის შედეგად. რადგან გვაინტერესებს ცვალებადობის აღმოჩენა, უნდა გამოვთვალოთ სხვაობა თითოეული წყვილისათვის. ვთქვათ, გვაქვს n წყვილი.

განვიხილოთ სხვაობები $d_i = x_i - y_i$, სადაც x_i პირველი შერჩევის i -ური მონაცემია; y_i მეორე შერჩევის i -ური მონაცემია. სულ გვექნება n სხვაობა.

თუ $n \geq 30$ ან სხვაობები ნორმალურადაა განაწილებული, კრიტერიუმის სტატისტიკა შემდეგი სახისაა:

$$T = \frac{\bar{X}_d - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

T სტატისტიკას აქვს t -განაწილება $n-1$ თავისუფლების ხარისხით. $\bar{X}_d$ და s_d არის, შესაბამისად, d_i სხვაობების საშუალო და სტანდარტული გადახრები. $\mu_d = \mu_1 - \mu_2$. ამ ტიპის T სტატისტიკას t -დამოკიდებული ეწოდება, განხილულ მეთოდს კი – ცდის პირთა შიდა სქემა.

მაგალითი 10.6. არჩეული იყო 2 ჰექტრის ფართის მქონე ათი ნაკვეთი. ყოველი ნაკვეთი გაყოფილი იყო ორ ნაწილად, ერთზე გამოიყენეს ძველი ტიპის სასუქი, მეორეზე კი – ახალი ტიპის. გაზომილი იყო ხახვის მოსავლიანობა ამ ნაკვეთებზე:

მოსავლიანობა										
ძველი ტიპის სასუქი	1.2	1.4	1.6	1.1	1.4	1.1	0.7	1.5	1.3	1.0
ახალი ტიპის სასუქი	1.4	1.2	1.5	1.1	1.6	1.2	1.0	1.4	1.3	1.5
სხვაობა	0.2	-0.2	-0.1	0.0	0.2	0.1	0.3	-0.1	0.0	0.5

ზრდის თუ არა ახალი ტიპის სასუქი მოსავლიანობას?

უნდა შევამოწმოთ $H_0: \mu_d = 0$ ჰიპოთეზა $H_A: \mu_d > 0$ ალტერნატივის საწინააღმდეგოდ. მნიშვნელოვნების დონედ ავიღოთ $\alpha = 0.05$.

დავუშვათ, რომ სხვაობები ნორმალურადაა განაწილებული. ეს მონაცემები დანყვილებულია, რადგან მონაცემთა ყოველი წყვილი ეკუთვნის ერთსა და იმავე ნაკვეთს.

ადვილად გამოვითვლით, რომ: $\bar{X}_d = 0.9$; $s_d = 0.738$; კრიტერიუმის სტატისტიკის მნიშვნელობა $t = \frac{0.9 - 0}{\frac{0.738}{\sqrt{10}}} = 3.86$; თავისუფლების ხარისხია $10 - 1 = 9$; $t_{9,0.05} = 1.833$;

კრიტიკული არეა $-[1.833, +\infty)$. ვინაიდან, $3.86 > 1.833$, ამიტომ ნულოვანი ჰიპოთეზა უნდა უარყყოთ: ახალი ტიპის სასუქი ზრდის მოსავლიანობას.

ორამოკრეფიანი ამოცანები ორი ალბათობის შედარებისას

ხშირად საჭიროა პოპულაციის პროპორციების (ალბათობების) შედარება. თუ ორივე შერჩევის მოცულობა ≥ 30 -ზე, Z სტატისტიკას, ცხთ-ის თანახმად, სტანდარტული ნორმალური განაწილება აქვს:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{as}{\approx} N(0,1).$$

$(1-\alpha) \cdot 100\%$ -იანი ნდობის ინტერვალი $p_1 - p_2$ -თვის არის:

$$\hat{p}_1 - \hat{p}_2 \mp z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

მაგალითი 10.7. 240 შემთხვევით შერჩეული მშობიარე ქალიდან პირველ სამშობიაროში 96-ს გაუკეთდა ეპიდურალური ანესთეზია. მეორე სამშობიაროში 176-დან 68-ს გაუკეთეს ეპიდურალური ანესთეზია. განსხვავდებიან თუ არა სამშობიაროები ამ პროცენტურის ჩატარების სიხშირით?

ამოხსნა: ვიპოვოთ პროპორციების წერტილოვანი შეფასებები:

$$\hat{p}_1 = \frac{96}{240} = 0.4, \quad \hat{p}_2 = \frac{68}{176} = 0.386.$$

95% -იანი ნდობის ინტერვალი $p_1 - p_2$ -თვის არის:

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 \mp 1.96 \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} &= \\ = 0.4 - 0.386 \mp 1.96 \cdot \sqrt{\frac{0.4 \cdot (1-0.4)}{240} + \frac{0.386 \cdot (1-0.386)}{176}} &= 0.014 \mp 0.095, \end{aligned}$$

ანუ მივიღებთ $(-0.081, 0.109)$ ინტერვალს, რომელიც მოიცავს ნულს. ეს კი ნიშნავს, რომ $p_1 = p_2$.

ჰიპოთეზათა შემოწმება

შესამოწმებელია $H_0: p_1 = p_2$ ერთ-ერთი ალტერნატივის საწინააღმდეგოდ:

$$H_A: p_1 \neq p_2 \quad \text{ან} \quad H_A: p_1 > p_2 \quad \text{ან} \quad H_A: p_1 < p_2.$$

ნულოვანი ჰიპოთეზის თანახმად, პოპულაციის პროპორციები ერთმანეთის ტოლია. p_1 -ის წერტილოვანი შეფასებაა $\hat{p}_1 = \frac{X_1}{n_1}$, სადაც X_1 „წარმატებათა“ რაოდენობაა პირველი შერჩევის n_1 ცდაში. X_1 -ს აქვს ბინომიალური განაწილება პარამეტრებით n_1, p_1 . $X_1 \equiv B(n_1, p_1)$. ანალოგიურად, $X_2 \equiv B(n_2, p_2)$. თუ $p_1 = p_2$, წარმატებათა $X_1 + X_2$ რაოდენობას $n_1 + n_2$ ცდაში აქვს ბინომიალური განაწილება $n_1 + n_2, p$ პარამეტრებით, სადაც $p_1 = p_2 = p$.

საერთო პროპორციის წერტილოვანი შეფასება:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}.$$

თუ $H_0: p_1 = p_2$, მაშინ კრიტერიუმის Z სტატისტიკას აქვს სტანდარტული ნორმალური განაწილება

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{as}{=} N(0,1).$$

მაგალითი 10.8. გამოვიყენოთ 10.7 მაგალითის მონაცემები და შევამოწმოთ $H_0: p_1 = p_2$ ჰიპოთეზა $H_A: p_1 \neq p_2$ ალტერნატივის საწინააღმდეგოდ, $\alpha = 0.05$.

ამოხსნა. კრიტიკული მნიშვნელობებია $\mp z_{\alpha/2} = \mp z_{0.025} = \pm 1.96$. შესაბამისად, კრიტიკული არეა $(-\infty, -1.96] \cup [1.96, +\infty)$. საერთო პროპორციის წერტილოვანი შეფასება:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{96 + 68}{240 + 176} = \frac{164}{416} = 0.394.$$

ამიტომ კრიტერიუმის სტატისტიკის მნიშვნელობა იქნება:

$$\begin{aligned} Z &= \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \\ &= \frac{0.4 - 0.386}{\sqrt{0.394(1-0.394)\left(\frac{1}{240} + \frac{1}{176}\right)}} = \frac{0.014}{0.0485} = 0.29. \end{aligned}$$

რადგან $-1.96 < 0.29 < 1.96$, ამიტომ ვერ უარვყოფთ ნულოვან ჰიპოთეზას. ეს გადაწყვეტილება ეთანხმება ნდობის ინტერვალზე დაყრდნობით მიღებულ გადაწყვეტილებას.

ტესტი პროპორციათა სხვაობისთვის

საჭიროა, რომ ორივე შერჩევისთვის და ორივე პოპულაციისთვის სრულდებოდეს:

$$0 \leq \hat{p}_i \leq 1, \quad np_i \geq 5 \quad \text{და} \quad n(1-p_i) \geq 5, \quad i=1,2.$$

თუ შესამოწმებელია $H_0: p_1 = p_2$ ჰიპოთეზა, მაშინ კრიტერიუმის სტატისტიკა:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{as}{\equiv} N(0, 1),$$

რომელსაც აქვს მიახლოებით სტანდარტული ნორმალური განაწილება, სადაც $\hat{p}$ ორ შერჩევაზე დაყრდნობით აგებული ნერთილოვანი შეფასებაა:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}.$$

პროპორციათა სხვაობის 95%-იანი ნდობის ინტერვალი:

$$\hat{p}_1 - \hat{p}_2 \mp 1.96 \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

(საზოგადოდ, $(1-\alpha) \cdot 100\%$ -იანი ნდობის ინტერვალი მოიცემა ფორმულით:

$$\hat{p}_1 - \hat{p}_2 \mp z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}).$$

კითხვები

- 30 კომპაქტდისკი შემთხვევით შერჩეული იყო რეპის კოლექციიდან, 30 – ჰიპ-პოპის კოლექციიდან. გაზომეს ხანგრძლივობა. შემაჯამებელი სტატისტიკები: რეპი დისკი – $n_1 = 30$, $\bar{x}_1 = 54$ წთ, $s_1 = 7$ წთ. ჰიპ-პოპი დისკი – $n_2 = 30$, $\bar{x}_2 = 48$ წთ, $s_2 = 8$ წთ.
 - შეამოწმეთ ჰიპოთეზა, რომ ყველა რეპიდისკი და ყველა ჰიპ-პოპ დისკი ერთნაირი ხანგრძლივობისაა. გამოიყენეთ 1%-იანი მნიშვნელოვნების დონე.
 - გამოთვალეთ p -სიდიდე.
 - ააგეთ 90%-იანი ნდობის ინტერვალი რეპიდისკის ხანგრძლივობისათვის და 90%-იანი ნდობის ინტერვალი ჰიპ-პოპისათვის. შეადარეთ ეს ინტერვალები.
- შეამოწმეს 48 ბავშვის ქცევის პრობლემები. ამ ბავშვების დედები ხშირად ამჟღავნებდნენ შფოთვის. 53 ბავშვის დედა – არა. შესწავლილ იქნა ამ ბავშვების ქცევის პრობლემები. ბავშვების ქცევის ქულების შემაჯამებელი სტატისტიკები ასეთია:

მშფოთავი დედები	$n_1 = 48$	$\bar{x}_1 = 7.92$	$s_1 = 3.45$
არამშფოთავი დედები	$n_2 = 53$	$\bar{x}_2 = 5.80$	$s_2 = 2.87$

- შეამოწმეთ ჰიპოთეზა: მშფოთავი დედების შვილებს ქცევის მეტი პრობლემა აქვთ. გამოიყენეთ $\alpha = 0.05$.
- გამოთვალეთ ამ ტესტის p -სიდიდე.
- ააგეთ 90%-იანი ნდობის ინტერვალი პოპულაციის ქცევის ქულების სხვაობისათვის, მშფოთავი და არამშფოთავი დედების შვილებისათვის.

- დ) თუ ორივე ჯგუფის მოცულობა ერთი და იგივე იქნება, რამდენი ბავშვი უნდა შევამოწმოთ თითოეულ ჯგუფში იმისათვის, რომ პოპულაციათა საშუალოების შეფასების სიზუსტე იყოს 0.7 და შეფასების სანდოობა 90% ?
3. გაზომეს 3 წლის 180 ბავშვის კონცენტრაციის უნარი (წუთებში). 120 ბავშვი დადიოდა საბავშვო ბაღში, 60 ბავშვს საბავშვო ბაღში არ უვლია. შემაჯამებელი სტატისტიკები ასეთია:

არ დადიან საბავშვო ბაღში	$n_1 = 60$	$\bar{x}_1 = 11.7$	$s_1 = 4.2$
დადიან საბავშვო ბაღში	$n_1 = 120$	$\bar{x}_1 = 12.5$	$s_1 = 4.6$

- ა) შეამოწმეთ $H_0: \mu_1 = \mu_2$ ჰიპოთეზა $H_1: \mu_1 \neq \mu_2$ ალტერნატივის წინააღმდეგ.
- ბ) ააგეთ 95%-იანი ნდობის ინტერვალის კონცენტრაციის უნარის სხვაობისათვის. ემთხვევა თუ არა ა) და ბ) პუნქტებში მიღებული გადაწყვეტილებები.
4. 2003 წელს 28-33 წლის 141 მამაკაცის საშუალო ხელფასი იყო 674 დოლარი, ხოლო სტანდარტული გადახრა – 174 დოლარი. 2002 წელს იმავე ასაკის 120 მამაკაცის საშუალო ხელფასი იყო 635 დოლარი, სტანდარტული გადახრა კი 182 დოლარი. გამოიყენეთ $\alpha = 0.05$ და გაარკვიეთ, იყო თუ არა ხელფასების მნიშვნელოვანი ზრდა.
5. სხვადასხვა მარკის ავტომობილები ერთი და იმავე მანძილის გავლისას ხარჯავენ ბენზინის განსხვავებულ რაოდენობას. ქვემოთ მოცემულია ორი მარკის ავტომობილის მიერ ერთ კილომეტრზე დახარჯული ბენზინის რაოდენობა ლიტრებში (ლ/კმ).

I მარკის ავტომობილი							
11.6	10.7	11.5	10.1	10.5	10.9	11.0	10.6
II მარკის ავტომობილი							
9.6	8.5	9.3	8.6	8.2	8.7	9.4	8.5

I მარკის ავტომობილი						
10.6	10.8	11.1	11.6	11.3	11.6	10.5
II მარკის ავტომობილი						
10.1	9.6	8.5	9.3	10.3	10.2	

- ა) ისარგებლეთ $\alpha = 0.05$ და შეამოწმეთ ჰიპოთეზა – განსხვავდებიან თუ არა ავტომობილები მოხმარებული ბენზინის რაოდენობით?
- ბ) ააგეთ ორი ბოქსპლოტი და შეადარეთ ერთმანეთს.
6. ტესტი, რომლის მეშვეობით ზომავენ ფსიქიკური ფუნქციონირების სიდიდეს, შეავსებინეს ერთი და იმავე ასაკის 24 ადამიანს, აქედან 12 პარკინსონით დაავადებული იყო, 12 კი – ჯანმრთელი.

	პარკინსონით დაავადებული ჯგუფი	საკონტროლო ჯგუფი
საშუალო	28.58	27.42
სტანდარტული გადახრა	2.065	1.782
n	12	12

ისარგებლეთ 10%-იანი მნიშვნელოვნების დონით და შეამოწმეთ ჰიპოთეზა – ახდენს თუ არა დაავადება გავლენას ფსიქიკის ფუნქციონირებაზე.

7. პორტ გადანსკის მახლობლად, ზღვაში, აიღეს წყლის 10 სინჯი, ხოლო 8 სინჯი – კურორტ ბალტიისკის მახლობლად. გაზომეს სტრონციუმის პროცენტული შემადგენლობა.

გდანსკი	$n_1 = 10$	$\bar{x}_1 = 4.81$	$s_1 = 0.81$
ბალტიისკი	$n_1 = 8$	$\bar{x}_1 = 4.58$	$s_1 = 0.87$

- ა) შეამოწმეთ ჰიპოთეზა დისპერსიების ტოლობის შესახებ.
- ბ) შეამოწმეთ ჰიპოთეზა: პორტ გადანსკის და კურორტ ბალტიისკის მახლობლად წყლის სინჯებში სტრონციუმის პროცენტული შემადგენლობა მნიშვნელოვნად არ განსხვავდება ერმანეთისაგან.
8. შევამოწმოთ ჰიპოთეზა: ქირურგიული ოპერაციის შემდეგ მწველებს უფრო დიდი დრო სჭირდებათ ანესთეზიიდან გამოსასვლელად, ვიდრე არამწველებს? 218 არამწველისგან 135-ს ერთ საათზე მეტი დრო დასჭირდა. 300 მწველისგან კი – 220 –ს.
- ა) შეამოწმეთ ეს ჰიპოთეზა, აიღეთ $\alpha = 0.01$.
- ბ) ააგეთ 95%-იანი ნდობის ინტერვალი მწველთა და არამწველთა პროპორციების სხვაობისათვის
9. 1992 წელს 625 ავტოსაგზაო შემთხვევიდან 273-ის სიკვდილის მიზეზი იყო მძღოლის ნასვამ მდგომარეობაში ყოფნა. 2002 წელს კი – 404-დან 109. ნიშნავს თუ არა ეს, რომ 2002 წელს შემცირდა ავტო-საგზაო შემთხვევების დროს სიკვდილიანობა, ნასვამი მძღოლების მიზეზით.
10. 500 შემთხვევით შერჩეული ავსტრალიელიდან 100-ის ასაკი იყო 15 წელზე ნაკლები. 200 ახალზელანდიელიდან 44-ის ასაკი 15 წელზე ნაკლებია. არის თუ არა ახალ ზელანდიაში უფრო ახალგაზრდა მოსახლეობა?
- ა) იპოვეთ საერთო პოპულაციის ნერტილოვანი შეფასება $\hat{p}$.
- ბ) გამოიყენეთ ეს შეფასება ჰიპოთეზის შესამოწმებლად, აიღეთ $\alpha = 0.05$.
11. ახალ ზელანდიაში ძალადობრივი დამნაშავე ქალების პროპორცია (30.45%) მეტია მამაკაცების პროპორციაზე (21.6%). ციხეში 79 ქალიდან 24 იჯდა მკვლელობისათვის, 1817 მამაკაციდან კი – 392.
- ა) გამოიყენეთ ეს მონაცემები და შეამოწმეთ, არის თუ არა მნიშვნელოვანი განსხვავება მოძალადე ქალებისა და მამაკაცების პროპორციებს შორის. გამოიყენეთ 10%-იანი მნიშვნელოვნების დონე.
- ბ) ნიშნავს თუ არა ეს, რომ ბნელ ადგილას ქალები უფრო საშიშნი არიან, ვიდრე მამაკაცები?

12. გაცხადდა რომ საკეისრო კვეთით დაბადებული ბავშვების პროპორცია წლიდან წლამდე 1.3%-ით იზრდება. 2002 წელს 667 ქალიდან 124-ს გაუკეთეს საკეისრო კვეთი. 2003 წელს კი 678-დან 135-ს. იყო თუ არა ამ წლებში მნიშვნელოვანი ზრდა? გამოიყენეთ $\alpha = 0.05$.
13. 1982 წელს გამოკითხეს 350 რესპონდენტი, თუ რამდენ საათს უყურებენ ისინი ტელევიზორს. საშუალო დრო იყო 4.1 საათი, სტანდარტული გადახრა – 3.3 საათი. 1994 წელს გამოკითხეს 1965 რესპონდენტი, ყურების საშუალო დრო იყო 2.8 საათი, სტანდარტული გადახრა – 2.0 საათი.
- ა) ააგეთ 95%-იანი ნდობის ინტერვალის საშუალო დროების სხვაობისათვის. თქვენი აზრით, მნიშვნელოვნად განსხვავდებიან თუ არა ეს დროები ერთმანეთისაგან? გამოიყენეთ $\alpha = 0.05$.
- ბ) როგორ ფიქრობთ, არის თუ არა მონაცემები ნორმალურად განაწილებული?
14. მოცემულია 7-კატეგორიანი პოლიტიკური იდეოლოგიის სკალაზე SPSS პროგრამით დამუშავებული მონაცემების ამონაბეჭდი:

Variable: IDEOLOGY				
YEAR	N	Mean	StdDev	Std Error
1978	627	4.03	1.26	0.0503
1994	2879	4.17	1.39	0.0259
Variances	T		Prob > T	
Unequal	2.474		0.0134	

- ა) არის თუ არა 1978 და 1994 წლის მონაცემები დამოუკიდებელი?
- ბ) ააგეთ პოლიტიკური იდეოლოგიის ორი წლის სხვაობის 95%-იანი ნდობის ინტერვალის.
- გ) შეიცვალა თუ არა პოლიტიკური იდეოლოგია 1978 წლიდან 1994 წლამდე? გამოიყენეთ $\alpha = 0.05$.
15. ახალგაზრდებში დანაშაულის შემთხვევების ხარისხი გაზომეს ოთხ კითხვაზე მიღებული პასუხის საშუალებით. ეს კითხვებია: 1. პქონდათ თუ არა შეხება პოლიცისტან? 2. იყვნენ თუ არა დაკავებული პოლიციის მიერ? 3. იყვნენ თუ არა ნასამართლეთ? 4. იყვნენ თუ არა დაკავებულნი ერთი წლის წინ? დანაშაულის შემთხვევების ხარისხის ქულა უდრიდა დადებითი პასუხების რაოდენობას, ანუ იცვლებოდა ნულიდან ოთხამდე. გამოკითხეს 108 ახალგაზრდა მარტოხელა დედების ოჯახებიდან და 44 ახალგაზრდა ოჯახებიდან, სადაც ორი ბიოლოგიური მშობელია. ახალგაზრდების მიერ დანაშაულის ჩადენის საშუალო ქულა მარტოხელა დედების ოჯახებიდან გამოვიდა – 1.9, სტანდარტული გადახრით – 1.5; ორი ბიოლოგიური მშობლის ოჯახებიდან კი საშუალო – 1.4, სტანდარტული გადახრით – 1.5.
- ა) ააგეთ 95%-იანი ნდობის ინტერვალის პოპულაციების საშუალოთა სხვაობისათვის.
- ბ) შეამოწმეთ ჰიპოთეზა საშუალოების ტოლობის შესახებ ორმხრივი ალტერნატივის წინააღმდეგ, $\alpha = 0.05$. იპოვეთ p -სიდიდე.
- გ) როგორ ფიქრობთ, ნორმალურადაა თუ არა განაწილებული დანაშაულების ქულები?

16. 1977 წელს ქვეყანაში სოციალურად დაუცველი ბავშვიანი ოჯახების 7.7% იყო სიღარიბის ზღვარს ქვემოთ, 1993 წელს კი 11.4%. რას უდრის სიტუაციის შესა-
დარებელი სტატისტიკის წერტილოვანი შეფასება?
17. იანვარში 1600 შემთხვევით გამოკითხული კანადელიდან 800 მხარს უჭერდა პრემიერმინისტრის გადაწყვეტილებებს. ერთი თვის შემდეგ, თებერვალში 1600 გამოკითხულიდან პრემიერმინისტრის გადაწყვეტილებებს მხარს უჭერდა 944.
 p_1 -ით აღვნიშნოთ იმ კანადელთა პროპორცია, ვინც მხარს უჭერდა პრემიერმი-
ნისტრს იანვარში, p_2 -ით კი – თებერვალში.
ა) იპოვეთ ამ პროპორციების წერტილოვანი შეფასებები. იპოვეთ ამ შეფასებე-
ბის სხვაობა.
ბ) ააგეთ 99%-იანი ნდობის ინტერვალის $p_2 - p_1$ -თვის. რა დასკვნას გამოიტანთ?
18. ცხრილში 10.1 მოცემულია გამოკითხვის შედეგები.

ცხრილი 10.1.

სიკვდილის შემდეგ ცხოვრების რწმენა			
სქესი	მჯერა	არ მჯერა	სულ
მდედრობითი	435	147	582
მამრობითი	375	134	509

- ა) იპოვეთ ქალების და მამაკაცების პროპორციების წერტილოვანი შეფასება, ვისაც სჯერა, რომ არსებობს ცხოვრება სიკვდილის შემდეგ.
- ბ) შეამოწმეთ $H_0: p_1 = p_2$ ჰიპოთეზა $H_A: p_1 \neq p_2$ ალტერნატივის წინააღმდეგ, $\alpha = 0.05$.
- გ) რა ტიპის შეცდომა შეიძლება იყოს დაშვებული?
19. ცხრილში 10.2 მოცემულია სტუდენტების ქულები მათემატიკაში ტრენინგის წინ და შემდეგ.

ცხრილი 10.2.

სტუდენტი	ქულა ტრენინგის წინ	ქულა ტრენინგის შემდეგ
1	60	70
2	73	80
3	42	40
4	88	94
5	66	79
6	77	86
7	90	93
8	63	71
9	55	70
10	96	97

- ა) გამოთვალეთ ქულათა საშუალო ტრენინგის წინ და ტრენინგის გავლის შემდეგ. გამოთვალეთ მიღებული საშუალოების სხვაობა. იპოვეთ წყვილების სხვაობების საშუალო და შეადარეთ საშუალოთა სხვაობას.
- ბ) შეამოწმეთ $H_0: \mu_d = 0$ ჰიპოთეზა $H_A: \mu_d > 0$ ალტერნატივის საწინააღმდეგოდ, $\alpha = 0.05$
20. არის თუ არა განსხვავება ბიჭებისა და გოგონების მიერ ერთ სემესტრში აღებული კრედიტების საშუალო რაოდენობას შორის ($\alpha = 0.05$).

ბიჭები	გოგონები
$n = 60$	$n = 60$
$\bar{X} = 14.24$ კრედიტი	$\bar{Y} = 15.65$ კრედიტი
$S_1 = 1.2$ კრედიტი	$S_2 = 1.56$ კრედიტი

21. საქველმოქმედო საქმიანობაზე მეტი ფული შემოდის კერძო ბიზნესიდან თუ სამთავრობო ორგანიზაციებიდან? აიღეთ $\alpha = 0.05$. 70 გამოკითხული ბიზნესმენიდან ერთი წლის წინ 23-მა გადარიცხა ფული საქველმოქმედო საქმიანობისათვის. 50 სამთავრობო ორგანიზაციაში მომუშავე თანამშრომლიდან კი – 19-მა.

მეათე თავში განვიხილეთ ორი დამოუკიდებელი ნორმალური შერჩევის საშუალოების შედარების ამოცანა. კერძოდ, შევამოწმეთ ჰიპოთეზა პოპულაციების საშუალოების ტოლობის შესახებ და ავაგეთ ნდობის ინტერვალი პოპულაციების საშუალოთა სხვაობისთვის. ამ მიზნით გამოვიყენეთ t -კრიტერიუმი.

ვთქვათ, ახლა გვაქვს ორზე მეტი შერჩევა და გვაინტერესებს შევამოწმოთ ჰიპოთეზა საშუალოთა ტოლობის შესახებ. ამ სიტუაციაში ვერ გამოვიყენებთ t -კრიტერიუმს, რადგანაც, მრავალჯერადი შედარების დროს (შერჩევების თითოეული წყვილის შედარებისას), შეცდომების ალბათობა ძალიან სწრაფად იზრდება შედარებების რიცხვის ზრდასთან ერთად.

მეთოდს, რომელიც გვაძლევს რამდენიმე საშუალოს შედარების საშუალებას, ეწოდება დისპერსიული ანალიზი (რომლის აბრევიატურაა – ANOVA). ერთი შეხედვით შეიძლება უცნაურად მოგვეჩვენოს, რომ ტესტს, რომელიც საშუალოთა შესადარებლად გამოიყენება, დისპერსიული ანალიზი ეწოდება. მიზეზი იმაში მდგომარეობს, რომ ეს მეთოდი ეყრდნობა დისპერსიების შეფარდებას, როგორც ცვალებადობის საზომს შერჩევების საშუალოებს შორის, ასევე თვით შერჩევის შიგნით.

იმისათვის, რომ გამოვიყენოთ აღნიშნული ტესტი – ANOVA, საჭიროა:

- ყველა პოპულაცია იყოს ნორმალური;
- პოპულაციების დისპერსიები იყოს ტოლი.

ერთფაქტორიანი დისპერსიული ანალიზი ცდისპირთა შორის სქემისათვის (ჰიპოთეზები)

გვინდა შევამოწმოთ ძირითადი

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

ჰიპოთეზა

$$H_A: \text{ყველა } \mu_k \text{ არ არის ერთმანეთის ტოლი}$$

ალტერნატივის წინააღმდეგ.

არ მოვითხოვთ, რომ ყველა საშუალო განსხვავებული იყოს ერთმანეთისგან. ნულოვან ჰიპოთეზას უარყოფთ მაშინაც, როცა, თუნდაც, საშუალოების რომელიმე ერთი წყვილი მაინც იქნება ერთმანეთისგან განსხვავებული.

კრიტერიუმის სტატისტიკა

გვაქვს დამოუკიდებელი შემთხვევითი შერჩევები k პოპულაციიდან. პირველი პოპულაციიდან გვაქვს n მონაცემი:

$$X_{11}, X_{12}, \dots, X_{1n},$$

მეორე პოპულაციიდანაც გვაქვს n მონაცემი:

$$X_{21}, X_{22}, \dots, X_{2n}$$

და ა. შ. k -ური პოპულაციიდანაც გვაქვს n მონაცემი:

$$X_{k1}, X_{k2}, \dots, X_{kn}.$$

თითოეულ შერჩევას აქვს თავისი საშუალო და სტანდარტული გადახრა. შევადგინოთ 11.1 ცხრილი. შემოვიღოთ აღნიშვნები:

$SST = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X})^2$ - არის X_{ij} მონაცემების მათი საერთო საშუალო მნიშვნელობიდან გადახრების კვადრატების სრული ჯამური სიდიდე (ტოტალ sum of squares). ნულოვანი ჰიპოთეზის სამართლიანობისას $MST = \frac{SST}{n-1}$ არის უცნობი σ^2 დისპერსიის ჩაუნაცვლებელი შეფასება, ანუ S^2 (გვაქვს k ჯგუფი და თითოეულ ჯგუფში n მონაცემი, $j = 1, 2, \dots, n$; $i = 1, 2, \dots, k$. სულ გვაქვს $N = nk$ მონაცემი).

ცხრილი 11.1

	ჯგუფი 1	ჯგუფი 2		ჯგუფი k	სულ
	X_{11}	X_{21}	...	X_{k1}	k
	X_{12}	X_{22}	...	X_{k2}	k
	...	...	...	...	...
	X_{1n}	X_{2n}	...	X_{kn}	k
შერჩევის მოცულობა	n	n	...	n	$N = kn$
საშუალო	$\bar{X}_1$	$\bar{X}_2$	...	$\bar{X}_k$	$\bar{X}$
სტანდარტული გადახრა	s_1	s_2	...	s_k	S

აქ $\bar{X}$ არის საშუალოების საშუალო, სრული საშუალო (ანუ ყველა მონაცემთა საშუალო):

$$\bar{X} = \frac{X_{11} + \dots + X_{kn}}{N} = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k}{k},$$

სადაც N შერჩევის მთლიანი რაოდენობაა, S არის შერჩევის მონაცემების სტანდარტული გადახრა, S^2 დისპერსიის შეფასებაა.

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

- ყოველი ჯგუფის ვარიაციათა ჯამია, მას შეიძლება შევხედოთ, როგორც ჯგუფებში ვარიაციის (ცვალებადობის) მახასიათებელს (sum of square within groups). ნულოვანი ჰიპოთეზის სამართლიანობის დროს $MSE = \frac{SSE}{n-k}$

უცნობი σ^2 დისპერსიის შეფასებაა.

$$SSR = \sum_{i=1}^k \sum_{j=1}^n (\bar{X}_i - \bar{\bar{X}})^2 = n \cdot \sum_{i=1}^k (\bar{X}_i - \bar{\bar{X}})^2 - \text{არის ჯგუფის საშუალო მნიშვნელობის}$$

საერთო საშუალოდან გადახრების კვადრატების ჯამი (sum of square between groups). მას შეიძლება შევხედოთ, როგორც ჯგუფთაშორისი ცვალებადობის მახასია-

თებელს. ნულოვანი ჰიპოთეზის სამართლიანობის დროს $MSR = \frac{SSR}{k-1}$ არის იგივე

უცნობი σ^2 დისპერსიის შეფასება.

თუ უცნობი σ^2 დისპერსიის აღნიშნული სამი შეფასება გვაძლევს დაახლოებით ერთნაირ რიცხვით მნიშვნელობებს, მაშინ ჯგუფის საშუალოებს შორის არ არის მნიშვნელოვანი სხვაობა. პირიქით, თუ დისპერსიის შეფასებები მნიშვნელოვნად განსხვავდებიან და, ამასთან, $\frac{SSR}{k-1}$ მნიშვნელოვნად ჭარბობს $\frac{SSE}{N-k}$ -ს, მაშინ უნდა უარვყოთ ნულოვანი ჰიპოთეზა, ანუ ჩავთვალოთ, რომ ზოგიერთი ჯგუფის საშუალოებს შორის მნიშვნელოვანი განსხვავებაა.

უცნობი σ^2 დისპერსიის ორი განხილული შეფასება ეფუძნება ჯგუფთა შორის SSR და ჯგუფებში შიგნით ცვალებადობას SSE . მათ ფარდობას ნულოვანი ჰიპოთეზის სამართლიანობის პირობებში აქვს F -განაწილება $k-1$ და $N-k$ თავისუფლების ხარისხებით. ეს საშუალებას გვაძლევს ჰიპოთეზის შესამოწმებლად გამოვიყენოთ:

$$F = \frac{\frac{SSR}{k-1}}{\frac{SSE}{N-k}} = \frac{MSR}{MSE} \equiv F(k-1, N-k)$$

სტატისტიკის დაკვირვებული (ჩვენ მიერ გამოთვლილი) მნიშვნელობა.

გადაწყვეტილების მიღების წესი

დავაფიქსიროთ მნიშვნელოვნების დონე α . ვიპოვოთ F -განაწილების შესაბამისი კრიტიკული მნიშვნელობა ფიშერის ზედა კრიტიკული წერტილების α -ალბათობის შესაბამისი ცხრილიდან; ვიპოვოთ $k-1$ და $N-k$ თავისუფლების ხარისხების მქონე ფიშერის განაწილების ზედა α -კრიტიკული წერტილი - $F_{crit} = F_{k-1, N-k, \alpha}$. თუ F სტატისტიკის დაკვირვებული მნიშვნელობაა f და $f > F_{crit}$, მაშინ $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ ჰიპოთეზას დავინუნებთ, წინააღმდეგ შემთხვევაში H_0 -ის უარყოფის საფუძველი არ გავაჩნია.

ANOVA ცხრილში არის სამი სტრიქონი: პირველი სტრიქონი შეიცავს ინფორმაციას ჯგუფთაშორის ცვალებადობაზე, მეორე - ჯგუფთა შიგნით ცვალებადობაზე, ხოლო მესამე სტრიქონი კი ტესტისთვის არ არის საჭირო, მაგრამ საჭიროა გამოთვლების სისწორის შესამოწმებლად:

$$SST = SSR + SSE, \quad N-1 = N-k+k-1.$$

თავისუფლების ხარისხი $N-1$, რომელიც ნაჩვენებია მესამე სტრიქონში, ეფუძნება N დაკვირვებას მინუს ერთი დამაკავშირებელი თანაფარდობა (ერთი ბმა) სრული საშუალოს შეფასებასთან. მეორე სტრიქონის თავისუფლების ხარისხი $N-k$ ეფუძნება N დაკვირვებას მინუს k დამაკავშირებელი თანაფარდობა k ჯგუფის საშუალოს შეფასებებთან.

პირველი სტრიქონის თავისუფლების ხარისხი $k-1$ ეფუძნება k ჯგუფის საშუალოს მინუს ერთი დამაკავშირებელი თანაფარდობა სრულ საშუალოსთან.

ცხრილი 11.2

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	SSR	$k-1$	$MSR = \frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$
ჯგუფთა შიგნით	SSE	$N-k$	$MSE = \frac{SSE}{N-k}$	
სულ	SST	$N-1$		

მაგალითი 11.1. ტყის ოთხ უბანზე შეაგროვეს სოკო. ცხრილში მოცემულია თითოეულ უბანზე 5 სხვადასხვა ადგილას ნაპოვნი სოკოს რაოდენობა:

უბანი	რაოდენობა					საშუალო
1	8	7	6	4	5	$\bar{X}_1 = 6$
2	14	14	12	7	8	$\bar{X}_2 = 11$
3	6	9	7	8	5	$\bar{X}_3 = 7$
4	4	7	4	3	2	$\bar{X}_4 = 4$
$k = 4, n = 5, N = nk = 20$						

შევამოწმოთ $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ჰიპოთეზა, H_A : ყველა μ_i არ არის ერთმანეთის ტოლი ალტერნატივის წინააღმდეგ.

გამოვთვალოთ საერთო საშუალო და ცვალებადობის მახასიათებლები. გვაქვს:

$$\bar{\bar{X}} = \frac{6+11+7+4}{4} = 7;$$

$$SSR = n[(\bar{X}_1 - \bar{\bar{X}})^2 + (\bar{X}_2 - \bar{\bar{X}})^2 + (\bar{X}_3 - \bar{\bar{X}})^2 + (\bar{X}_4 - \bar{\bar{X}})^2] =$$

$$= 5[(6-7)^2 + (11-7)^2 + (7-7)^2 + (4-7)^2] = 5(1+16+0+9) = 130$$

$$\begin{aligned}
 SSE &= \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 = (8-6)^2 + (7-6)^2 + (6-6)^2 + (4-6)^2 + (5-6)^2 + \\
 &+ (14-11)^2 + (14-11)^2 + (12-11)^2 + (7-11)^2 + (8-11)^2 + \\
 &+ (6-7)^2 + (9-7)^2 + (7-7)^2 + (8-7)^2 + (5-7)^2 + \\
 &+ (4-4)^2 + (7-4)^2 + (4-4)^2 + (3-4)^2 + (2-4)^2 = 78, \\
 k-1 &= 4-1 = 3, \quad N-k = 20-4 = 16.
 \end{aligned}$$

შევავსოთ ANOVA ცხრილი.

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	130	3	43.33	8.888
ჯგუფთა შიგნით	78	16	4.875	
სულ	208	19		

ტესტი ცალმხრივია. ვიპოვოთ კრიტიკული მნიშვნელობა F განაწილების ცხრილიდან. $F_{\alpha, n} = F_{3, 16, 0.05} = 3.2389$. ტესტის სტატისტიკის 8.888 მნიშვნელობა მეტია 3.2389 კრიტიკულ მნიშვნელობაზე, ამიტომ ნულოვანი ჰიპოთეზა უნდა უარყოფოთ. დასკვნა: სხვადასხვა უბნებზე შეგროვილი სოკოს საშუალო რაოდენობები განსხვავებულია ერთმანეთისაგან.

ტესტი არ განსაზღვრავს კონკრეტულად რომელი საშუალოები განსხვავდებიან ერთმანეთისაგან. როცა გვაქვს ოთხი შერჩევის საშუალო, შეგვიძლია გავაკეთოთ ექვსი შედარება ორ საშუალოს შორის:

$\bar{X}_1$ შევადაროთ $\bar{X}_2$ -ს;

$\bar{X}_1$ შევადაროთ $\bar{X}_3$ -ს;

$\bar{X}_1$ შევადაროთ $\bar{X}_4$ -ს;

$\bar{X}_2$ შევადაროთ $\bar{X}_3$ -ს;

$\bar{X}_2$ შევადაროთ $\bar{X}_4$ -ს;

$\bar{X}_3$ შევადაროთ $\bar{X}_4$ -ს.

შედარებები უნდა მოვახდინოთ მხოლოდ მაშინ, თუ ნულოვანი ჰიპოთეზა უარყოფილია. შედარების მრავალი კრიტერიუმი არსებობს, ჩვენ გამოვიყენებთ ტუკის კრიტერიუმს, რომელიც გამოითვლება ფორმულით:

$$CD = q \sqrt{\frac{MSE}{n}}.$$

ტუკის კრიტერიუმი ინარჩუნებს პირველი გვარის შეცდომის დაშვების ალბათობას α -ს ტოლად ყველა წყვილთაშორის შედარებაში.

თუ ორ საშუალოს შორის სხვაობის აბსოლუტური მნიშვნელობა მეტია ან ტოლია ტუკის კრიტერიუმის მნიშვნელობაზე, მაშინ პოპულაციების საშუალოები მნიშვნელოვნად განსხვავდებიან ერთმანეთისაგან.

q მნიშვნელობა დამოკიდებულია არჩეულ მნიშვნელოვნების დონეზე, ჯგუფების (შესადარებელი საშუალოების) k რაოდენობაზე და $N-k$ თავისუფლების ხარისხზე. მისი მნიშვნელობა უნდა მოვძებნოთ სტიუდენტის რანგობრივი სტატიკის q -ს მნიშვნელობების ცხრილიდან. ჩვენს მაგალითში: $\alpha = 0.05$, $k = 4$ და $N - k = 16$. ამიტომ ცხრილიდან დავინახავთ, რომ $q = 4.05$. გარდა ამისა, $n = 5$ და $MSE = 4.875$. შესაბამისად:

$$CD = q \sqrt{\frac{MSE}{n}} = 4.05 \cdot \sqrt{\frac{4.875}{5}} = 3.999.$$

განსხვავებულია იმ პოპულაციების საშუალოები, რომელთა შერჩევით საშუალოს შორის სხვაობის აბსოლუტური მნიშვნელობა მეტია ან ტოლია 3.999-ზე. ჩვენს მაგალითში გვაქვს:

$$\bar{X}_1 - \bar{X}_2 = |6 - 1| = 5 > 3.999 \rightarrow \mu_1 \neq \mu_2.$$

$$\bar{X}_1 - \bar{X}_3 = |6 - 7| = 1 < 3.999.$$

$$\bar{X}_1 - \bar{X}_4 = |6 - 4| = 2 < 3.999.$$

$$\bar{X}_2 - \bar{X}_3 = |11 - 7| = 4 > 3.999 \rightarrow \mu_2 \neq \mu_3.$$

$$\bar{X}_2 - \bar{X}_4 = |11 - 4| = 7 > 3.999 \rightarrow \mu_2 \neq \mu_4.$$

$$\bar{X}_3 - \bar{X}_4 = |7 - 4| = 3 < 3.999.$$

დასკვნა: მეორე უბანზე სოკოს საშუალო რაოდენობა მნიშვნელოვნად განსხვავდება ყველა დანარჩენ სამ უბანზე სოკოს საშუალო რაოდენობისგან.

განხილულ სტატისტიკურ მეთოდს ეწოდება ერთფაქტორიანი დისპერსიული ანალიზი ცდისპირთა შორის სქემისთვის. ის გამოიყენება სამი ან მეტი პოპულაციის საშუალოების შესადარებლად, შერჩევები ალბულის ერთმანეთისაგან დამოუკიდებლად. ამ მაგალითში ფაქტორი არის უბანი და ამ ფაქტორს აქვს 4 დონე.

ცდისპირთა შიდა ორფაქტორიანი სქემა და დისპერსიული ანალიზი

მაგალითი 11.2. შევამოწმოთ ჰიპოთეზა, რომ არ არსებობს განსხვავება სახლების შეფასებაში, რომელიც წარდგენილია სამი განსხვავებული შემფასებელი ექსპერტის მიერ. აფასებენ სახლის მახასიათებლებს: ზომა, ხარისხი, სახლის განლაგება და ა. შ. ასეთ მახასიათებლებს ხშირად ბლოკებს უწოდებენ. ცხრილში მოყვანილია სახლის სამი ექსპერტის შეფასება 5 მახასიათებლის მიხედვით:

	შემფასებელი			
მახასიათებელი	1	2	3	$\bar{X}_j$
1	78.0	82.0	79.0	79.67
2	102.0	102.0	99.0	101.00
3	68.0	74.0	70.0	70.67
4	83.0	88.0	86.0	85.67
5	95.0	99.0	92.0	95.33
$\bar{X}_i$	85.2	89.0	85.2	

გამოვთვალოთ საერთო საშუალო :

$$\bar{\bar{X}} = \frac{85.2 + 89.0 + 85.2}{3} = \frac{79.67 + 101.0 + 70.67 + 85.67 + 95.33}{5} = 86.47 .$$

უნდა გამოვიყენოთ ორ ფაქტორიანი დისპერსიული ანალიზი იმისათვის, რომ შევამოწმოთ ჰიპოთეზა, არსებობს თუ არა განსხვავება სამი ექსპერტის შეფასებაში.

შესამოწმებელი გვაქვს:

ძირითადი ჰიპოთეზა: $H_0: \mu_1 = \mu_2 = \mu_3$,

ალტერნატიული ჰიპოთეზა: H_A : ყველა საშუალო ერთმანეთის ტოლი არ არის.

ისევე როგორც ერთფაქტორიან ანალიზში, შემოვიღოთ მონაცემთა დისპერსიის შეფასებები:

$$SST = \sum_{i=1}^k \sum_{j=1}^r (X_{ij} - \bar{\bar{X}})^2 ,$$

სადაც k სვეტების (ექსპერტების) რაოდენობაა, r -სტრიქონების, ბლოკების (სახლის მახასიათებლების) რაოდენობა, X_{ij} ინდივიდუალური შეფასება (1000 დოლარი), $\bar{\bar{X}}$ სრული (საერთო) საშუალო (86.47). გვაქვს:

$$SST = (78.0 - 86.47)^2 + (102.0 - 86.47)^2 + \dots + (92.0 - 86.47)^2 = 1829.73 .$$

ერთფაქტორიან ANOVA-ში სრულდებოდატოლობა:

$$SST = SSR + SSE$$

ორფაქტორიანი ANOVA-ს შემთხვევაში მონაცემთა სრული ცვალებადობა იყოფა სამ ნაწილად ორის ნაცვლად:

$$SST = SSR + SSE + SSB ,$$

სადაც SST = არის X_{ij} მონაცემების საერთო საშუალო მნიშვნელობიდან გადახრების კვადრატების სრული ჯამური სიდიდე, SSR - გადახრების კვადრატების ჯამი ჯგუფებს შორის (ექსპერტებს შორის), SSB - გადახრების კვადრატების ჯამი

ბლოკებს შორის (მახასიათებლებს შორის), SSE – გადახრების კვადრატების ჯამი ჯგუფებს შიგნით.

ჩვენს მაგალითში:

$$SSR = \sum_{i=1}^I r \left(\bar{X}_i - \bar{\bar{X}} \right)^2,$$

სადაც $\bar{X}_i$ თითოეული ჯგუფის საშუალოა, r – სტრუქტონების, ბლოკების რაოდენობა – $r = 5$, k – ჯგუფების (ექსპერტების) რაოდენობა – $k = 3$:

$$SSR = 5(85.2 - 86.47)^2 + 5(89.0 - 86.47)^2 + 5(85.2 - 86.47)^2 = 48.13,$$

$$SSB = \sum k \left(\bar{X}_j - \bar{\bar{X}} \right)^2 = 3(79.67 - 86.47)^2 + \dots + 3(95.33 - 86.47)^2 = 1758.42,$$

$$SSE = SST - SSR - SSB = 1829 - 48.13 - 1758.42 = 23.18.$$

გამოსათვლელი გვაქვს ორი F სიდიდე, რადგანაც შესაძომნებელი გვაქვს ბლოკებთან დაკავშირებული მეორადი ჰიპოთეზა: ახდენს თუ არა შეფასებაზე გავლენას ის, თუ რა მახასიათებლით შევაფასებთ სახლს.

მეორადი ჰიპოთეზა:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5,$$

H_A : ყველა ბლოკის საშუალო არ არის ერთმანეთის ტოლი.

ჩვენ ჯამონმებთ ტოლია თუ არა საშუალო შეფასება მახასიათებლების მიხედვით. თუ H_0 ჰიპოთეზა უარყოფილია, ეს ნიშნავს იმას, რომ ბლოკირება გამართლებულია.

შევავსოთ ANOVA ცხრილი.

ჯგუფთა შორის	SSR	$k-1$	$\frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$
ბლოკებს შორის	SSB	$r-1$	$\frac{SSB}{r-1}$	$F = \frac{MSB}{MSE}$
ჯგუფთა შიგნით	SSE	$(k-1)(r-1)$	$\frac{SSE}{(k-1)(r-1)}$	
სულ	SST	$N-1$		

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	48.13	2	$\frac{48.13}{2} = 24.065$	8.31
ბლოკებს შორის	1758.42	4	$\frac{1758.42}{4} = 439.605$	151.72
ჯგუფთა შიგნით	23.18	8	$\frac{23.18}{8} = 2.898$	
სულ	1829.73	14		

შენიშვნა: $SSE = SST - (SSR + SSB)$

F -ის კრიტიკული მნიშვნელობა, როცა $\alpha = 0.05$ და თავისუფლების ხარისხებია 4 და 8, არის 3.84. ვინაიდან $8.31 > 3.84$ -ზე, ამიტომ მეორად ნულოვან ჰიპოთეზას უარყოფთ: მახასიათებლის არჩევა გაელენას ახდენს შეფასებაზე.

რადგანაც F -ის კრიტიკული მნიშვნელობა, როცა $\alpha = 0.05$ და თავისუფლების ხარისხებია 2 და 8, არის 4.46 და $8.31 > 4.46$, ამიტომ ძირითადი ჰიპოთეზაც უარყოფილი იქნება: სამი ექსპერტი განსხვავებულად აფასებს სახლს.

რადგან ნულოვანი ჰიპოთეზა უარყოფილია, უნდა გავიგოთ კონკრეტულად რომელი საშუალოები განსხვავდებიან ერთმანეთისაგან. ამისათვის გამოვთვალოთ საშუალოების სხვაობების აბსოლუტური მნიშვნელობები და შევადაროთ CD კრიტიკულ სხვაობას, რომელიც გამოითვლება ფორმულით:

$$CD = t_{crit} \cdot \sqrt{MSE} \cdot \sqrt{\frac{2}{n}},$$

სადაც $n = 5$ ჯგუფში მონაცემთა რაოდენობაა, $t_{crit} = t_{(k-1)(r-1), \alpha/2} = t_{8, 0.025} = 2.306$, $MSE = 2.898$. გვაქვს:

$$CD = t_{8, \alpha/2} \cdot \sqrt{MSE} \cdot \sqrt{\frac{2}{n}} = 2.306 \cdot \sqrt{2.898} \cdot \sqrt{\frac{2}{5}} = 2.483.$$

თუ ორ საშუალოს შორის სხვაობის აბსოლუტური მნიშვნელობა მეტია ან ტოლია CD კრიტიკულ სხვაობაზე, მაშინ პოპულაციების საშუალოები მნიშვნელოვნად განსხვავდებიან ერთმანეთისაგან. გვაქვს:

სხვაობა	გადაწყვეტილება
$ 85.2 - 89.0 = 3.8 > 2.483$	$\mu_1 \neq \mu_2$
$ 85.2 - 85.2 = 0.0 < 2.483$	$\mu_1 = \mu_2$
$ 89.0 - 85.2 = 3.8 > 2.483$	$\mu_2 \neq \mu_3$

მეორე ექსპერტი განსხვავებულად აფასებს.

ცდისპირთა შორის ორფაქტორიანი სქემა და დისპერსიული ანალიზი

მაგალითი 11.3. ფსიქოლოგები სწავლობენ მუსიკის გავლენას ადამიანის მიერ სხვადასხვა ამოცანის შესრულებაზე. 12 სტუდენტ გოგონას და 12 სტუდენტ ვაჟს სთხოვეს გაეკეთებინათ მოკლე სტატიის კორექტურა (სტატიაში იყო 25 შეცდომა). ერთი ჯგუფი ასრულებდა სამუშაოს მუსიკის თანხლების გარეშე, მეორე კი პოპულარული მელოდიის თანხლებით. აღმოჩენილი შეცდომების რაოდენობა მოყვანილია ცხრილში:

აღმოჩენილი შეცდომების რაოდენობა			
ფაქტორი A			
		მუსიკის გარეშე A_1	მუსიკის თანხლებით A_2
		4	16
		5	14
მამრობითი B_1		7	17
		6	19
		10	16
		4	20
$\bar{X}_{A_1 B_1} = 6, \quad \bar{X}_{A_2 B_1} = 17, \quad \bar{X}_{B_1} = 11.5$			
მდედრობითი B_2	187		
	136		
	1610		
	165		
$\bar{X}_{A_1 B_2} = 15, \quad \bar{X}_{A_2 B_2} = 7, \quad \bar{X}_{B_2} = 11$			
$\bar{X}_{A_1} = 10.5, \quad \bar{X}_{A_2} = 12, \quad \bar{X} = 11.25$			

ფაქტორი A – მუსიკის ტიპი, ფაქტორი B – სქესი. A ფაქტორს აქვს ორი დონე, B ფაქტორსაც აქვს ორი დონე.

$\bar{X}_{A_1 B_1}$ – ვაჟების მიერ აღმოჩენილი შეცდომების საშუალო რაოდენობა მუსიკის გარეშე, ტოლია 6-ის (უჯრედის საშუალო);

$\bar{X}_{A_1 B_2}$ – გოგონების მიერ აღმოჩენილი შეცდომების საშუალო რაოდენობა მუსიკის გარეშე, ტოლია 15-ის (უჯრედის საშუალო);

$\bar{X}_{A_1}$ – შეცდომების საშუალო რაოდენობა, აღმოჩენილი მუსიკის გარეშე, ტოლია 10.5-ის (სვეტის საშუალო);

$\bar{X}_{A_2 B_1}$ – ვაჟების მიერ აღმოჩენილი შეცდომების საშუალო რაოდენობა მუსიკის თანხლებით, არის 17 (უჯრედის საშუალო);

$\bar{X}_{A_2 B_2}$ – გოგონების მიერ აღმოჩენილი შეცდომების საშუალო რაოდენობა მუსიკის თანხლებით არის 7 (უჯრედის საშუალო);

$\bar{X}_{A_2}$ – შეცდომების საშუალო რაოდენობა, აღმოჩენილი მუსიკის თანხლებით, არის 12 (სვეტის საშუალო);

$\bar{X}_{A_1}$ – შეცდომების საშუალო რაოდენობა, აღმოჩენილი ვაჟების მიერ, არის 11.5 (სტრიქონის საშუალო);

$\bar{X}_{B_2}$ – შეცდომების საშუალო რაოდენობა, აღმოჩენილი გოგონების მიერ, არის 11 (სტრიქონის საშუალო);

$$\bar{\bar{X}} = \frac{\bar{X}_{A_1} + \bar{X}_{A_2}}{2} = \frac{\bar{X}_{B_1} + \bar{X}_{B_2}}{2} (\text{სრული საშუალო}) = 11.25.$$

შესამოწმებელია სამი ჰიპოთეზა:

1. მუსიკა არ მოქმედებს აღმოჩენილი შეცდომების რაოდენობაზე – $H_0: \mu_{A_1} = \mu_{A_2}$, $H_A: \mu_{A_1} \neq \mu_{A_2}$;
2. სქესზე არ არის დამოკიდებული აღმოჩენილი შეცდომების რაოდენობა – $H_0: \mu_{B_1} = \mu_{B_2}$, $H_A: \mu_{B_1} \neq \mu_{B_2}$;
3. სქესი და მუსიკის ტიპი ერთმანეთისაგან დამოუკიდებელია.
ამ სამი ჰიპოთეზის შესამოწმებლად საჭიროა ორფაქტორიანი დისპერსიული ანალიზის გამოყენება.

მონაცემთა სრული ცვალებადობა დავეყოთ ოთხ ნაწილად:

$$SST = SSA + SSB + SSA \cdot B + SSE,$$

სადაც SST სრული ცვალებადობაა, N მონაცემთა სრული რაოდენობაა, n მონაცემთა რაოდენობაა ჯგუფში, a არის A ფაქტორის დონეების რაოდენობა, b არის B ფაქტორის დონეების რაოდენობა, SSA არის ცვალებადობა A ფაქტორის გამო (მუსიკის ტიპი), SSB არის ცვალებადობა B ფაქტორის გამო (სქესი), $SSA \cdot B$ არის ცვალებადობა, გამომწვეული A და B ფაქტორის ურთიერთქმედებით, SSE არის ცვალებადობა ჯგუფის შიგნით:

$$SSA = N(\bar{X}_{A_1} - \bar{\bar{X}})^2 = 24(10.5 - 11.25)^2 = 13.5;$$

$$SSB = N(\bar{X}_{B_1} - \bar{\bar{X}})^2 = 24(11.5 - 11.25)^2 = 1.5;$$

$$SSA \cdot B = N(\bar{X}_{A_1 B_1} - \bar{X}_{A_1} - \bar{X}_{B_1} + \bar{\bar{X}})^2 = 24(6 - 10.5 - 11.5 + 11.25)^2 = 541.5;$$

$$SSE = (4-6)^2 + (5-6)^2 + \dots + (10-7)^2 + (5-7)^2 = 96.0.$$

თავისუფლების ხარისხებია:

$$df A = a-1, \quad df B = b-1,$$

$$df A \cdot B = (a-1)(b-1),$$

ჯგუფთა შიგნით $df = ab(n-1)$, ამასთანავე:

$$a-1 + b-1 + (a-1)(b-1) + ab(n-1) = abn-1 = N-1.$$

შევავსოთ ორფაქტორიანი ANOVA ცხრილი:

ცვალებადობის წყარო	SS	df	MS	F
მუსიკა	13.5	1	13.5	2.81
სქესი	1.5	1	1.5	0.31
ურთიერთქმედება	541.5	1	541.5	112.81
ჯგუფთა შიგნით	96.0	20	4.8	
სულ	652.5	23		

$$F_A = \frac{MSA}{MSE}, \quad F_B = \frac{MSB}{MSE}, \quad F_{AB} = \frac{MSA \cdot B}{MSE},$$

$$F_{crit} = F_{1,20,0.05} = 4.35.$$

გადაწყვეტილებები:

1. მუსიკის ტიპი (ფაქტორი A) არ ახდენს გავლენას ($2.81 < 4.35$);
2. სქესი (ფაქტორი B) არ ახდენს გავლენას ($0.31 < 4.35$);
3. მუსიკის ტიპი და სქესი ურთიერთქმედებაში ($112.81 > 4.35$).

ამიტომ, უნდა გავარკვიოთ, რომელი უჯრედის საშუალოები განსხვავდებიან ერთმანეთისაგან. ამისათვის გამოვთვალოთ ტუკის კრიტერიუმის მნიშვნელობა:

$$CD = q \sqrt{\frac{MSE}{n}}.$$

სტიუდენტის რანგობრივი სტატისტიკის q -ს მნიშვნელობების ცხრილიდან, იმის გათვალისწინებით, რომ საქმე გვაქვს ორფაქტორიან სქემასთან: $2 \cdot 2$, $a = 2$, $b = 2$, ეპოულობთ, რომ $q = 3.58$. ამიტომ:

$$CD = 3.58 \cdot \sqrt{\frac{4.8}{6}} = 3.2.$$

გამოვთვალოთ შერჩევით საშუალოს შორის სხვაობის აბსოლუტური მნიშვნელობები:

$$|6 - 17| = 11, \quad |15 - 7| = 8, \quad |6 - 15| = 9, \quad |17 - 7| = 10.$$

ყველა სხვაობა მეტია 3.2-ზე.

გადაწყვეტილება

1. აღმოჩენილი შეცდომების საშუალო რაოდენობა მუსიკის გარეშე და მუსიკის თანხლებით მნიშვნელოვნად არ განსხვავდება ერთმანეთისაგან.
2. შეცდომების საშუალო რაოდენობა, აღმოჩენილი ვაუების მიერ, მნიშვნელოვნად არ განსხვავდება შეცდომების საშუალო რაოდენობისაგან, აღმოჩენილი ქალების მიერ.

3. მნიშვნელოვანია ურთიერთქმედება: მუსიკის ტიპის ეფექტი დამოკიდებულია სქესზე. ვაჟები, მუსიკის მოსმენის დროს, უფრო ბევრ შეცდომას ამჩნევენ, გოგონები კი, უფრო ბევრ შეცდომას პოულობენ, როცა არ უსმენენ მუსიკას. განსხვავება ქალებსა და ვაჟებს შორის დამოკიდებულია მუსიკის ტიპზე: ქალები საშუალოდ უფრო ბევრ შეცდომას ამჩნევენ, როცა არ უსმენენ მუსიკას, ვაჟები კი – როცა უსმენენ.

ურთიერთქმედების ანალიზი გვიჩვენებს, რომ მუსიკა გავლენას ახდენს აღმოჩენილი შეცდომების რაოდენობაზე, მაგრამ ეფექტის მიმართულება დამოკიდებულია სქესზე. უფრო მეტიც, ქალები და მამაკაცები განსხვავდებიან აღმოჩენილი შეცდომების მხრივ, მაგრამ განსხვავება დამოკიდებულია მუსიკის ტიპზე.

შეჯამება

ორი საშუალოს შესადარებლად იყენებენ t -კრიტერიუმს. თუ შესადარებელია ორზე მეტი საშუალო, გამოიყენება სტატისტიკური მეთოდი, ცნობილი, როგორც დისპერსიული ანალიზი.

თუ ყველა შერჩევა აღებულია ნორმალური პოპულაციიდან და პოპულაციების დისპერსიები ტოლია, ნულოვანი ჰიპოთეზის შესამოწმებლად, რომლის თანახმად ყველა შერჩევა აღებულია ერთი და იმავე პოპულაციიდან ალტერნატივის

წინააღმდეგ, იყენებენ $F = \frac{MSR}{MSE}$ სტატისტიკას. ამ სტატისტიკას აქვს F -განაწილება $k-1$ და $N-k$ თავისუფლების ხარისხებით.

თუ დასტურდება, რომ საშუალოები ერთმანეთის ტოლი არ არის, საჭიროა მრავალჯერადი შედარების კრიტერიუმის გამოყენება, იმის დასადგენად, თუ რომელი საშუალოები განსხვავდებიან მნიშვნელოვნად ერთმანეთისაგან.

არსებობს სიტუაციები, სადაც ერთფაქტორიანი დისპერსიული ანალიზი არ არის საკმარისი. მაგალითად, როცა გვინდა შევაფასოთ სხვადასხვა ტიპის აკუმულატორის მუშაობის ვადა, ჩვენ უნდა მოვათავსოთ სხვადასხვა ტიპის აკუმულატორი მანქანაში და გავიმეოროთ ეს ცდა სხვადასხვა ტიპის მანქანისთვის. უნდა ვაკონტროლოთ მანქანების ცვალებადობა, იმისათვის, რომ შევამოწმოთ აკუმულატორების მუშაობის ვადები. ასეთი ტიპის ამოცანებისათვის იყენებენ ცდისპირთა შიდა ერთფაქტორიან სქემას.

ცდისპირთა შორის ორფაქტორიანი სქემის გამოყენება საჭიროა მაშინ, როცა ცდის შედეგები მიიღება ორი ფაქტორის ზემოქმედებით, მაგალითად, როცა გვინტერესებს, განსხვავდება თუ არა ქალებისა და მამაკაცების (პირველი ფაქტორი) რეაქცია ორ განსხვავებულ სტიმულზე (მეორე ფაქტორი).

კითხვები

1. მოქმედებს თუ არა სმენის პრობლემები სტუდენტების შეფასებაზე. ქვემოთ მოცემულია სტუდენტების მიერ ნაკარნახევ ლიტერატურულ ტესტში მიღებული ქულები.

სმენა		
ნორმალური	საშუალო	ცუდი
31	23	12
29	27	18
32	26	17
38	24	13
36	20	15
14	20	15

- ა) რას უდრის n , k , N .
- ბ) შეამოწმეთ $\mu_1 = \mu_2 = \mu_3$ ჰიპოთეზა „სმენის პრობლემები გავლენას ახდენენ მიღებულ ქულებზე“ ალტერნატივის წინააღმდეგ. აიღეთ $\alpha = 0.05$.
- გ) რას უდრის საშუალო ქულის საუკეთესო შეფასება სტუდენტებისთვის ცუდი სმენით. რას უდრის საშუალო ქულის საუკეთესო შეფასება, როცა არ ვიცით რომელ ჯგუფს ეკუთვნის სტუდენტი.
2. შემთხვევით შერჩეული იყო ხუთ-ხუთი სტუდენტი საცხოვრებელი ადგილის მიხედვით, ქვემოთ მოცემულია მათ მიერ მიღებული ქულები.

სოფელი	დაბა	ქალაქი	დედაქალაქი
4	30	5	25
16	10	40	35
27	6	37	40
33	24	23	7
15	25	10	23

- ა) შეამოწმეთ ჰიპოთეზა, მოქმედებს თუ არა საცხოვრებელი ადგილი მომზადების დონეზე. $\alpha = 0.05$.
- ბ) რა ტიპის შეცდომა შეიძლება იყოს დაშვებული?
- გ) რას უდრის სტუდენტის საშუალო ქულის შეფასება?

3. ერთი და იმავე დონის 15 სტუდენტი შემთხვევით გადაანაწილეს ინგლისურის სწავლების სამ ჯგუფში. ერთსა და იმავე ტესტში საკონტროლოში მიღებული ქულებია:

სწავლების მეთოდი			
	1	2	3
ქულები	19	14	15
	13	13	10
	16	16	11
	17	12	12
	20	10	12

ასდენს თუ არა სწავლების მეთოდი გავლენას ტესტში მიღებულ ქულებზე?
 $\alpha = 0.05$.

4. ავსტრალიის ბაყაყები 30 წელს ცოცხლობენ და მთელი სიცოცხლის განმავლობაში აგრძელებენ ზრდას. 1984 წელს 100 ბაყაყი გადაიყვანეს ორ განსხვავებულ ახალ საცხოვრებელ ადგილას. 1996 წელს სამივე საცხოვრებელ ადგილზე დაჭერილი და აწონილი იყო სამ-სამი ბაყაყი. შეამოწმეთ, იქონია თუ არა გავლენა ახალმა ადგილმა მათ წონაზე. აიღეთ $\alpha = 0.05$.

საცხოვრებელი ადგილი	წონა		
ძირითადი	1.67	0.43	2.72
ახალი 1	2.93	0.80	4.18
ახალი 2	2.55	3.80	3.60

5. სამ ახლომდებარე ტბაში დაიჭირეს და აწონეს ხუთ-ხუთი ერთსადიმავე ჯიშის თევზი. განსხვავებულია თუ არა თევზის საშუალო წონები, $\alpha = 0.01$.

პირველი ტბა	285	315	260	290	420
მეორე ტბა	305	340	310	320	270
მესამე ტბა	500	380	480	350	400

- ა) ჩამოაყალიბეთ ნულოვანი და ალტერნატიული ჰიპოთეზები და შეავსეთ ANOVA ცხრილი.
- ბ) რას უდრის ამ ჯიშის თევზის საშუალო წონის ნერტილოვანი შეფასება?
- გ) რას უდრის მესამე ტბის თევზის საშუალო წონის ნერტილოვანი შეფასება?
6. ტყის ოთხ უბანზე გაზომილი იყო ხავსის სისქე (სმ²). A და B უბანი იყო ქარისგან დაუცველი, D და C - ქარისგან დაცული, D უბანი უფრო მშრალია, ვიდრე C უბანი. ჩამოაყალიბეთ ნულოვანი და ალტერნატიული ჰიპოთეზები და შეავსეთ ANOVA ცხრილი.

უბანი სიგანე	A	B	C	D
	7	6	8	7
	2	4	4	4
	4	6	5	2

7. შეავსეთ გამოტოვებული ადგილები ANOVA ცხრილებში და უპასუხეთ კითხვებს.

ცხრილი 1

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	50	1	-	-
ჯგუფთა შიგნით	270	26	-	-
სულ				

ცხრილი 2

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	-	3	6	-
ჯგუფთა შიგნით	-	60	-	-
სულ	170			

- ა) რამდენი რესპონდენტი მონაწილეობდა თითოეულ კვლევაში?
 ბ) თუ $\alpha = 0.05$, იპოვეთ F -ის კრიტიკული მნიშვნელობები.
 გ) რა გადაწყვეტილებას მიიღებთ თითოეული ცხრილის მიხედვით?
8. ინკუბატორში მოათავსეს კვერცხები სხვადასხვა ტემპერატურულ რეჟიმში. ქვემოთ მოყვანილია გამოჩეკილი წინილების წონები:

18°C	82	82	81	82	73
21°C	91	84	90	85	85
22°C	81	83	81	84	81

- ა) ჩამოაყალიბეთ ნულოვანი და ალტერნატიული ჰიპოთეზები და შეავსეთ ANOVA ცხრილი.
 ბ) მოქმედებს თუ არა ტემპერატურული რეჟიმი წინილების წონაზე? ისარგებლეთ $\alpha = 0.05$ მნიშვნელოვნების დონით.
9. შეამჩნიეს, რომ ავსტრალიაში მსჯავრდებულის მიღებული სასჯელი ძარცვისთვის (20-25 წლის მამაკაცები) ეროვნებაზეა დამოკიდებული. შეამოწმეთ ამ ინფორმაციის სისწორე.

	სასჯელი (თვე):					
ევროპელები	12	18	12	12	12	36
წყნარი ოკეანე	18	24	16	24	22	30
მაიორი	24	24	36	24	36	44
სხვები	18	12	16	18	12	12

10. მოქმედებს თუ არა მდინარის ნაპირზე მომუშავე ტყის დასამუშავებელი სანა-
რმოები მდინარის ტემპერატურაზე? ცხრილში წარმოდგენილია წყლის 5 სინჯის
ტემპერატურა მდინარის სხვადასხვა ადგილას:

	ტემპერატურა					$\bar{x}$
5 კმ ობიექტიდან მაღლა	7	7	15	11	9	9.8
ობიექტთან	19	25	22	19	23	21.6
5 კმ ობიექტიდან დაბლა	14	18	18	19	19	17.6
10 კმ ობიექტიდან დაბლა	12	17	12	18	18	15.4
25 კმ ობიექტიდან დაბლა	11	7	10	11	15	10.8

11. გამოიკვლიეს ძილის თავისებურებები 8, 10 და 12 წლის ბავშვებში. ცხრილში
მოყვანილია ძილის ხანგრძლივობა (წთ) დაძინებიდან გაღვიძებამდე:

ასაკი		
8	10	12
580	497	525
525	515	506
562	543	475
590	478	493
575	567	537
603	532	532
594	517	480
536	534	472
612	556	463
592	574	515

- ჩამოაყალიბეთ სტატისტიკური ჰიპოთეზები.
- რისი ტოლია F -ის კრიტიკული მნიშვნელობა 0.05 დონეზე?
- რა გადაწყვეტილებას მიიღებთ სტატისტიკური ჰიპოთეზების შესახებ?
- საჭიროა თუ არა მრავალჯერადი კრიტერიუმის გამოყენება?
- რისი ტოლია ტუკის CD მნიშვნელობა?
- რომელ წყვილთა შორის განსხვავება მნიშვნელოვანია?

12. ფსიქოლოგები სწავლობენ მუსიკის გავლენას ადამიანის მიერ სხვადასხვა ამოცანის შესრულებაზე. 24 ადამიანს სთხოვეს წაეკითხათ მოკლე სტატიის კორექტურა. პირველი ჯგუფის წევრები ამ ამოცანას მუსიკის თანხლების გარეშე ასრულებდნენ, მეორე ჯგუფის წევრები უსმენდნენ პოპულარულ მელოდიას, მესამე ჯგუფის წევრები ასრულებდნენ ამოცანას მძიმე როკის თანხლებით. აღმოჩენილი შეცდომების რაოდენობა 50 შესაძლებლობიდან ქვემოთაა მოყვანილი:

მუსიკის გარეშე	პოპულარული მელოდია	მძიმე როკი
40	34	26
41	39	24
39	38	19
36	40	23
35	34	18
32	35	21
31	29	23
34	36	29

- ა) ახდენს თუ არა მუსიკა გავლენას აღმოჩენილი შეცდომების რაოდენობაზე, ისარგებლეთ $\alpha = 0.05$.
- ბ) თუ საჭიროა გამოიყენეთ ტუკის კრიტერიუმი.
13. 340 დემოკრატი, 313 დამოუკიდებელი, 290 რესპუბლიკელი გამოკითხეს მათი პოლიტიკური იდეოლოგიის შესახებ. პოლიტიკური იდეოლოგიის სკალა შედგებოდა 7 კატეგორიიდან: 1 – ექსტრემალური ლიბერალი, 2 – ლიბერალი, 3 – ოდნავ ლიბერალი, 4 – ზომიერი, 5 – ოდნავ კონსერვატორი, 6 – კონსერვატორი, 7 – ექსტრემალური კონსერვატორი. გამოკითხვის შედეგები მოყვანილია ცხრილში.

**ცხრილი 3. პოლიტიკური იდეოლოგია განპირობებული
პარტიული იდენტიფიკაციით**

პარტია	პოლიტიკური იდეოლოგია									
	1	2	3	4	5	6	7	n	$\bar{x}$	s
დემოკრატი	11	50	60	139	35	39	6	340	3.82	1.32
დამოუკიდებელი	8	33	47	142	37	40	6	313	3.99	1.27
რესპუბლიკელი	2	19	30	99	65	61	14	290	4.53	1.28

დამუშავების შედეგები მოყვანილია ANOVA ცხრილში.

ANOVA ცხრილი

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	85.38	2	42.69	25.6
ჯგუფთა შიგნით	1570.84	940	1.67	
სულ	1656.22	942		

- ა) ჩამოაყალიბეთ ჰიპოთეზები.
 ბ) თუ $\alpha = 0.05$, რა გადაწყვეტილებას მიიღებთ?
 გ) საჭიროა თუ არა ტუკის კრიტერიუმის გამოყენება?
 დ) საშუალოთა რომელი წყვილები განსხვავდებიან მნიშვნელოვნად ერთმანეთისაგან?
 ამონაწერი კომპიუტერული პროგრამიდან:

Dependent Variable: IDEOI.OGY					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Between	2	85.382	42.691	25.55	0.0001
Within	940	1570.84	1.67		
Total	942	1656.22			

14. არსებობს ე. წ. საცხოვრებელი სეგრეგაციის ინდექსი. ეს ინდექსი იცვლება ნულოდან 100-მდე, სადაც უფრო დიდი რიცხვები ნიშნავს უფრო დიდ სეგრეგაციას (1990 წელს აშშ-ის ნაციონალური ინდექსი იყო 65). ცხრილში მოცემულია სეგრეგაციის ინდექსის მნიშვნელობები ქალაქების მიხედვით.

ჩრდილოეთი	ცენტრალური	სამხრეთი	დასავლეთი
ბუფალო 84	გარი 91	ნიუ ორლეანი 73	რიჩმონდი 38
ნეუარკი 83	დეტროიტი 89	ორლანდო 64	დალასი 64
ფილადელფია 83	ჩიკაგო 87	მაიამი 75	ანაჰაიმი 43
პიტსბურგი 74	მილუოკი 85	ატლანტა 72	ჰიუსტონი 59

შესაბამისი ANOVA ცხრილი:

ცვალებადობის წყარო	SS	df	MS	F
ჯგუფთა შორის	3107.0	3		
ჯგუფთა შიგნით	622.0	12		
სულ				

- ა) დაასრულეთ ANOVA ცხრილი.
 ბ) ჩამოაყალიბეთ ჰიპოთეზები.
 გ) თუ $\alpha = 0.05$, რა გადაწყვეტილებას მიიღებთ?
 დ) რომელი საშუალოები განსხვავდებიან ერთმანეთისაგან მნიშვნელოვნად?

15. ცხრა პირველკურსელი სტუდენტი გადაანალიზეს იაპონური ენის სწავლების ჯგუფებში. პირველი ჯგუფის სტუდენტებს ჰქონდათ მაღალი ქულები უნარებში, მეორე ჯგუფის სტუდენტებს – საშუალო, მესამე ჯგუფის სტუდენტებს – დაბალი. პირველ „კვიზში“ მიღებული ქულები (მაქსიმალური ქულა 10):

პირველი ჯგუფი	მეორე ჯგუფი	მესამე ჯგუფი
10	6	3
9	4	2
11	5	4

- ა) ჩამოაყალიბეთ ჰიპოთეზები.
 ბ) რა გადაწყვეტილებას მიიღებთ? $\alpha = 0.05$.
 გ) მოქმედებს თუ არა უნარებში მიღებული ქულა „კვიზის“ ქულებზე?

16. თავისი ნაკეთობების რეკლამისათვის ფირმა უშვებდა სამ განსხვავებულ კატალოგს. გაყიდვების მენეჯერს დაავალეს გაერკვია, რომელი ტიპის კატალოგი იზიდავს მეტ მყიდველს. მენეჯერს ესმოდა, რომ განსხვავება კლიენტებს შორის დამალავდა სხვაობას კატალოგებს შორის, ამიტომ მიმართა ცდისპირთა შიდა-დისპერსიულ ანალიზს. შეარჩიეს 100 კლიენტი და ყოველ მათგანს გაუგზავნეს სამ-სამი კატალოგი. შესაბამისი შესყიდვები (დოლარი) მოყვანილია ქვემოთ:

კლიენტი	კატალოგი 1	კატალოგი 2	კატალოგი 3
1	44	27	90
2	67	36	80
3	50	20	112
4	80	40	230
5	36	18	60
6	50	30	80
7	110	140	209
8	80	40	140
9	20	12	40
10	30	12	58

- ა) ჩამოაყალიბეთ ნულოვანი და ალტერნატიული ჰიპოთეზები.
 ბ) ააგეთ ANOVA ცხრილი. შეამოწმეთ ჰიპოთეზა, აიღეთ $\alpha = 0.01$.
 გ) გაარკვიეთ, რომელი კატალოგი იზიდავს მეტ კლიენტს.

17. სამი ქარხანა უშვებს მანქანის საბურავებს. შეარჩიეს 20 მანქანა და დაამონტაჟეს თითოეული ქარხნის საბურავი. 12.000 მილის გავლის შემდეგ შეამოწმეს საბურავების პროფილი. მიიღეს შემდეგი მონაცემები: $SST = 200$, $SSR = 50$, $SSB = 70$.

ა) ჩამოაყალიბეთ ჰიპოთეზები.

ბ) ააგეთ ANOVA ცხრილი. შეამოწმეთ ჰიპოთეზა, აიღეთ $\alpha = 0.05$.

გ) გაარკვიეთ, არის თუ არა განსხვავება გამოშვებულ პროდუქციას შორის.

დ) დაასაბუთეთ, რატომ მიმართეს ცდისპირთა შიდა ანალიზს ამ კვლევაში.

18. ექსტრავერტებს ახასიათებენ, როგორც ღია, სოციალურ და ხალისით აღსავსე ადამიანებს, ხოლო ინტროვერტებს – როგორც ჩაკეტილ და ნაკლებად სოციალურ ტიპებს. განსხვავებულად რეაგირებენ თუ არა ექსტრავერტები და ინტროვერტები ხმაურზე? ცდისპირები ასრულებდნენ დაეალებას: წაკითხულის შინაარსის გადმოცემას ხმადაბალი და ხმამაღალი ხმაურის ფონზე. ცდისპირების მიღებული ქულებია:

პიროვნების ტიპი (ფაქტორი A)		
	ინტროვერტული	ექსტრავერტული
	დონე A	დონე A
	10	15
	8	14
	12	12
	15	15
ხმადაბალი	10	11
B ფაქტორის B _i დონე	9	8
	14	12
	13	8
	14	13
	7	10
	12	9
	11	11
ხმაურის ფონი (ფაქტორი B)	6	13
	8	8
	11	9
	4	14
	5	13
	8	12

პიროვნების ტიპი (ფაქტორი A)		
ხმამალალი B_2 დონე	9	15
	10	11
	9	9
	10	11
	7	12
	11	10

რა დამოკიდებულება არსებობს პიროვნების ტიპისა და ხმაურის ფონის ეფექტს შორის? თუ ხდება ურთიერთქმედება, აღწერეთ ის.

19. დაასრულეთ ANOVA ცხრილები და უპასუხეთ კითხვებს.

ცვალებადობის წყარო	SS	df	MS	F
დავალების ტიპი (ფაქტორი A)	-	3	5.0	-
ხმაურის დონე (ფაქტორი B)	16.0	-	-	-
$A \cdot B$	36.0	6	-	-
ჯგუფთა შიდა	-	60	2.0	
სულ		-	-	

ცვალებადობის წყარო	SS	df	MS	F
დავალების ტიპი (ფაქტორი A)	40.0	-	10.0	
ხმაურის დონე (ფაქტორი B)	-	3	40.0	
$A \cdot B$	960.0	-	80.0	
ჯგუფთა შიდა	3200.0	160		

- A ფაქტორის რამდენი დონე იყო?
 - B ფაქტორის რამდენი დონე იყო?
 - სულ რამდენი ცდისპირი მონაწილეობდა ცდაში?
 - თუ $\alpha = 0.05$, როგორ გადაწყვეტილებებს გააკეთებთ ნულოვანი ჰიპოთეზისათვის?
20. 1980 წლის ზამთრის ოლიმპიადის შემდეგ გააანალიზეს ფიგურულ სრიალში მონაწილეთა შედეგები და მსაჯების გადაწყვეტილებები. კვლევის საგანი იყო, რამდენად „პატრიოტული“ ან „ნეიტრალური“ იყვნენ მსაჯები თავისი თანამემამულე კონკურსანტების მიმართ.

ცვალებადობის წყარო	SS	df	MS	F
კონკურსანტი (A)	40.22	10	4.02	205.9
მსაჯი (B)	0.23	8	0.03	1.45
ურთიერთქმედება	1.58	80	0.02	
შიდა ცვალებადობა	5.80	297	0.02	
სულ	47.83	395		

- ა) რამდენი იყო კონკურსანტი?
 ბ) რამდენი იყო მსაჯი?
 გ) არის თუ არა განსხვავება კონკურსანტებს შორის?
 დ) არის თუ არა განსხვავება მსაჯებს შორის?
 ე) თუ ურთიერთქმედება მნიშვნელოვანია, ნიშნავს თუ არა ეს „პატრიოტულ“ ან „ნეიტრალურ“ გადაწყვეტილებას თანამემამულის მიმართ?
21. 8 მაგისტრანტის ნაშრომი შემოწმებულია იყო 4 ექსპერტის მიერ. მიღებული იყო შემდეგი შეფასებები

მაგისტრანტი	ექსპერტი 1	ექსპერტი 2	ექსპერტი 3	ექსპერტი 4
1	16	7	16	21
2	12	3	13	26
3	10	9	17	23
4	9	10	20	25
5	10	5	22	30
6	13	8	17	24
7	11	7	16	22
8	14	12	23	21

- ა) გამოთვალეთ SST.
 ბ) გამოთვალეთ SSR.
 გ) გამოთვალეთ SSB.
 დ) გამოთვალეთ SSE.
 ე) შეამოწმეთ ძირითადი ჰიპოტეზა: არსებობს თუ არა განსხვავება ოთხი ექსპერტის შეფასებაში.
 ვ) შეამოწმეთ მეორადი ჰიპოტეზა: არსებობს თუ არა განსხვავება მაგისტრანტების საშუალო შეფასებას შორის. ისარგებლეთ $\alpha = 0.05$ დონით.

არაპარამეტრული სტატისტიკური კრიტერიუმები

მე-8 თავიდან ამ თავამდე, პოპულაციის პარამეტრების შესახებ დასკვნების გასაკეთებლად, განიხილებოდა პარამეტრული სტატისტიკური კრიტერიუმები.

1-კრიტერიუმის და დისპერსიული ანალიზის პარამეტრული კრიტერიუმების გამოყენებისათვის დაშვებული იყო, რომ:

1. პოპულაციებში, რომლიდანაც აღებულია შერჩევა, მონაცემები ნორმალურადაა განაწილებული;
2. სხვადასხვა პოპულაციაში მონაცემების დისპერსიები ერთმანეთის ტოლია;
3. აზრი აქვს მიღებული მონაცემების საშუალოს გამოთვლას (პარამეტრულ კრიტერიუმს ვერ გამოიყენებთ, როცა მონაცემები სახელდების სკალაზეა).

არაპარამეტრული სტატისტიკური კრიტერიუმები ამონებენ ჰიპოთეზებს, რომლებიც არ ეხება პოპულაციის პარამეტრებს. მათ უფრო ნაკლები დაშვებები აქვთ პოპულაციის განაწილების შესახებ და ამიტომ მათ ზოგჯერ განაწილებისგან თავისუფალ კრიტერიუმებსაც უწოდებენ.

ხი-კვადრატ (χ^2) თანხმობის კრიტერიუმი

ვთქვათ, გვაქვს 100 სტუდენტის სია, ზოგი მათგანი პირველ კურსზე სწავლობს, ზოგი მეორე, მესამე ან მეოთხეკურსელია. შერჩევის მოცულობა $n = 100$ და ჩვენ ვანალიზებთ მათ კურსების მიხედვით, $k = 4$ კატეგორიად.

ზოგადად, თუ n მონაცემი ნაწილდება k კატეგორიად, ამ k კატეგორიაში უნდა მოხვდეს ყველა შესაძლო მონაცემი. მაშინ, თუ p_i -ით აღვნიშნავთ i -ურ კატეგორიაში მოხვედრილ მონაცემთა წილს, გვექნება $\sum p_i = 1$.

ჰიპოთეზები

შესამოწმებელია $H_0: p_1 = \bar{p}_1, p_2 = \bar{p}_2, \dots, p_k = \bar{p}_k$ ჰიპოთეზა, სადაც $\sum \bar{p}_i = 1$. $\bar{p}_i$ – პირველ კატეგორიაში მოხვედრის ალბათობაა, $\bar{p}_2$ – მეორე კატეგორიაში მოხვედრის ალბათობაა და ა. შ.

ალტერნატიული ჰიპოთეზაა H_A : ნულოვანი ჰიპოთეზის რომელიმე ერთი ტოლობა მაინც არ სრულდება.

ამ ჰიპოთეზის სპეციალური სახეა $H_0: p_i = \frac{1}{k}$ – ყველა კატეგორიაში მოხვედრის ალბათობა ერთი და იგივეა.

კრიტერიუმის სტატისტიკა

გვაქვს n მონაცემი. აღვნიშნოთ o_i -ით i -ურ კატეგორიაში მოხვედრილ მონაცემთა რაოდენობა (სიხშირე). o_i – დაკვირვებული სიხშირეა (observed frequency). ცხადია, რადგან გვაქვს n მონაცემი, დაკვირვებულ სიხშირეთა ჯამი იქნება n ($\sum o_i = n$).

დაკვირვებულ სიხშირეებთან ერთად განიხილავენ ე. წ. მოსალოდნელ სიხშირეებს (expected frequency) – e_i , პირობაში, რომ სამართლიანია H_0 ჰიპოთეზა.

p_i არის პირველ კატეგორიაში მოხვედრის ალბათობა. პირველ კატეგორიაში მოხვედრილ ინდივიდთა რაოდენობა ბინომურად არის განაწილებული პარამეტრებით n და p_i (წარმატების ალბათობა) და საშუალოთი $n \cdot p_i$. ამ სიდიდეს პირველი კატეგორიის შესაბამისი მოსალოდნელი სიხშირე ეწოდება (იგულისხმება, რომ სრულდება ნულოვანი ჰიპოთეზა).

i -ური კატეგორიის შესაბამისი მოსალოდნელი სიხშირეა $e_i = n \cdot p_i$, სადაც p_i არის i -ურ კატეგორიაში მოხვედრის ალბათობა:

$$\sum p_i = 1. \quad \sum o_i = n.$$

თუ H_0 ჰიპოთეზა სამართლიანია, მოსალოდნელია, რომ დაკვირვებული სიხშირეები ახლოს იქნებიან მოსალოდნელ სიხშირეებთან. ამიტომ გამოსათვლელია $o_i - e_i$, სხვაობები და კრიტერიუმის სტატისტიკის სახეა:

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}.$$

ამ სტატისტიკას აქვს ზი-კვადრატ განაწილება თავისუფლების ხარისხით $k-1$. ეს სტატისტიკა ყოველთვის არაუარყოფითია. როცა ყველა დაკვირვებული სიხშირე უდრის მოსალოდნელ სიხშირეებს, სტატისტიკა ნულის ტოლია. თავისუფლების ხარისხი უდრის კატეგორიების რაოდენობას k მინუს ერთი (დაკავშირებული $\sum p_i = 1$ ტოლობასთან).

უარყოფის არეები

ზი-კვადრატ განაწილების კრიტიკული მნიშვნელობები $\chi^2_{k-1, \alpha}$ მოცემული α -თვის უნდა მოვძებნოთ ზი-კვადრატ განაწილების ზედა კრიტიკული წერტილების ცხრილიდან. კრიტიკული არე მარჯვენა ცალმხრივია – $[\chi^2_{k-1, \alpha}, +\infty)$.

გადაწყვეტილება

თუ კრიტერიუმის სტატისტიკის მნიშვნელობა მეტია ან ტოლი კრიტიკულ მნიშვნელობაზე $\chi^2_{k-1, \alpha}$, მაშინ ნულოვანი ჰიპოთეზა უნდა უარყვოდ. წინააღმდეგ შემთხვევაში ამის საფუძველი არ გაგვანია.

შენიშვნა. სწორი გადანყვეტილების მისაღებად საჭიროა მოსალოდნელი სიხშირეების, სულ მცირე, 80% იყოს 5-ზე მეტი. თუ მოსალოდნელი სიხშირეების 20%-ზე მეტი ნაკლებია 5-ზე, საჭიროა კატეგორიების გაერთიანება (იხ. მაგალითი 12.4).

მაგალითი 12.1. მოთამაშე ეჭვობს, რომ მისი კამათელი არ არის „სწორი“. მან გააგორა კამათელი 120-ჯერ და მიიღო შემდეგი შედეგი:

ქულა	1	2	3	4	5	6
სიხშირე	23	18	24	19	25	11

შევამოწმოთ, არის თუ არა კამათელი „ნესიერი“. სანდოობის დონე იყოს $\alpha = 0.05$.

ამოხსნა: H_0 – ნულოვანი ჰიპოთეზა ნიშნავს, რომ $p_i = \frac{1}{n}$ ყველა i -თვის, ანუ კამათელის ყველა შედეგი არის ტოლშესაძლებელი. ალტერნატიული ჰიპოთეზის თანახმად, ყველა p_i ერთმანეთის ტოლი არ არის. ყველა მოსალოდნელი სიხშირე არის:

$$e_i = n \cdot p_i = 120 \cdot \frac{1}{6} = 20.$$

კრიტერიუმის სტატისტიკის მნიშვნელობა იქნება:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} = \frac{(23-20)^2}{20} + \frac{(18-20)^2}{20} + \dots + \frac{(11-20)^2}{20} = 6.8.$$

ამ სტატისტიკას აქვს ხი-კვადრატ განაწილება $k-1 = 6-1 = 5$ თავისუფლების ხარისხით. კრიტიკული მნიშვნელობა, როცა $\alpha = 0.05$, არის $\chi^2_{5,0.05} = 11.070$ და ჩვენ ვერ უარვყოფთ ნულოვან ჰიპოთეზას.

თუ შევამჩნევთ, რომ კენტი ქულები უფრო ხშირად ჯდება, ვიდრე ლუნები და შევამოწმებთ ჰიპოთეზას პროპორციების ტოლობის შესახებ, მოთამაშის ეჭვები შეიძლება გამართლდეს. შეამოწმეთ!

ხი-კვადრატ თანხმობის კრიტერიუმში შეიძლება გამოვიყენოთ შემდეგი ამოცანების გადასაწყვეტად: არის თუ არა მონაცემები მიღებული ბინომური პოპულაციიდან $p = 0.5$ პარამეტრით? ან, არის თუ არა პოპულაცია ნორმალურად განაწილებული? და სხვა.

დამოუკიდებლობის ხი-კვადრატ კრიტერიუმი

ორ სიდიდეს შორის კავშირის არსებობის დადგენა სხვადასხვაგვარად შეიძლება: თუ მონაცემები ინტერვალის სკალაზეა გაზომილი, ნრფივი კავშირის დასადგენად, გამოვიყენებთ პირსონის კორელაციის კოეფიციენტს; თუ რიგის

სკალაზე – სპირმენის კორელაციის კოეფიციენტს. ორი სახელდების სკალაზე გაზომილი სიდიდეების კავშირის დადგენა ხდება ზი-კვადრატ დამოუკიდებლობის კრიტერიუმით. მონაცემები, ამ შემთხვევაში, წარმოდგენილია ე. წ. შეუღლების ცხრილის სახით.

მაგალითი 12.2. დამოკიდებულია თუ არა სახლის გაყიდვის ვადა სახლის ღირებულებაზე. შეგროვილი იყო შემდეგი მონაცემები, რომელიც წარმოდგენილია შეუღლების ცხრილის სახით (იხ. ცხრილი 12.1).

ცხრილი 12.1

გაყიდვის ფასი	რამდენ დღეში გაიყიდა სახლი		
	60 დღე ან ნაკლები	61 დღე ან მეტი	სულ
75 000 დოლარზე ნაკლები	18	12	30
75 000-დან 125 000-მდე	14	31	45
125 000 დოლარზე მეტი	4	11	15
სულ	36	54	90

H_0 : ალბულო ორი ცვლადი დამოუკიდებელია,

H_1 : ალბულო ორი ცვლადი დამოკიდებულია.

რას ნიშნავს ამ ამოცანაში, რომ ცვლადები დამოკიდებულია? ეს ნიშნავს, რომ გაყიდვამდე უკვე ცნობილია, რომ ფასზეა დამოკიდებული, სწრაფად გაიყიდება სახლი თუ არა. დამოუკიდებლობა ნიშნავს, რომ ფასზე არ არის დამოკიდებული სახლის გაყიდვის ვადები.

ზოგად შემთხვევაში შეუღლების ცხრილს აქვს შემდეგი სახე:

ცხრილი 12.2

ცვლადი B	ცვლადი A				სულ
	A_1	A_2	...	A_c	
B_1	o_{11}	o_{12}	...	o_{1c}	r_1
B_2	o_{21}	o_{22}	...	o_{2c}	r_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_r	o_{r1}	o_{r2}	...	o_{rc}	r_r
სულ	c_1	c_2	...	c_c	n

ij არის ij უჯრაში მონაცემთა რაოდენობა, სადაც i სტრიქონის ნომერია, j კი სვეტის ნომერი; r_i არის i -ურ სტრიქონში მონაცემთა მთლიანი რაოდენობა, c_j არის j -ურ სვეტში მონაცემთა მთლიანი რაოდენობა, n კი მონაცემთა სრული რაოდენობაა, $n = \sum r_i = \sum c_j = \sum o_{ij}$.

ჰიპოთეზები

ნულოვანი ჰიპოთეზა ნიშნავს: ცვლადები დამოუკიდებელია. ალტერნატიული ჰიპოთეზა ნიშნავს, რომ ცვლადები დამოკიდებულია.

კრიტერიუმის სტატისტიკა

ნულოვანი ჰიპოთეზის თანახმად, ცვლადები დამოუკიდებელია. ეს კი ნიშნავს, რომ A, B , უჯრედში მოხვედრის ალბათობა უდრის i -ურ სტრიქონში და j -ურ სვეტში მოხვედრის ალბათობათა ნამრავლს:

$$P(A, B_i) = P(A_i)P(B_j) = \frac{r_i}{n} \cdot \frac{c_j}{n} = \frac{r_i \cdot c_j}{n^2}.$$

ij უჯრაში მოხვედრილი მონაცემების მოსალოდნელი e_{ij} სიხშირე უდრის n -ჯერ ამ უჯრაში მოხვედრის ალბათობას

$$n \cdot \frac{r_i \cdot c_j}{n^2} = \frac{r_i \cdot c_j}{n}.$$

თუ ნულოვანი ჰიპოთეზა სრულდება, მაშინ მოსალოდნელი სიხშირეები გამოითვლება ფორმულით:

$$e_{ij} = \frac{r_i \cdot c_j}{n}.$$

კრიტერიუმის სტატისტიკა ეფუძნება მოსაზრებას, რომ, თუ ნულოვანი ჰიპოთეზა სამართლიანია, დაკვირვებული სიხშირეები ახლოს იქნებიან მოსალოდნელ სიხშირეებთან, ხოლო, თუ ალტერნატიული ჰიპოთეზა სწორია, მაშინ სხვაობა დაკვირვებულ და მოსალოდნელ სიხშირეებს შორის დიდია.

კრიტერიუმის სტატისტიკას აქვს სახე:

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

და ამ სტატისტიკას აქვს ზი-კვადრატ განაწილება თავისუფლების ხარისხით $(r-1) \cdot (c-1)$.

უარყოფის არე

თუ ყველა დაკვირვებული სიხშირე უდრის თავის შესაბამის მოსალოდნელ სიხშირეს, სხვაობა მათ შორის ნულის ტოლია და კრიტერიუმის სტატისტიკაც ნუ-

ლის ტოლია. რაც უფრო დიდია განსხვავება დაკვირვებულ და მოსალოდნელ სიხშირეებს შორის, მით უფრო დიდ მნიშვნელობას მიიღებს კრიტერიუმის სტატისტიკა. კრიტიკული მნიშვნელობა χ^2_{α} უნდა ვიპოვოთ ხი-კვადრატ განაწილების კრიტიკული ნერტილების ცხრილიდან. უარყოფის არეა – $[\chi^2_{\alpha}, +\infty)$. კრიტერიუმი მარჯვენა ცალმხრივია.

გადაწყვეტილება

თუ კრიტერიუმის სტატისტიკის მნიშვნელობა მეტია კრიტიკულ მნიშვნელობაზე, ნულოვან ჰიპოთეზას უარყოფთ, α მნიშვნელოვნების დონით, ალტერნატივის სასარგებლოდ. წინააღმდეგ შემთხვევაში ნულოვანი ჰიპოთეზის უარყოფის საფუძველი არ გაგვაჩნია.

მაგალითი 12.3. ცხრილში მოყვანილია პოლიკლინიკაში განკურნების შედეგები სქესის მიხედვით (ერთი და იგივე დაავადება).

ცხრილი 12.3

	აბებით მკურნალობა	კონსულ- ტაციები	მკურნალობის გარეშე	სულ
მამრობითი	10/6	8/9	6/9	24
მდედრობითი	6/10	16/15	18/15	40
სულ	16	24	24	64

კრიტერიუმის სტატისტიკის მნიშვნელობაა:

$$\chi^2 = \frac{(10-6)^2}{6} + \frac{(8-9)^2}{9} + \frac{(6-9)^2}{9} + \frac{(6-10)^2}{10} + \frac{(16-15)^2}{15} + \frac{(18-15)^2}{15} = 6.044.$$

თუ $\alpha = 0.05$, $(r-1) \cdot (c-1) = 1 \cdot 2 = 2$ თავისუფლების ხარისხის შესაბამისი კრიტიკული მნიშვნელობა არის $\chi^2_{0.05} = 5.991$ და ვინაიდან $6.044 > 5.991$, ნულოვან ჰიპოთეზას უარყოფთ. სქესსა და განკურნების შედეგებს შორის არსებობს კავშირი. აბებით მკურნალობა უფრო შედეგიანია კაცებისათვის. ქალებს განკურნებისათვის უფრო ხშირად მკურნალობა არ სჭირდებათ.

ჰიპოთეზათა შემოწმება 2×2 შეუღლების ცხრილის შემთხვევაში

2×2 ცხრილის შემთხვევაში, კორექციის გათვალისწინებით, კრიტერიუმის სტატისტიკა გამოითვლება ფორმულით:

$$\chi^2 = \sum \frac{(|o_{ij} - e_{ij}| - 0.5)^2}{e_{ij}}$$

და ამ სტატისტიკას აქვს ხი-კვადრატ განაწილება თავისუფლების ხარისხით 1.

ამ შემთხვევაში ყველა უჯრისთვის $|o_{ij} - e_{ij}|$ სხვაობა ერთი და იგივეა. ამ სხვაობას უნდა გამოვაკლოთ 0.5, მერე ავიყვანოთ კვადრატში, გავყოთ მოსალოდნელ სიხშირეზე და მიღებული რიცხვები შევკრიბოთ.

მაგალითი 12.4. მონაცემები შეგროვებული იყო მცირე და საშუალო ბიზნესის მფლობელებისგან, მათი მომავალი საქმიანობის პროგნოზის შესახებ:

ცხრილი 12.4

	მფლობელის პროგნოზი			
	გაუმჯობესდება	იგივე დარჩება	გაუარესდება	სულ
მცირე ბიზნესის მფლობელი	6	18	8	28
საშუალო ბიზნესის მფლობელი	3	12	13	23
სულ	9	30	21	60

არის თუ არა ბიზნესის ზომა და მფლობელის რწმენა ერთმანეთთან კავშირში?

ნულოვანი ჰიპოთეზა: მფლობელის პროგნოზი და ფირმის ზომა ერთმანეთისაგან დამოუკიდებელია, ისინი დაკავშირებულია ერთმანეთთან – ალტერნატივის სანინაალმდეგოდ.

მოსალოდნელი სიხშირეები:

$$e_{ij} = \frac{r_i \cdot c_j}{n},$$

$$e_{11} = \frac{28 \cdot 9}{60} = 4.2, \quad e_{12} = \frac{28 \cdot 30}{60} = 14, \quad e_{13} = \frac{21 \cdot 28}{60} = 9.8,$$

$$e_{21} = \frac{9 \cdot 32}{60} = 4.8, \quad e_{22} = \frac{30 \cdot 32}{60} = 16, \quad e_{23} = \frac{21 \cdot 32}{60} = 11.2.$$

ორი მოსალოდნელი სიხშირე ნაკლებია 5-ზე. ეს კი დაუშვებელია და ამიტომ უნდა მოვახდინოთ კატეგორიების გაერთიანება. ახალი ცხრილი მიიღებს სახეს:

ცხრილი 12.5

	მფლობელის პროგნოზი		სულ
	გაუმჯობესდება ან იგივე დარჩება	გაუარესდება	
მცირე	24/20.8	8/11.2	32
საშუალო	15/18.2	13/9.8	28
სულ	39	21	60

მივიღეთ შეუღლების 2×2 ცხრილი. კრიტერიუმის სტატისტიკის მნიშვნელობა იქნება:

$$\chi^2 = \sum \frac{(|o_{ij} - e_{ij}| - 0.5)^2}{e_{ij}} = \frac{2.7^2}{20.8} + \frac{2.7^2}{11.2} + \frac{2.7^2}{18.2} + \frac{2.7^2}{9.8} = 1.146.$$

სტატისტიკის თავისუფლების ხარისხი - $(2-1) \cdot (2-1) = 1$. ხი-კვადრატ განაწილების ცხრილიდან, კრიტიკული მნიშვნელობა, როცა $\alpha = 0.05$, არის $\chi^2_{1,0.05} = 3.841$. ვინაიდან $2.146 < 3.841$, ამიტომ ნულოვან ჰიპოთეზას 0.05 დონეზე ვერ უარვყოფთ. ჩვენ ვერ აღმოვაჩინეთ მნიშვნელოვანი კავშირი ბიზნესის ზომასა და მფლობელის აზრს შორის.

მაგალითი 12.5. მიკროსაფინანსო ბანკის მენეჯერს დაავალეს შეემუშავებინა ავტომანქანის სესხის გაცემის კრიტერიუმები. ორი წლის განმავლობაში გაცემული იყო 400 სესხი. მენეჯერმა დაახარისხა სესხები შემდეგნაირად:

კატეგორია	სიხშირე
კარგი გადახდელი (დროზე იხდის)	300
საშუალო (ავკიანებს გადახდას)	60
ცუდი (საჭიროა ზომების მიღება)	40

კლიენტებზე მოძიებულ იქნა დამატებითი ინფორმაცია, აქირავენ, თუ ყიდულობს ის სახლს. მონაცემები წარმოდგენილი იყო შეუღლების ცხრილის სახით, რადგანაც მენეჯერს აინტერესებდა არის, თუ არა მიღებული ცვლადები ერთმანეთზე დამოკიდებული. $\alpha = 0.01$.

ცხრილი 12.6

	აქირავენ	ყიდულობს	სულ
კარგი	140	160	300
საშუალო	20	40	60
ცუდი	20	20	40
სულ	180	220	400 = N

180 კლიენტი აქირაევებს სახლს, 220 – ყიდულობს. 140 „კარგი“ კლიენტი აქირაევებს სახლს და ა. შ.

H_0 : სტრიქონის და სვეტის ცვლადები დამოუკიდებელია,

H_A : სტრიქონის და სვეტის ცვლადები დამოკიდებულია.

თუ ნულოვანი ჰიპოთეზა სრულდება, მაშინ კარგი გადამხდელის ალბათობა, რომელიც აქირაევებს სახლს და კარგი გადამხდელის ალბათობა, რომელიც ყიდის სახლს, ტოლია. ეს ორი ალბათობა უდრის კარგი გადამხდელის ალბათობას ქირა-ყიდვის გარეშე.

$$p(\text{კარგი}) = \frac{\text{კარგი გადამხდელების რაოდენობა}}{\text{სესხების რაოდენობა}} = \frac{300}{400} = 0.75,$$

თუ სრულდება ნულოვანი ჰიპოთეზა

$$p(\text{კარგი/აქირაევებს}) = 0.75,$$

$$p(\text{კარგი/ყიდულობს}) = 0.75.$$

180 კაციდან, ვინც სახლს აქირაევებს, 75% ანუ 135 უნდა იყოს კარგი გადამხდელი. 220 კაციდან, ვინც სახლს ყიდულობს, 75% ანუ 165 უნდა იყოს კარგი გადამხდელი. ასეთი მსჯელობით ვიპოვით სხვა დანარჩენ მოსალოდნელ სიხშირებს. ცხრილი 12.6 მიიღებს სახეს:

ცხრილი 12.7

	აქირაევებს	ყიდულობს	სულ
კარგი	140/135	160/165	300
საშუალო	20/27	40/33	60
ცუდი	20/18	20/22	40
სულ	180	220	400 = N

თუ სრულდება ნულოვანი ჰიპოთეზა, განსხვავება დაკვირვებულ და მოსალოდნელ სიხშირებს შორის პატარაა:

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = \frac{(140-135)^2}{135} + \frac{(160-165)^2}{165} + \frac{(20-27)^2}{27} + \frac{(40-33)^2}{33} + \frac{(20-18)^2}{18} + \frac{(20-22)^2}{22} = 4.04.$$

სვეტის ცვლადს აქვს 2 დონე (აქირაევებს, ყიდულობს),

სტრიქონის ცვლადს – სამი დონე (კარგი, საშუალო, ცუდი),

თავისუფლების ხარისხი = $(2-1)(3-1) = 2$, როცა $\alpha = 0.01$, $\chi^2_{crit} = 9.21$.

რადგან კრიტერიუმის სტატისტიკის მნიშვნელობა $4.04 < 9.21$, ნულოვან ჰიპოთეზას ვერ უარვყოფთ. ცვლადები დამოუკიდებელია.

ცნობილი იყო კლიენტების შემოსავალი და შედგენილი იყო შეუღლების ცხრილი.

ცხრილი 12.8

	შემოსავალი (ათასი დოლარი)				სულ
	0-10	10-20	20-30	30-ზე მეტი	
კარგი	12/37.5	43/54.75	87/84	158/123.75	300
საშუალო	18/7.5	20/10.95	18/16.8	4/24.75	60
ცუდი	20/5	10/7.3	7/11.2	3/16.5	40
სულ	50	73	112	165	400

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = \frac{(12-37.5)^2}{37.5} + \frac{(43-54.75)^2}{54.75} + \dots + \frac{(3-16.5)^2}{16.5} = 127.72.$$

თავისუფლების ხარისხი $= (4-1)(3-1) = 6$, $\alpha = 0.01$, $\chi^2_{\text{crit}} = 16.812$.

რადგან $127.72 > 16.812$, ნულოვანი ჰიპოთეზა უნდა უარყვოდ. კლიენტები მაღალი შემოსავლით უფრო კარგი გადამხდელები არიან.

როცა χ^2 სტატისტიკის მნიშვნელობა მეტია კრიტიკულ მნიშვნელობაზე, ნულოვანი ჰიპოთეზა უნდა უარყვოდ. ეს კი ნიშნავს, რომ დაკვირვებულ და მოსალოდნელ სიხშირებს შორის განსხვავება მნიშვნელოვანია. მაგრამ ეს სტატისტიკა არ გვიჩვენებს, რომელი უჯრედები აჩვენებენ გადახრას დამოუკიდებლობისგან და რომელი არა. სხვაობას $o_{ij} - e_{ij}$ ნაშთი ეწოდება.

განვიხილოთ ისევ 12.3 ცხრილი. ნაშთი $o_{11} - e_{11} = 10 - 6 = 4$ დადებითია, ნაშთი $o_{21} - e_{21} = 6 - 10 = -4$ უარყოფითია. ნაშთი დადებითია, როცა დაკვირვებული სიხშირე მეტია მოსალოდნელ სიხშირეზე და უარყოფითია, როცა პირიქით – ნაკლებია. რამდენად დიდი უნდა იყოს ნაშთი, რომ ეს მეტყველებდეს დამოკიდებულებაზე?

იმისათვის, რომ ვუპასუხოთ ამ კითხვას, უნდა გამოვთვალოთ სტატისტიკა:

$$i = \frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}(1-r_i)(1-c_j)}},$$

სადაც r_i არის i -ურ სტრიქონში მონაცემთა მთლიანი რაოდენობა, c_j არის j -ურ სვეტში მონაცემთა მთლიანი რაოდენობა.

თუ ამ სტატისტიკის მნიშვნელობა მეტია ან ნაკლებია 2-ზე, ჩავთლით, რომ ამ უჯრედის დაკვირვებული და მოსალოდნელი სიხშირეები მნიშვნელოვნად განსხვავდებიან. 12.3 მაგალითში ასეთია 4 ნაშთი: აბებით მკურნალობა, როგორც ქალებისთვის ასევე მამაკაცებისთვის; მკურნალობის გარეშე, როგორც ქალებისთვის ასევე მამაკაცებისთვის გვიჩვენებს მნიშვნელოვან გადახრას დამოუკიდებლობისგან.

თუ χ^2 სტატისტიკაზე დაყრდნობით ნულოვანი ჰიპოთეზა უარყოფილია, ცვლადებს შორის არის სტატისტიკური კავშირი. მაგრამ იმისათვის, რომ გავარკვიოთ, რამდენად ძლიერია ეს კავშირი, უნდა განვიხილოთ შანსების შეფარდება.

შანსების შეფარდება

მაგალითი 12.6. განვიხილოთ 1993 წლის მონაცემების შეუღლების ცხრილი მსხვერპლისა და დამნაშავეს რასის მიხედვით (ერთი მსხვერპლი, ერთი დამნაშავე)

ცხრილი 12.9

დამნაშავეს რასა	მსხვერპლის რასა		სულ
	თეთრი	შავი	
თეთრი	4686	304	4990
შავი	849	5393	6242

შემოვიღოთ აღნიშვნები:

p_1 - ალბათობა იმისა, რომ მსხვერპლი თეთრია, როცა დამნაშავე თეთრია (განიხილება როგორც წარმატების ალბათობა);

$1 - p_1$ - შავი მსხვერპლის ალბათობა, როცა დამნაშავე თეთრია (მარცხის ალბათობა),

p_2 - ალბათობა იმისა, რომ მსხვერპლი თეთრია, როცა დამნაშავე შავია (წარმატების ალბათობა),

$1 - p_2$ - შავი მსხვერპლის ალბათობა, როცა დამნაშავე შავია (მარცხის ალბათობა).

წარმატების შანსი განიმარტება როგორც შეფარდება

$$\text{შანსი} = \frac{p}{1-p} = \frac{\text{წარმატების ალბათობა}}{\text{მარცხის ალბათობა}}.$$

თეთრი დამნაშავეისათვის თეთრი მსხვერპლის პროპორციაა $\frac{4686}{4990} = 0.939$ და

შავი მსხვერპლის პროპორციაა $\frac{304}{4990} = 0.061$.

თეთრი დამნაშავეისათვის შანსი, რომ მსხვერპლი იქნება თეთრი, არის

$$\frac{0.939}{0.061} = 15.4.$$

სიდიდე:

$$\frac{\frac{4686}{4990}}{\frac{304}{4990}} = \frac{4686}{304} = 15.4$$

ნიშნავს, რომ თეთრი დამნაშავეისათვის ყოველ ერთ შავ მსხვერპლზე მოდის 15.4 თეთრი.

შავი დამნაშავეისათვის შანსი, რომ მსხვერპლი იქნება თეთრი, არის $\frac{849}{5393} = 0.157$. ეს კი ნიშნავს, რომ ერთ შავ მსხვერპლზე მოდის 0.157 თეთრი.

ანალოგიურად, რადგან $\frac{5393}{849} = \frac{1}{0.157} = 6.4$, შავი დამნაშავეისთვის ერთ თეთრ მსხვერპლზე მოდის 6.4 შავი.

შანსების შეფარდება:

2×2 ცხრილში ორი სტრიქონის შანსების შეფარდებას ეწოდება შანსების შეფარდება

$$\theta = \frac{\text{შანსი თეთრი დამნაშავეისთვის}}{\text{შანსი შავი დამნაშავეისთვის}} = \frac{15.4}{0.157} = 97.9.$$

თეთრი მსხვერპლის შანსი თეთრი დამნაშავეებისთვის 97.9-ჯერ მეტია, ვიდრე თეთრი მსხვერპლის შანსი შავი დამნაშავეებისთვის.

$\theta = 97.9$ არ ნიშნავს, რომ p_1 ალბათობა 97.9-ჯერ მეტია p_2 -ზე. $\theta = 97.9$ ნიშნავს, რომ პირველ სტრიქონში შანსი 97.9-ჯერ მეტია მეორე სტრიქონის შანსზე.

შანსების შეფარდების თვისებები 2×2 შეუღლების ცხრილებისათვის

1. $\theta = \frac{o_{11} \cdot o_{22}}{o_{21} \cdot o_{12}};$
2. θ -მ შეიძლება მიიღოს ნებისმიერი დადებითი მნიშვნელობა;
3. როცა $p_1 = p_2$ და ორივე სტრიქონის შანსებიც ტოლია, მაშინ ცვლადები დამოუკიდებელია. $\theta = 1$, როცა ცვლადები დამოუკიდებელია;
4. რაც უფრო ძლიერია კავშირი ორ ცვლადს შორის, მით უფრო განსხვავდება θ ერთისაგან. როცა $\theta > 1$, წარმატების შანსი პირველ სტრიქონში მეტია წარმატების შანსზე მეორე სტრიქონში. როცა $\theta < 1$, წარმატების შანსი პირველ სტრიქონში ნაკლებია წარმატების შანსზე მეორე სტრიქონში.

მან-უიტნის კრიტერიუმი

მაგალითი 12.7. საგზაო სამსახურის მენეჯერის ეჭვით, ქალაქის შიგნით გზების გასარემონტებლად ნაკლებ ასფალტს ხარჯავენ, ვიდრე გზებზე ქალაქის გარეთ. შეგროვილი იყო 12 კვირის მონაცემები (იარდი ასფალტი ერთ მილზე).

ცხრილი 12.10

ქალაქის შიგნით		ქალაქის გარეთ	
	რანგი		რანგი
460	2	600	6
830	16	652	9
720	12	603	7
930	20	594	5
500	4	1402	23
620	8	1111	21
703	11	902	18
407	1	700	10
1521	24	827	15
900	17	490	3
750	13	904	19
800	14	140	22
	რანგების ჯამი		რანგების ჯამი
	142		158

მიუხედავად იმისა, რომ მონაცემები ინტერვალის სკალეზეა გაზომილი, მენეჯერმა არ გამოიყენა t -კრიტერიუმი (ორი დამოუკიდებელი შერჩევისათვის), რადგანაც გრძნობდა, რომ მონაცემებს არ ექნებოდათ ნორმალური განაწილება. ასეთი ტიპის ამოცანებისათვის არსებობს მან-უიტნის არაპარამეტრული სტატისტიკური კრიტერიუმი ცდისპირთა შორის სქემაში. ეს კრიტერიუმი არის ორი დამოუკიდებელი ჯგუფისათვის t -კრიტერიუმის არაპარამეტრული ანალოგი.

ჰიპოთეზები

ნულოვანი ჰიპოთეზა მან-უიტნის U ტესტში:

H_0 : ორივე შერჩევა აღებულია ერთი და იმავე პოპულაციიდან,

H_A : შერჩევები აღებულია სხვადასხვა პოპულაციიდან (პოპულაციები განსხვავდებიან მდებარეობის მიმართებაში).

ჩვენი ამოცანის ჰიპოთეზები:

H_0 : მოხმარებული ასფალტის მონაცემებს ქალაქის შიგნით და ქალაქის გარეთ აქვთ ერთი და იგივე განაწილება,

H_A : ქალაქის გარეთ აღინიშნება ასფალტის მეტი მოხმარების ტენდენცია, ვიდრე ქალაქის შიგნით.
ავილთ $\alpha = 0.05$.

კრიტერიუმის სტატისტიკა

იმისათვის, რომ გამოვიყენოთ მან-უიტნის კრიტერიუმი, ორივე შერჩევის მონაცემები უნდა გავაერთიანოთ და მივუწეროთ რანგები, როგორც ნაჩვენებია 12.10 ცხრილში. მან-უიტნის კრიტერიუმის ლოგიკა მდგომარეობს იმაში, რომ, თუ ერთი ჯგუფის რანგების ჯამი მნიშვნელოვნად განსხვავდება მეორე ჯგუფის რანგების ჯამისაგან, შეგვიძლია დავასკვნათ, რომ პოპულაციის მდებარეობის პარამეტრები განსხვავებულია.

გამოსათვლელი გვაქვს U სტატისტიკა ორივე შერჩევისთვის.

$$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} \text{ მინუს პირველი შერჩევის რანგების ჯამი,}$$

$$U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} \text{ მინუს მეორე შერჩევის რანგების ჯამი,}$$

სადაც n_1 და n_2 პირველი და მეორე შერჩევის მოცულობებია. შერჩევის რანგების ჯამი მოცემულია ცხრილში 12.10.

$$U_1 = 12(12) + \frac{12(13)}{2} - 142 = 80,$$

$$U_2 = 12(12) + \frac{12(13)}{2} - 158 = 64.$$

მიაქციეთ ყურადღება, რომ $U_1 + U_2 = n_1 n_2$.

კრიტერიუმის სტატისტიკა U = მინიმუმი U_1 -სა და U_2 -ს შორის,
 U = უმცირესი 80-სა და 64-ს შორის = 64.

უარყოფის არეები

თუ ავირჩევთ ორმხრივ კრიტერიუმს და კრიტერიუმის სტატისტიკის მნიშვნელობა იქნება ნაკლები მან-უიტნის კრიტიკული წერტილების ცხრილში ნაპოვნ კრიტიკულ მნიშვნელობაზე, მაშინ ნულოვან ჰიპოთეზას უარვყოფთ.

ჩვენს მაგალითში, 0.05 მნიშვნელოვნების დონისათვის, $n_1 = 12$ და $n_2 = 12$ მოცულობების შესაბამისი U განაწილების კრიტიკული მნიშვნელობა იქნება $U_{crit} = 37$. უარყოფის არეა - $U \leq U_{crit}$.

გადანყვეტილება

კრიტერიუმის სტატისტიკის მნიშვნელობა - 64 მეტია კრიტიკულ მნიშვნელობაზე - 37, ამიტომ ნულოვან ჰიპოთეზას ვერ უარვყოფთ, ვერ ვიტყვი, რომ ქალაქგარეთ ასფალტი იფლანგება. მიაქციეთ ყურადღება, რომ ამ კრიტერიუმით

ნულოვანი ჰიპოთეზის უარყოფა ხდება მაშინ, თუ გამოთვლილი კრიტერიუმის სტატისტიკის მნიშვნელობა ნაკლებია ან ტოლი კრიტიკულ მნიშვნელობაზე.

უილკოქსონის ნიშნის რანგების კრიტერიუმი

უილკოქსონის ნიშნის რანგების კრიტერიუმი არის არაპარამეტრული კრიტერიუმი ცდის პირებს შიგნით სქემისთვის. ის, დაწყვილებული მონაცემებისათვის t -კრიტერიუმის არაპარამეტრული ანალოგია.

მაგალითი 12.8. ორი დღის განმავლობაში კაფეში აკვირდებოდნენ ერთი და იმავე 11 კლიენტის სადილობას. პირველ დღეს უკრავდა ნელი მუსიკა, მეორე დღეს – სწრაფი. ჩაინერეს თითოეული ცდისპირის მიერ ნუთში მიღებული ლუკმების რაოდენობა:

ცხრილი 12.11

ცდისპირი	მუსიკის ტემპი		d	$ d $	$ d $ -ს რანგი	ნიშნის რანგი
	სწრაფი	ნელი				
1	3	4	-1	1	2	2 ₋
2	4	6	-2	2	4.5	4.5 ₋
3	2	3	-1	1	2	2 ₋
4	3	3	0	0		
5	3	8	-5	5	9	9 ₋
6	4	2	+2	2	4.5	4.5 ₊
7	1	4	-3	3	6.5	6.5 ₋
8	5	4	+1	1	2	2 ₊
9	3	6	-3	3	6.5	6.5 ₋
10	5	9	-4	4	8	8 ₋
11	2	8	-6	6	10	10 ₋

ჰიპოთეზები

H_0 : ერთმანეთთან დაკავშირებული პირველი და მეორე დღის მონაცემების განაწილებები იდენტურია.

H_1 : ერთმანეთთან დაკავშირებული პირველი და მეორე დღის მონაცემების განაწილებები იდენტური არ არის.

კრიტერიუმის სტატისტიკა

რანგების ჯამებს შორის უმცირესი ჯამი

$$T = \text{უმცირესი } \sum R_- \text{ და } \sum R_+ \text{ შორის,}$$

სადაც $\sum R_-$ უარყოფითი სხვაობების შესაბამისი რანგების ჯამია, ხოლო $\sum R_+$ – დადებითი სხვაობების შესაბამისი რანგების ჯამია.

12.11 ცხრილიდან

$$\sum R_- = 2+4.5+2+9+6.5+6.5+8+10=48.5, \sum R_+ = 4.5+2=6.5,$$

ე. ი.

$$T = \text{უმცირესი } \sum R_- \text{ -სა და } \sum R_+ \text{ -ს შორის} = 6.5.$$

უარყოფის არე

უარყოფის არე არის კრიტერიუმის სტატისტიკის ის მნიშვნელობები, რომლებიც ნაკლებია ან ტოლია კრიტიკულ მნიშვნელობაზე.

კრიტიკული მნიშვნელობები დამოკიდებულია მნიშვნელოვნების α დონეზე და ნულისგან განსხვავებული სხვაობების მქონე წყვილების რაოდენობაზე.

T -ს კრიტიკული მნიშვნელობები მოცემულია უილკოქსონის ნიშნიანი რანგების კრიტერიუმისათვის T -ს კრიტიკული მნიშვნელობების ცხრილში. როცა $\alpha = 0.05$ და ნულისგან განსხვავებული სხვაობების რაოდენობა 10-ის ტოლია, კრიტიკული მნიშვნელობა უდრის 8-ს. მიაქციეთ ყურადღება, რომ ამ ტესტში ნულოვან ჰიპოთეზას უარყოფენ, როცა გამოთვლილი სტატისტიკის მნიშვნელობა ნაკლებია ან ტოლი ცხრილში ნაპოვნ კრიტიკულ მნიშვნელობაზე.

გადაწყვეტილება

კრიტერიუმის სტატისტიკის მნიშვნელობა – 6.5 ნაკლებია კრიტიკულ მნიშვნელობაზე – 8. ამიტომ ნულოვანი ჰიპოთეზა უნდა უარყოფილი. მიღებული ლუკმების რაოდენობის განაწილება, მუსიკის სხვადასხვა ტემპის პირობებში, ერთმანეთისგან განსხვავდება. ნელ ტემპთან შედარებით, სწრაფი ტემპის დროს მეტ ლუკმას იღებენ.

კრუსკალ-უოლისის კრიტერიუმი

ერთფაქტორიანი დისპერსიული ანალიზი გამოიყენება ორზე მეტი საშუალოს შესადარებლად. ამ ანალიზის გამოყენება შეიძლება, თუ შერჩევები ალებულია ნორმალური პოპულაციებიდან და პოპულაციების დისპერსიები ტოლია.

კრუსკალ-უოლისის კრიტერიუმი არის ერთფაქტორიანი ANOVA კრიტერიუმების არაპარამეტრული ანალოგი. ამ კრიტერიუმის გამოყენებისას არ არის აუცილებელი ნორმალურობის მოთხოვნა.

მაგალითი 12.9. შესამონმებელი იყო სამი ტიპის ელექტროჯაგრის მუშაობის ვადები (საათი). $\alpha = 0.05$.

ბრენდი 1	40	37	51	20	46	93	27	25	48
ბრენდი 2	69	113	217	92	65	49	122	117	105
ბრენდი 3	5	14	10	17	36	52	13	68	141

ჰიპოთეზები

H_0 : სამი ბრენდის მუშაობის ვადა არ განსხვავდება ერთმანეთისგან.

H_A : სამი ბრენდის მუშაობის ვადა განსხვავებულია.

კრიტერიუმის სტატისტიკა

გავაერთიანოთ სამივე შერჩევის მონაცემები და მოვახდინოთ მათი რანჟირება:

	რანგი									რანგების ჯამი
ბრენდი 1	11	10	15	6	12	21	8	7	13	103
ბრენდი 2	19	23	27	20	17	14	25	24	22	191
ბრენდი 3	1	4	2	5	9	16	3	18	26	84

კრიტერიუმის სტატისტიკას აქვს სახე:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1),$$

სადაც N მონაცემთა სრული რაოდენობაა, k შერჩევების რაოდენობაა, R_i არის i -ური შერჩევის რანგების ჯამი, n_i არის i -ური შერჩევის მოცულობა.

H სტატისტიკას აქვს ზი-კვადრატ განაწილება თავისუფლების ხარისხით $k-1$.

უარყოფის არე – კრიტერიუმის სტატისტიკის ის მნიშვნელობებია, რომლებიც მეტია ან ტოლი კრიტიკულ მნიშვნელობაზე $\chi_{k-1, \alpha}^2$.

გადაწყვეტილება

თუ კრიტერიუმის სტატისტიკის მნიშვნელობა მეტია კრიტიკულ მნიშვნელობაზე, ნულოვან ჰიპოთეზას უარეყოფთ. ჩვენს მაგალითში:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) =$$

$$= \frac{12}{27(27+1)} \left[\frac{103^2}{9} + \frac{191^2}{9} + \frac{84^2}{9} \right] - 3(27+1) = 11.49$$

კრიტიკული მნიშვნელობა ზი-კვადრატ განაწილების ცხრილიდან, როცა $\alpha = 0.05$ და თავისუფლების ხარისხია $k-1 = 3-1 = 2$, უდრის 5.991-ს.

რადგან $11.49 > 5.991$ -ზე, ნულოვან ჰიპოთეზას უარეყოფთ. ელექტროჯაგრი-სების მუშაობის ვადები განსხვავებულია.

არსებობს შეზღუდვები კრუსკალ-უოლისის კრიტერიუმის გამოყენებისას: შერჩევის მოცულობები უნდა იყოს ხუთზე მეტი, შერჩევები პოპულაციიდან დამოუკიდებლად არის აღებული, პოპულაციები განსხვავდებიან მხოლოდ მდებარეობის პარამეტრებით.

თუ შერჩევებში ბევრი მონაცემი მეორდება, ანუ არსებობენ „ბმები“, H სტატისტიკა უნდა გაიყოს საკორექციო L მაშრაველზე:

$$L = 1 - \frac{\sum_{i=1}^k (t_i^3 - t_i)}{N^3 - N},$$

სადაც g ჯგუფების რაოდენობაა ბმებით, t გამეორებული მონაცემების რაოდენობა, N კი მონაცემთა სრული რაოდენობაა.

მაგალითი 12.10. დამოკიდებულია თუ არა გამოცდაზე მიღებული ქულები გამომცდელის გუნდზე? ქვემოთ მოცემულია 36 აბიტურიენტის ქულები:

გამომცდელის პირველი გუნდი	გამომცდელის მეორე გუნდი	გამომცდელის მესამე გუნდი
90	90	55
65	88	55
72	83	90
83	83	83
65	65	65
50	65	65
65	83	60
83	83	55
83	90	65
55	90	72
72	65	90
65	72	65

ჩამოვყალიბოთ ნულოვანი და ალტერნატიული ჰიპოთეზა:

H_0 : აბიტურიენტების საშუალო ქულები არ განსხვავდება ერთმანეთისაგან.

H_A : აბიტურიენტების საშუალო ქულები განსხვავდება ერთმანეთისაგან.

$$\alpha = 0.10.$$

კრუსკალ-უოლისის პროცედურის პირველი ნაბიჯია ქულების გადაყვანა რანგებში. გასათვალისწინებელია, რომ ბევრი ქულა მეორდება. რანგები შემდეგნაირია:

	პირველი გუნდი	მეორე გუნდი	მესამე გუნდი
	33.5	33.5	3.5
	12.0	30.0	3.5
	19.5	25.5	33.5
	25.5	25.5	25.5
	12.0	12.0	12.0
	1.0	12.0	12.0
	12.0	25.5	6.0
	25.5	25.5	3.5
	25.5	33.5	12.0
	3.5	33.5	19.5
	19.5	12.0	33.5
	12.0	19.5	12.0
რანგების ჯამი	201.5	288.0	176.5

ქულა 55 მეორდება 4-ჯერ; ქულა 65 – 11-ჯერ; ქულა 72 – 4-ჯერ; ქულა 83 – 6-ჯერ; ქულა 90 – 6-ჯერ. საკორექციო მამრავლი L ტოლი იქნება:

$$L = 1 - \frac{\sum (r^3 - r)}{N^3 - N} =$$

$$= 1 - \frac{(4^3 - 4) + (11^3 - 11) + (4^3 - 4) + (8^3 - 8) + (6^3 - 6)}{36^3 - 36} = 0.9538.$$

კრუსკალ-უოლისის სტატისტიკის მნიშვნელობა:

$$H = \frac{\frac{12}{N(N+1)} \sum \frac{R^2}{b} - 3(N+1)}{L} =$$

$$= \frac{\frac{12}{36 \cdot 37} \left(\frac{201.5^2}{12} + \frac{288.0^2}{12} + \frac{176.5^2}{12} \right) - 3 \cdot 37}{0.9538} = \frac{5.1400}{0.9538} = 5.389,$$

ხი-კვადრატ განაწილების ცხრილიდან, როცა $\alpha = 0.1$ და თავისუფლების ხარისხი $k - 1 = 2$, კრიტიკული მნიშვნელობა უდრის 4.605-ს. რადგან $H = 5.389 > 4.605$, ამიტომ ნულოვან ჰიპოთეზას უარეყოფთ. შემფასებელი გუნდები ერთნაირად არ აფასებენ აბიტურიენტების ცოდნას.

შეჯამება

თანხმობის ხი-კვადრატ კრიტერიუმი გამოიყენება იმისათვის, რომ დადგინდეს, რამდენად კარგად შეესაბამება კატეგორიებში მოთავსებული მონაცემები ამ კატეგორიისთვის მიღებულ მოსალოდნელ სიხშირეებს.

თანხმობის ხი-კვადრატ კრიტერიუმი გამოითვლება ფორმულით

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}.$$

დამოუკიდებლობის ხი-კვადრატ კრიტერიუმი გამოიყენება ორი ცვლადის კავშირის დასადგენად, როცა მონაცემები სახელდების სკალაზეა მიღებული. დამოუკიდებლობის ხი-კვადრატ კრიტერიუმი გამოითვლება ფორმულით:

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}.$$

მან-უიტნის U კრიტერიუმი არაპარამეტრული კრიტერიუმია ცდისპირებს შორის სტევისათვის (ორი დამოუკიდებელი ჯგუფისთვის t -კრიტერიუმის ანალოგი).

უილკოქსონის ნიშნის რანგების კრიტერიუმი არაპარამეტრული კრიტერიუმია ცდისპირებს შიდა სტევისთვის (დანჯვილებული მონაცემებისათვის t -კრიტერიუმის ანალოგი).

კრუსკალ-უოლისის კრიტერიუმი არის ერთფაქტორიანი ANOVA კრიტერიუმის არაპარამეტრული ანალოგი.

კითხვები

- ერთი თვის განმავლობაში დათვალეს დღეში მარკეტში მოსული მყიდველების რაოდენობა.

კვირის დღე	დაკვირვებული სიხშირე
ორშაბათი	1525
სამშაბათი	1711
ოთხშაბათი	1655
ხუთშაბათი	1497
პარასკევი	1603
შაბათი	1801
სულ	9792

შეამოწმეთ ჰიპოთეზა: დამოკიდებულია თუ არა მყიდველების რაოდენობა კვირის დღეზე, თუ $\alpha = 0.05$.

2. უსაფრთხოების მენეჯერის პიპოთეზით, უბედური შემთხვევების რაოდენობა დამოკიდებულია მუშის პროფესიონალიზმის დონეზე. ერთი წლის განმავლობაში შეგროვილი იყო შემდეგი მონაცემები:

პროფესიული დონე	GS-9	GS-7	GS-5 და GS-3	სულ
შემთხვევათა რაოდენობა	3	7	8	18

- ა) შეამოწმეთ, არის თუ არა მოსალოდნელი სიხშირე მეტი 5-ზე;
 ბ) შეამოწმეთ პიპოთეზა, აიღეთ $\alpha = 0.05$.
3. ფირმა ყიდის მაცივრებს თავის ოთხ ფილიალში. ერთი თვის გაყიდვის შედეგები მოცემულია ცხრილში:

ფილიალი 1	ფილიალი 2	ფილიალი 3	ფილიალი 4
305	287	112	196

- შეამოწმეთ პიპოთეზა: განსხვავდება თუ არა გაყიდვის მოცულობები ფილიალების მიხედვით? $\alpha = 0.05$.
4. სასტუმროში სამი ტიპის ნომერია: ორადგილიანი, ოთხადგილიანი და ლუქსი. მენეჯერმა დათვალა ამ ნომრებზე დაკვეთების რაოდენობა:

ერთოთახიანი	18
ოროთახიანი	21
ლუქსი	11

- შეამოწმეთ პიპოთეზა: თანაბრად არის თუ არა განაწილებული დაკვეთები ამ ნომრებზე? $\alpha = 0.05$.
5. დამოკიდებულია თუ არა მძღოლის სქესზე ავტოსაგზაო შემთხვევების რაოდენობა? სამი წლის დაკვირვების მონაცემებია:

ავტოსაგზაო შემთხვევათა რაოდენობა	მძღოლის სქესი	
	მამრობითი	მდედრობითი
0	240	160
1	80	40
2	32	18
3	11	9
სამზე მეტი	5	4

შეამოწმეთ პიპოთეზა, ისარგებლეთ $\alpha = 0.05$.

6. დამოკიდებულია თუ არა მიმდინარე ანგარიშზე გარიგების რაოდენობა კლიენტების ოჯახურ მდგომარეობაზე. აიღეთ $\alpha = 0.05$.

ოჯახური მდგომარეობა	გარიგების რაოდენობა				
	0-10	11-20	21-30	31-40	40-ზე მეტი
მარტოხელა	13	23	19	20	11
დაოჯახებული	6	15	33	45	27
გაყრილი	4	19	22	20	15
სხვა	2	11	8	5	2

7. შეგროვილი იყო ინფორმაცია ერთ-ერთი ფირმის თანამშრომლებზე: რამდენი წელი მუშაობენ ისინი ამ ფირმაში და გაფიცვის შემთხვევაში, სავარაუდოდ, რამდენი კვირა იქნებიან გაფიცულები.

მუშაობის ვადა	გაფიცვის სავარაუდო დრო		
	ერთ კვირამდე	1-4 კვირა	ოთხ კვირაზე მეტი
ერთ წლამდე	23	6	3
1-2 წელი	19	15	8
2-5 წელი	20	23	19
5-10 წელი	4	21	29
10 წელზე მეტი	2	5	18

ამ მონაცემებზე დაყრდნობით, შეგიძლიათ თუ არა დაასკვნათ, რომ გაფიცვის სავარაუდო დრო დამოკიდებულია მუშაობის სტაჟზე. აიღეთ $\alpha = 0.05$.

8. შეგროვილი იყო შემდეგი ორი დამოუკიდებელი შერჩევა:

შერჩევა 1	405	450	290	370	345	460
შერჩევა 2	300	340	400	250	270	410

შერჩევა 1	425	275	380	330	500	215
შერჩევა 2	435	390	225	210	395	315

- ა) ამ შერჩევებზე დაყრდნობით, შეგიძლიათ, თუ არა დაასკვნათ, რომ შერჩევები ალებულია ერთი და იმავე პოპულაციიდან. გამოიყენეთ $\alpha = 0.05$ და t -კრიტერიუმი. რა დაშვებები უნდა იყოს შესრულებული, რომ გამოიყენოთ ეს კრიტერიუმი?
- ბ) ვთქვათ, არ გვსურს შევამოწმოთ ა)-ში მოთხოვნილი დაშვებები. რომელი არაპარამეტრული ტესტი უნდა გამოვიყენოთ შერჩევის საშუალოების ტოლობის შესამოწმებლად? შეამოწმეთ. აიღეთ $\alpha = 0.05$.

9. 27 პროფესიონალს და 24 მომხმარებელს სთხოვეს შეეფასებინათ ახალი კომპიუტერული პროგრამა ასბალიან სკალაზე, სადაც 100 ნიშნავდა უმაღლეს კმაყოფილებას. განსხვავდება თუ არა პროფესიონალების და მომხმარებლების შეფასება? $\alpha = 0.05$.

პროფესიონალები	მომხმარებლები
$n = 27$	$n = 24$
რანგების ჯამი = 348	რანგების ჯამი = 300

შენიშვნა: როცა $n_1 > 10$ $n_2 > 10$, n -ს აქვს ნორმალური განაწილების

$$\text{საშუალო} = \frac{n_1 n_2}{2}, \quad \text{სტანდარტული გადახრა} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

10. ეფექტურია თუ არა სწრაფი კითხვის შემსწავლელი კურსები? შეამოწმებს 15 მამაკაცის კითხვის სისწრაფე კურსის დასაწყისში და კურსის დასრულების შემდეგ. დამუშავების შედეგად მიიღეს შემდეგი მონაცემები:

უარყოფითი რანგების ჯამი = 180 ;

დადებითი რანგების ჯამი = 15.

თუ უფრო მაღალი რანგი ნიშნავს წამში უფრო ბევრი სიტყვის წაკითხვას, როგორ შეაფასებთ ამ კურსებს? აიღეთ $\alpha = 0.05$.

11. შეამოწმებს სამი წამყვანი ფირმის ფეხსაცმლის ხარისხი 40-ბალიან სკალაზე, სადაც 40 ნიშნავდა უმაღლეს ხარისხს. მიიღეს შემდეგი მონაცემები:

ფირმა 1	21	25	36	35	33	23	31	32
ფირმა 2	17	15	34	22	16	19	30	20
ფირმა 3	29	38	28	27	14	26	39	36

- ა) რა დაშვებები უნდა სრულდებოდეს, რომ გამოვიყენოთ ერთფაქტორიანი დისპერსიული ანალიზი?
- ბ) ვთქვათ, არ გვსურს შევამოწმოთ ა)-ში მოთხოვნილი დაშვებები, რომელი არაპარამეტრული კრიტერიუმი უნდა გამოვიყენოთ ფირმების საშუალო რეიტინგის ტოლობის შესახებ? შეამოწმეთ ეს ჰიპოთეზა. აიღეთ $\alpha = 0.05$.
12. არსებობს თუ არა განსხვავება კალათბურთის გუნდის და ბეისბოლის ნაციონალური ლიგის მოთამაშეების საშუალო ანაზღაურებას შორის?
- ა) თუ არ სრულდება ნორმალურობის დაშვება, რა კრიტერიუმს გამოვიყენებთ?
- ბ) გამოვიყენოთ დამუშავების შედეგები და შეამოწმეთ ჰიპოთეზა, თუ $\alpha = 0.05$.

NBA	NFL	ბეისბოლი
$n = 20$	$n = 30$	$n = 40$
$R = 1710$	$R = 1100$	$R = 1340$

13. გარკვეული დეტალის დიამეტრის ზომები ნორმალურადაა განაწილებული საშუალოთი $\mu = 3$ სმ და სტანდარტული გადახრით $\sigma = 0.001$ სმ. გაზომეს 500 დეტალის დიამეტრი და მიღებულ იქნა შემდეგი მონაცემები:

ცხრილი 12.12

დიამეტრი	დაკვირვებული სიხშირე	მოსალოდნელი სიხშირე
2.995 ნაკლები	3	0
2.995-დან 2.996-მდე	4	0.015
2.996-დან 2.997-მდე	5	0.635
2.997-დან 2.998-მდე	19	10750
2.998-დან 2.999-მდე	98	67.950
2.999-დან 3.000-მდე	146	170.650
3.000-დან 3.001-მდე	124	170.650
3.001-დან 3.002-მდე	83	67950
3.002-დან 3.003-მდე	11	10750
3.003-ზე მეტი	7	0.650
სულ	500	

შევამოწმოთ ჰიპოთეზები:

H_0 : დიამეტრები ნორმალურადაა განაწილებული $\mu = 3$ სმ საშუალოთი, $\sigma = 0.001$ სმ სტანდარტული გადახრით.

H_A : ნორმალურობა არ სრულდება.

იმისათვის, რომ ვისარგებლოთ χ^2 -კრიტერიუმით, თითოეული ინტერვალისათვის გამოსათვლელია მოსალოდნელი სიხშირეები:

$$P(\text{დიამეტრის ზომა } 2.995\text{-ზე ნაკლებია}) = \\ = P\left(Z < \frac{2.995 - 3}{0.001}\right) = P(Z < -5) = 0.$$

ანალოგიურად გამოვთვლით, მაგალითად:

$$P(2.997 < X < 2.998) = P(-3 < Z < -2) = 0.4987 - 0.4772 = 0.0215.$$

χ^2 -კრიტერიუმი კარგად არ მუშაობს, თუ მოსალოდნელი სიხშირეების 20% 5-ზე ნაკლებია. ამ შემთხვევაში უნდა გავაერთიანოთ ინტერვალები:

ცხრილი 12.13

დიამეტრი	დაკვირვებული სიხშირე	მოსალოდნელი სიხშირე
2.998-მდე	31	11.4
2.998-დან-2.999-მდე	98	67.95
2.999-დან-3.000-მდე	146	170.65
3.000-დან-3.001-მდე	124	170.65
3.001-დან-3.002-მდე	83	67.95
3.002-ზე მეტი	18	11.40

$$\chi^2 = \sum \frac{(o_y - e_y)^2}{e_y} = 70.45.$$

თავისუფლების ხარისხი = ინტერვალების (კატეგორიების) რაოდენობა მინუს 1 = 5. $\alpha = 0.05$, $\chi^2_{crit} = 11.07$. ნულოვანი ჰიპოთეზა უნდა უარყვით ამის შედეგად დია-მეტრები არ არის ნორმალურად განაწილებული და ამიტომ ამოღებულ უნდა იქნეს პროდუქცია, ხოლო დაზგები, რომელზედაც მზადდებოდა ეს პროდუქცია, უნდა დარეგულირდეს ან შეიცვალოს.

14. სადაზღვევო კომპანია სწავლობს, დამოკიდებულია, თუ არა ავტოსაგზაო შემთხვევების რაოდენობა მძღოლის ასაკზე, სქესზე, საოჯახო სტატუსზე და ა. შ. 1000 კლიენტზე დაყრდნობით შეაქსეს შეუღლების ცხრილი:

ავტო-საგზაო შემთხვევებზე	ასაკი				
	25 წელზე ნაკლები	25-40	40-55	55-ზე მეტი	სულ
აქვთ სარჩელი	93	72	53	63	281
არ აქვთ სარჩელი	115	155	265	184	719
სულ	208	227	318	247	1000 = N

ამ მონაცემებზე დაყრდნობით დაასკენის თუ არა სადაზღვევო კომპანია, რომ ახალგაზრდებს მოსალოდნელზე მეტი ავტოსაგზაო შემთხვევები აქვთ და ამიტომ უნდა გაიზარდოს თუ არა მათთვის სადაზღვევო პრემია (გადასახადი)? $\alpha = 0.05$.

დამოკიდებულია თუ არა სტატისტიკაში მიღებული საბოლოო ქულა ლექტორზე?

ლექტორი 1	90	65	72	83	65	50
ლექტორი 2	90	88	83	83	65	65
ლექტორი 3	55	55	90	83	65	65

ლექტორი 1	65	83	83	55		65
ლექტორი 2	83	83	90	90	65	72
ლექტორი 3	60	55	65	72	90	65

შეამონმეთ ეს პიპოთეზა იმ პირობებში, როდესაც პოპულაციები არ არის ნორმალურად განაწილებული. იგულისხმეთ, რომ პოპულაციების დისპერსიები ტოლია.

H_0 : სამივე ლექტორის ჯგუფში მიღებულია ერთნაირი საშუალო ქულები.

H_a : ჯგუფების საშუალო ქულები ერთმანეთისგან განსხვავებულია.

პირველი ნაბიჯი კრუსკალ-უოლისის კრიტერიუმის გამოყენებისას არის შერჩევების გაერთიანება და რანჟირება.

მიღებული რანგებია:

ლექტორი 1	ლექტორი 2	ლექტორი 3
33.5	33.5	3.5
12.0	30.0	3.5
19.5	25.5	33.5
25.5	25.5	25.5
12.0	12.0	12.0
1.0	12.0	12.0
12.0	25.5	6.0
25.5	25.5	3.5
25.5	33.5	12.0
3.5	33.5	19.0
19.5	12.0	33.5
12.0	19.5	12.0
ჯამი = 201.5	ჯამი = 288	ჯამი = 176.5

მაშინ კრიტერიუმის სტატისტიკის მნიშვნელობა იქნება:

$$\begin{aligned}
 H &= \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) - \frac{\sum_{i=1}^k (r_i^3 - r_i)}{N^3 - N} = \\
 &= \frac{\frac{12}{36 \cdot 37} \left[\frac{201.5^2}{12} + \frac{288.0^2}{12} + \frac{176.5^2}{12} \right] - 3 \cdot 37}{1 - \frac{(4^3 - 4) + (11^3 - 11) + (4^3 - 4) + (8^3 - 8) + (6^3 - 6)}{36^3 - 36}} = \\
 &= \frac{5.1400}{0.9538} = 5.389.
 \end{aligned}$$

19.5 მეორდება 4-ჯერ, 12 მეორდება 11-ჯერ, 3.5 მეორდება 4-ჯერ, 25.5 მეორდება 8-ჯერ, 33.5 მეორდება 6-ჯერ. თავისუფლების ხარისხი $k-1=3-1=2$, თუ $\alpha=0.05$, $\chi^2_{2,0.05}=5.991$.

კრიტერიუმის სტატისტიკის მნიშვნელობა 5.389 ნაკლებია კრიტიკულ მნიშვნელობაზე 5.991. ამიტომ ნულოვან პიპოთეზას ვერ უარეყოფთ.

15. დამოკიდებულია თუ არა ალკოჰოლის მიღება იმაზე, ადამიანი თამბაქოს მომხმარებელია თუ არა? დაამუშავეთ შეუღლების ცხრილი და შეამოწმეთ პიპოთეზა. $\alpha = 0.05$.

		ენვეა	
		კი	არა
სვამს	კი	1449	500
	არა	46	281

რომელი უჯრედის დაკვირვებული სიხშირე აჩვენებს ყველაზე დიდ გადახრას დამოუკიდებლობისაგან?

16. 100000 პაციენტზე დაყრდნობით შეადგინეს შეუღლების ცხრილი, სადაც ერთი ცვლადი დაავადების სტატუსია, მეორე ცვლადი კი ამ დაავადებაზე ტესტის შედეგი.

	სადიაგნოსტიკო ტესტის შედეგი	
	დადებითი	უარყოფითი
დაავადებული	475	25
ჯანმრთელი	4975	94525

- რამდენი ადამიანია დაავადებული? რამდენია ჯანმრთელი?
 - დაავადებული ადამიანებისათვის რას უდრის დადებითი სადიაგნოსტიკო ტესტების პროპორცია? უარყოფითი ტესტების პროპორცია?
 - დაავადებული ადამიანისათვის რას უდრის შანსი, რომ ტესტი დაადასტურებს დაავადებას?
 - ჯანმრთელი ადამიანისთვის რას უდრის დადებითი ტესტების პროპორცია? უარყოფითი ტესტების პროპორცია?
 - ჯანმრთელი ადამიანისთვის რას უდრის ცრუ დადებითი ტესტის შანსი?
 - დაავადებული ადამიანისთვის დაავადების აღმოჩენის შანსი რამდენჯერ მეტია ცრუ შედეგის მიღების შანსზე ჯანმრთელი ადამიანისთვის?
 - დამოკიდებული არის თუ არა ტესტის შედეგი დაავადების სტატუსზე?
 - თუ ტესტის შედეგი დადებითია, აქედან ადამიანების რამდენი %-ია დაავადებული? ჯანმრთელი?
17. არის თუ არა ახალგაზრდების მისწრაფებები დამოკიდებული ოჯახის შემოსავალზე? იხილეთ ცხრილი

ოჯახის შემოსავალი	მისწრაფება			
	მხოლოდ სკოლა	ბაკალავრიატი	მაგისტრატურა	დოქტორანტურა
დაბალი	9	44	13	10
საშუალო	11	52	23	22
მაღალი	9	41	12	27

ამ ცხრილის დამუშავების შედეგი წარმოდგენილია SSPS ამონაბეჭდით.

statistic	df	Value	Prob
chi-Square	6	8.871	0.181

- ა) რას უდრის კრიტერიუმის სტატისტიკის მნიშვნელობა?
 ბ) დამოკიდებულია თუ არა მისწრაფება ოჯახის შემოსავალზე?
18. შეუღლების ცხრილში ერთი ცვლადი პოლიტიკური კუთვნილებაა, მეორე ცვლადი პოლიტიკური იდეოლოგია.

პოლიტიკური კუთვნილება	პოლიტიკური იდეოლოგია				
	ძალიან ლიბერა- ლური	ოდნავ ლიბე- რალური	ზომიერი	ოდნავ კონსერვა- ტორი	ძალიან კონსერ- ვატორი
დემოკრატი	80	81	171	41	55
რესპუბლიკელი	30	46	148	84	99

- ა) დამოკიდებულია თუ არა ეს ცვლადები?
 ბ) რომელი უჯრედის დაკვირვებული სიხშირე აჩვენებს ყველაზე დიდ გადახრას დამოუკიდებლობისგან?

რეგრესია და ნინასნარმეტყველება

(მარტივი რეგრესიის მოდელი)

ორ რაოდენობრივ ცვლადს შორის დამოკიდებულება შეიძლება იყოს ფუნქციონალური ანუ ისეთი, როცა ერთი ცვლადის მნიშვნელობების მიხედვით ცალსახად შეგვიძლია გამოვთვალოთ მეორე ცვლადის შესაბამისი მნიშვნელობები. ასეთი დამოკიდებულების მრავალ მაგალითს სკოლის პროგრამაში ეცნობოდით. მაგალითად, ვიცით, რომ დროს, სიჩქარესა და განვლილ მანძილს შორის დამოკიდებულება, თანაბარი მოძრაობისას, გამოითვლება ფორმულით $s = v \cdot t$. ამ ფორმულის მიხედვით უტყუარად შეგვიძლია გამოვთვალოთ განვლილი მანძილი, თუ ცნობილია სიჩქარე და დრო. ასეთივე მდგომარეობაა სამკუთხედის ფართობის გამოთვლისას. ვიცით, რომ ფართობი სიმაღლისა და ფუძის სიგრძეების ნამრავლის

ნახევრის ტოლია $S = \frac{ah}{2}$ (აქ a ფუძის სიგრძეა, h - სიმაღლისა). ცალსახად შეგ-

ვიძლია კვადრატის ფართობის გამოთვლა $S = a^2$ ფორმულით, სადაც a კვადრატის გვერდის სიგრძეა. ასეთი მაგალითები თქვენთვის მრავლადაა ცნობილი.

მეორე მხრივ, შესაძლებელია ცვლადებს შორის კავშირი არ იყოს ზუსტად ფუნქციონალური. მაგალითად, დათესილი მინდვრის ფართობით მოსავლის რაოდენობა ზუსტად, ფუნქციონალურად, არ შეიძლება აღინეროს. თუმცა შეგვიძლია ვივარაუდოთ, რომ მეტი ფართობის დათესვა იწვევს აღებული მოსავლის ზრდის ტენდენციას. ანუ ვვარაუდობთ, რომ ეს სიდიდეები (ფართობი და მოსავლიანობა) ერთმანეთთანაა დაკავშირებული. ასეთი ტიპის კავშირები, როცა ერთი ცვლადის ზრდა ან კლება, იწვევს მეორე ცვლადის ზრდის ან კლების ტენდენციას, აღინერება ამ ცვლადებს შორის კორელაციური კავშირით. თუ ერთი ცვლადის ცვლილება არ იწვევს მეორე ცვლადის ცვლილებას და, პირიქით, ვამბობთ, რომ ცვლადები არაკორელირებულნი არიან. სხვა შემთხვევაში ისინი კორელირებულნი არიან. თუ რამდენად ძლიერია ეს კავშირი, ცალკე კვლევის საგანია. კორელაციური კავშირის სიდიდეს კორელაციის კოეფიციენტით ვზომავთ. კორელაციის კოეფიციენტის დასადგენად ვატარებთ ცდას და მოვიპოვებთ მონაცემებს.

მეხუთე თავში, შერჩევებზე დაყრდნობით, ორ რაოდენობრივ ცვლადს შორის კავშირის დასადგენად, შემოღებული იყო პირსონის კორელაციის კოეფიციენტი:

$$\begin{aligned} r &= \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \\ &= \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}. \end{aligned}$$

ეს სიდიდე პოპულაციის უცნობი კორელაციის ρ კოეფიციენტის ნერტილოვანი შეფასებაა. r -ის საშუალებით შეიძლება გაკეთდეს დასკვნები ρ -ს მნიშვნელობის შესახებ.

დასკვნები კორელაციის კოეფიციენტის შესახებ

განვიხილოთ შემდეგი ჰიპოთეზის შემოწმების ამოცანა:

$$H_0: \rho = 0, \quad H_A: \rho \neq 0,$$

სადაც ρ პოპულაციის კორელაციის კოეფიციენტია.

თქვათ, (X, Y) სიდიდეების განაწილება ნორმალურია. მაშინ H_0 -ის სამართლიანობისას სტატისტიკას

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

აქვს t -განაწილება, თავისუფლების ხარისხით $n-2$. ამგვარად, t -განაწილების ცხრილის საშუალებით შეიძლება შემოწმდეს ეს ჰიპოთეზა.

სახელდობრ, H_0 -ს უარყოფთ, თუ:

$$|t| \geq t_{n-2, \alpha/2} \quad (t \leq -t_{n-2, \alpha/2} \quad \text{ან} \quad t \geq t_{n-2, \alpha/2})$$

და დავასკვნით, რომ ორ ცვლადს შორის ნამდვილად არსებობს კავშირი; წინააღმდეგ შემთხვევაში H_0 ჰიპოთეზის უარყოფის საფუძველი არ გაგვაჩნია.

მე-5 თავში შევისწავლეთ, რომ წრფივი კავშირი ორ ცვლადს შორის აღინერება განტოლებით:

$$Y' = \alpha + \beta X,$$

სადაც Y დამოკიდებული ცვლადია, X დამოუკიდებელი ცვლადია, α და β უცნობი სიდიდეებია (პოპულაციის პარამეტრებია). α წრფის თანაკვეთაა oy ღერძთან, β წრფის დახრილობა. ამ მოდელს მ ა რ ტ ი ვ ი რ ე გ რ ე ს ი ი ს მოდელი ეწოდება (ერთი დამოუკიდებელი ცვლადი X და ერთი დამოკიდებული ცვლადი Y).

უმცირეს კვადრატთა მეთოდის გამოყენებით დავადგენთ, რომ α და β უცნობი პარამეტრების შეფასებებია:

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}, \quad a = \bar{Y} - b\bar{X}.$$

$\hat{Y} = a + bX$ წრფეს ეწოდება რეგრესიის წრფის შეფასება, ან უმცირეს კვადრატთა მეთოდით აგებული წრფე. თუ ამ წრფის განტოლებაში X -ის ნაცვლად ჩავსვამთ მის დაკვირვებულ მნიშვნელობებს, შესაბამის $\hat{Y}_i$ მნიშვნელობებს ეწოდება Y ცვლადის პროგნოზირებული მნიშვნელობები, ხოლო $e_i = Y_i - \hat{Y}_i$ სიდიდეებს კი - ნაშთები (ან ცდომილებები).

ჰიპოთეზის შემოწმება დახრილობის შესახებ

ორ ცვლადს შორის მნიშვნელოვანი კავშირის დადგენის ალტერნატიული გზა მდგომარეობს რეგრესიის წრფის პარამეტრების შესახებ ჰიპოთეზათა შემოწმებაში.

შესამოწმებელია, რომ $\beta = 0$, ალტერნატივის საწინააღმდეგოდ – $\beta \neq 0$,

$$H_0: \beta = 0, \quad H_A: \beta \neq 0.$$

თუ რეგრესიის წრფის დახრილობა ნულის ტოლია, ეს ნიშნავს, რომ არ არსებობს წრფივი დამოკიდებულება ორ ცვლადს შორის. ამ შემთხვევაში „ნამდვილი“ დამოკიდებულება მოცემულია შემდეგი სახით:

$$y_i = \alpha.$$

Y ცვლადი არ არის დამოკიდებული დამოუკიდებელ X ცვლადზე და, უბრალოდ, წარმოადგენს შემთხვევით შერჩევას $\mu = \alpha$ საშუალოთი და მუდმივი დისპერსიით. ამიტომ, როცა ვამოწმებთ ამ ჰიპოთეზას, ფაქტობრივად ვამოწმებთ, არის თუ არა წრფივი კავშირი Y და X ცვლადებს შორის. როცა უარვყოფთ ნულოვან ჰიპოთეზას, ვრწმუნდებით, რომ X წარმოადგენს საჭირო ინფორმაციას Y -თვის.

კრიტერიუმის სტატისტიკა

ვთქვათ, Y_i არის Y ცვლადის დაკვირვებული მნიშვნელობები, $i = 1, 2, \dots, n$; n – შერჩევის მოცულობაა.

$\hat{Y}_i$ არის გამოთვლილი მნიშვნელობები, რომლებიც მიღებულია $\hat{Y} = a + bX$ განტოლებიდან, თუ ამ განტოლებაში ჩასმულია X -ის მნიშვნელობები X_i , $i = 1, 2, \dots, n$.

$Y_i - \hat{Y}_i = e_i$ – სხვაობას დაკვირვებულ და გამოთვლილ მნიშვნელობებს შორის ეწოდება **ნ ა შ თ ი ა ნ ც დ ო მ ი ლ ე ბ ა**.

ჰიპოთეზის შესამოწმებლად უნდა დავუშვათ, რომ ნაშთები ნორმალურადაა განაწილებული ნულის ტოლი საშუალოთი და მუდმივი σ^2 დისპერსიით (ამ დაშვების შესამოწმებლად აგებენ $(e_i, \hat{Y}_i)$ გრაფიკულ დამოკიდებულებას). თუ e_i ნორმალურადაა განაწილებული, b დახრილობის კოეფიციენტი იქნება ნორმალურად განაწილებული საშუალოთი β და დისპერსიით, რომელიც გამოითვლება ფორმულით:

$$se(b) = \frac{\sigma}{\sqrt{\sum (X - \bar{X})^2}}.$$

σ^2 -ის შეფასება გამოითვლება ფორმულით:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (Y - \bar{Y})^2 - b \sum (X - \bar{X})(Y - \bar{Y})}{n-2}.$$

კრიტერიუმის სტატისტიკაა:

$$T = \frac{b - \beta}{\frac{\hat{\sigma}}{\sqrt{\sum (X - \bar{X})^2}}}$$

და ამ სტატისტიკას აქვს t -განაწილება თავისუფლების ხარისხით $n-2$ (2 აკლდება, რადგან შესაფასებელი გვაქვს ორი — α და β პარამეტრი).

უარყოფის არე

უარყოფის არეს ავაგებთ სტიუდენტის ზედა კრიტიკული წერილების ცხრილის გამოყენებით, რადგან კრიტერიუმის სტატისტიკას აქვს t -განაწილება თავისუფლების ხარისხით $n-2$.

გადანყვეტილება

თუ კრიტერიუმის სტატისტიკის მნიშვნელობა მოხვდა კრიტიკულ არეში, α მნიშვნელოვნების დონით, ნულოვან ჰიპოთეზას უარყოფთ. თუ კრიტერიუმის სტატისტიკის მნიშვნელობა არ მოხვდა კრიტიკულ არეში, ნულოვან ჰიპოთეზას ვერ უარყოფთ, ეს კი ნიშნავს, რომ დამოკიდებული ცვლადი არ არის წრფივად დამოკიდებული დამოუკიდებელ ცვლადზე.

ჰიპოთეზის შემოწმება თანაკვეთის შესახებ

თუ ნულოვანი ჰიპოთეზა სრულდება, მაშინ a შეფასების საშუალოა α და სტანდარტული გადახრა გამოითვლება ფორმულით:

$$se(a) = \sigma \sqrt{\frac{\frac{1}{n} \sum X^2}{\sum (X - \bar{X})^2}}.$$

კრიტერიუმის სტატისტიკას აქვს სახე:

$$T = \frac{a - \alpha}{\frac{\hat{\sigma}}{\sqrt{\sum (X - \bar{X})^2}}}$$

და ამ სტატისტიკას აქვს t -განაწილება თავისუფლების ხარისხით $n-2$. უარყოფის არეს ჩვენ ავაგებთ სტიუდენტის განაწილების ზედა კრიტიკული წერილების ცხრილის გამოყენებით.

მაგალითი 13.1. X არის რეკლამაზე დახარჯული ფული (ათასი დოლარი), Y მოგება (ათასი დოლარი).

X :	Y :
0.8	22
1.0	28
1.6	22
2.0	26
2.2	34
2.6	18
3.0	30
3.0	38
4.0	30
4.0	40
4.0	50
4.6	46
$\Sigma = 32.8$	$\Sigma = 384$

არის თუ არა კავშირი ამ ორ ცვლადს შორის? გამოთვალეთ პირსონის კორელაციის კოეფიციენტი. შედეგსოთ შემდეგი ცხრილი:

XY :	X^2 :	Y^2 :
17.6	0.64	484
28.0	1.00	784
35.2	2.56	484
52.0	4.00	676
74.8	4.84	1156
46.8	6.76	324
90.0	9.00	90
114.0	9.00	1444
120.0	16.00	900
160.0	16.00	1600
200.0	21.16	2500
21.6		2116
$\Sigma = 1.150$	$\Sigma = 106.96$	$\Sigma = 13368$

ფორმულაში

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} =$$

$$= \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}.$$

ჩასმის შემდეგ მივიღებთ:

$$r = \frac{12 \cdot (11500) - 384 \cdot (32.8)}{\sqrt{(12 \cdot (106.96) - (32.8)^2)(12 \cdot (13.368) - (384)^2)}} = 0.7343.$$

რეკლამაზე დახარჯულ ფულსა და მოგებას შორის არსებობს ნრფივი დადებითი კავშირი.

დასკვნები კორელაციის კოეფიციენტის შესახებ

მიუხედავად იმისა, რომ კორელაციის კოეფიციენტი საკმაოდ დიდია, $r = 0.7343$, უნდა გვახსოვდეს, რომ ის არის გამოთვლილი მხოლოდ 12 მონაცემზე დაყრდნობით და ყოველთვის არსებობს შეცდომის ალბათობა. იმისათვის, რომ დაერწმუნდეთ არსებობს თუ არა კავშირი მოგებასა და რეკლამას შორის, უნდა შევამოწმოთ ჰიპოთეზა: უდრის თუ არა ნულს პოპულაციის კორელაციის კოეფიციენტი ρ .

$$H_0: \rho = 0, \quad H_A: \rho \neq 0.$$

ამისათვის კი გამოსათვლელია სტატისტიკის მნიშვნელობა:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.7343}{\sqrt{\frac{1-0.5392}{10}}} = 3.421.$$

ამ სტატისტიკას აქვს t -განაწილება თავისუფლების ხარისხით $n-2$. თუ $\alpha = 0.05$, $t_{crit} = 2.228$. რადგან $t = 3.421 > 2.228$, H_0 -ს უარეყოფთ. რეკლამაზე დახარჯულ ფულსა და მოგებას შორის არსებობს ნრფივი კავშირი, ე. ი. რაც უფრო მეტ ფულს დავხარჯავთ რეკლამაზე, მით უფრო მეტი მოგებაა მოსალოდნელი.

ორ ცვლადს შორის კავშირის დასადგენად მივმართოთ ალტერნატიულ გზას. საპროგნოზო განტოლებას აქვს სახე:

$$\hat{Y} = a + bX,$$

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} =$$

$$= \frac{1150 - \frac{32.8 \cdot 384}{12}}{106.96 - \frac{32.8^2}{12}} = 5.801,$$

$$a = \bar{Y} - b\bar{X} = 32 - 5.801 \cdot (2.733) = 16.143.$$

უმცირეს კვადრატთა მეთოდით მიღებულ რეგრესიის განტოლებას აქვს შემდეგი სახე:

$$\hat{Y} = 16.143 + 5.801 \cdot X$$

შევამოწმოთ ჰიპოთეზა:

$$H_0: \beta = 0, H_A: \beta \neq 0$$

პოპულაციის დისპერსიის შეფასება:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum e_i^2}{n-2} = \frac{\sum (Y - \bar{Y})^2 - b \sum (X - \bar{X}) \sum (Y - \bar{Y})}{n-2} = \\ &= \frac{1080 - 5.801 \cdot 100.4}{12-2} = 49.76,\end{aligned}$$

$$\sum e_i^2 = 497.556,$$

$$T = \frac{\frac{b - \beta}{\hat{\sigma}}}{\sqrt{\frac{\sum (X - \bar{X})^2}{n}}} = \frac{\frac{5.801 - 0}{\sqrt{49.76}}}{\sqrt{17.31}} = 3.421,$$

$\hat{\sigma}$ შეფასების სტანდარტული შეცდომაა.

თავისუფლების $n-2=12-2=10$ ხარისხისა და $\alpha=0.05$ მნიშვნელოვნების დონისათვის, კრიტიკული მნიშვნელობა იქნება $t_{n-2, \alpha/2} = t_{10, 0.025} = 2.228$. ვინაიდან $3.421 > 2.228$ -ზე, ნულოვანი ჰიპოთეზა უნდა უარყვოდ. რეკლამაზე დახარჯულ ფულსა და მოგებას შორის არსებობს წრფივი დამოკიდებულება. რაც უფრო დიდ ფულს ეხარჯავთ რეკლამაზე, მით უფრო დიდი მოგება.

შევამოწმოთ ჰიპოთეზა თანაკვეთის შესახებ.

შევამოწმოთ, მაგალითად, $H_0: a = 1, H_A: a \neq 1$.

კრიტერიუმის სტატისტიკა:

$$T = \frac{\frac{a - \alpha}{\hat{\sigma}}}{\sqrt{\frac{1}{n} \frac{\sum X^2}{\sum (X - \bar{X})^2}}} = 2.99.$$

ვინაიდან $2.99 > 2.228$, ნულოვანი ჰიპოთეზა $\alpha=0.05$ მნიშვნელოვნების დონით უნდა უარყვოდ.

საპროგნოზო ინტერვალები

განტოლება,

$$\hat{Y} = a + bX$$

გვაძლევს საშუალებას დავინახოთ რა არის მოსალოდნელი, თუ X მიიღებს გარკვეულ მნიშვნელობას. მაგალითად, თუ X რეკლამაზე დახარჯული ფულია, რა მოგებას მოიტანს ის?

სიდიდეს - $\hat{Y}_0 = a + bX_0$ აქვს საშუალო $\alpha + \beta X_0$ და სტანდარტული გადახრა

$$se(Y_0) = \sigma \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X - \bar{X})^2}}.$$

ნდობის ინტერვალს Y_0 -თვის აქვს სახე:

$$Y_0 \pm t\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X - \bar{X})^2}},$$

სადაც t სიდიდე უნდა ვიპოვოთ t -განაწილების ზედა კრიტიკული წერტილების ცხრილში, $n-2$ თავისუფლების ხარისხის და $(100-\alpha)\%$ სანდოობის შესაბამისად. ეს ინტერვალი დამოკიდებულია იმაზე, თუ კონკრეტულად რომელი მნიშვნელობისათვისაა ის ასაგები. $(X_0 - \bar{X})^2$ ზრდასთან ერთად იზრდება სტანდარტული გადახრა და ინტერვალი უფრო განიერი ხდება იმ X_0 -თვის, რომლებიც უფრო შორს არიან გადახრილი $\bar{X}$ -გან და უფრო ვიწროა იმ მნიშვნელობებისათვის, რომლებიც ახლოს არიან $\bar{X}$ -თან.

დავუბრუნდეთ 13.1 მაგალითს და ვუპასუხოთ შემდეგ კითხვებს:

- საშუალოდ რა მოგებას უნდა ველოდოთ, თუ რეკლამაზე დავხარჯავთ 2100 დოლარს?
- ააგეთ 95%-იანი ნდობის ინტერვალი ამ მოგებისათვის.

ამოხსნა: ა) საპროგნოზო განტოლებაში ჩასმის შემდეგ მივიღებთ:

$$\hat{Y} = a + bX = 16143 + 5.801 \cdot 2.1 = 28.325.$$

საშუალოდ მოგება შეადგენს 28.325 დოლარს.

ბ) ამ მოგების 95%-იანი ნდობის ინტერვალია:

$$28.325 \pm 2.228 \cdot 7.05 \sqrt{1 + \frac{1}{12} + \frac{(2.1 - 2.73)^2}{17.31}} = 28.325 \pm 16.52,$$

ანუ (11.810; 44.840).

დეტერმინაციის კოეფიციენტი.

მოდელის ვარგისიანობის შემოწმება

მაგალითი 13.2. განვიხილოთ ისევ 13.1 მაგალითი. მოგებებში სრული ცვალებადობა მოცემულია ფორმულით:

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 13368 - \frac{(384)^2}{12} = 1080,$$

სადაც SST დამოკიდებული ცვლადის (მოგებების) სრული ცვალებადობაა, „ შერჩევის მოცულობაა, Y_i დამოკიდებული ცვლადის i -ური მნიშვნელობაა, $\bar{Y}$ დამოკიდებული ცვლადის შერჩევის საშუალოს მნიშვნელობაა.

უმცირეს კვადრატთა მეთოდის გამოყენებით ვაგებთ წრფეს, რომლისთვისაც მინიმალურია ნაშთების კვადრატების ჯამი:

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 497.556,$$

სადაც Y_i არის დამოკიდებული ცვლადის i -ური დაკვირვებული მნიშვნელობა, $\hat{Y}_i$ რეგრესიის განტოლებაში X ჩასმის შემდეგ მიღებული (გამოთვლილი) მნიშვნელობა.

SSE რეგრესიის წრფისგან გადახრების კვადრატების ჯამი წარმოადგენს დამოკიდებული ცვლადის ცვალებადობის ნაწილს, რომელიც არ არის ახსნილი რეგრესიის წრფით.

ცვალებადობის ნაწილი, რომელიც ახსნილია რეგრესიის წრფით, გამოითვლება როგორც

$$SSR = SST - SSE = 1080 - 497.556 = 582.444.$$

სრულ ცვალებადობაში რეგრესიით ახსნილ ცვალებადობის წილს ეწოდება დეტერმინაციის კოეფიციენტი:

$$R^2 = \frac{SSR}{SST} = \frac{582.444}{1080} = 0.5393.$$

ეს ნიშნავს, რომ ამ შერჩევაში მოგებების ცვალებადობის 53.93% აიხსნება რეკლამაზე დახარჯული თანხით.

R^2 -ის თვისებები

R^2 იღებს მნიშვნელობებს ნულსა და ერთს შორის. თუ ცვლადებს შორის ზუსტი წრფივი კავშირია, მაშინ $R^2 = 1$. ამ დროს ყველა დაკვირვებული მნიშვნელობა ძვეს რეგრესიის წრფეზე. რაც უფრო ძლიერია კავშირი ცვლადებს შორის, მით უფრო ახლოსაა R^2 სიდიდე ერთთან.

როგორც ნაჩვენებია მე-5 თავში, არსებობს დეტერმინაციის კოეფიციენტის გამოთვლის ალტერნატიული გზა:

$$R^2 = r^2, \quad r^2 = 0.7343^2 = 0.5392,$$

სადაც R^2 დეტერმინაციის კოეფიციენტია, r პირსონის კორელაციის კოეფიციენტია.

თუ R^2 მცირეა, ე. ი. რეგრესიით ახსნილია სრული გაბნევის მცირე წილი, მოსაძებნია სხვა ალტერნატიული მოდელი (ვთქვათ, არანრფივი ან მრავლობითი წრფივი რეგრესიის მოდელი და სხვა), რომელიც უფრო ეფექტურად ახსნის დამოკიდებული ცვლადის დაკვირვებული მნიშვნელობების სრულ გაბნევას თავისი საშუალო მნიშვნელობის მიმართ.

რეგრესიული ანალიზი მრუდწირული (არანრფივი) კავშირისთვის

თუ გაბნევის დიაგრამა გვიჩვენებს, რომ კავშირი ორ ცვლადს შორის არ არის წრფივი და ახლოსაა ერთ-ერთ ქვემოთ მოყვანილ დამოკიდებულებასთან:

$$Y = X^2, \quad Y = \sqrt{X} \quad \text{ან} \quad Y = \log X \quad \text{ან სხვა,}$$

მაშინ ხელოვნურად, ცვლადების ტრანსფორმაციით, კავშირი შეიძლება გარდაიქმნას წრფივ დამოკიდებულებად.

მაგალითი 13.3. დამოკიდებულება გამოშვებულ პროდუქციას და ზეგანაკვეთურ სამუშაო საათებს შორის მოცემულია ცხრილით:

გამოშვებული პროდუქცია (ცალი)	200	500	1000	1300	1300	800
ზეგანაკვეთი (საათი)	25	25	75	175	200	75

გამოშვებული პროდუქცია (ცალი)	200	1400	600	1100	900	900	1200
ზეგანაკვეთი (საათი)	50	225	75	125	75	75	175

გამოშვებული პროდუქცია (ცალი)	400	30	160	1100	1200	700	1000
ზეგანაკვეთი (საათი)	50	50	250	150	50	100	

წრფივი რეგრესიული ანალიზის შედეგები მოცემულია 13.1 ცხრილში 3.1.

ცხრილი 13.1

REGRESSION EQUATION			
VARIABLE	B	ST. ERROW	T-VALUE
X	0.1526	0.01858	8.20
INTERSEPT	-24.78		
CORRELATION COEFF.	0.888		
R SQUARE	0.789		
STANDARD ERROR OF ESTIMATE	32.08		

რეგრესიის განტოლებას აქვს სახე:

$$\hat{Y} = -24.78 + 0.1526X.$$

გაბნევის დიაგრამა გვიჩვენებს, რომ კავშირი ამ ორ ცვლადს შორის შესაძლებელია უფრო კარგად აისახოს არანრფივი დამოკიდებულებით.

X -ის ნაცვლად შემოვიღოთ ახალი ცვლადი X^2 და ჩავატაროთ რეგრესიული ანალიზი. ქვემოთ მოცემულია ამონაბეჭდი SPSS პროგრამიდან (იხ. 13.2 ცხრილი).

რეგრესიის განტოლებას აქვს სახე:

$$\hat{Y} = 17.224 + 0.0001X^2.$$

თუ შევადარებთ 13.1 და 13.2 ცხრილების შედეგებს, ვნახავთ, რომ ცვლადის ტრანსფორმაციის შედეგად R^2 გაიზარდა 86.8%-დან 93.3%-მდე. სტანდარტული შეცდომა შემცირდა 32.08-დან 21.55-მდე. მეორე წრფე უფრო სასარგებლოა პროგნოზისთვის, მას ახლავს შეფასების ნაკლები შეცდომა.

ცხრილი 13.2

REGRESSION EQUATION			
VARIABLE	B	ST. ERROW B	T-VALUE
X SQUARE	0.0001	0.0000076	13.08
INTERSEPT	17.224		
CORRELATION COEFF.	0.9513		
R SQUARE	0.905		
STANDARD ERROW OF ESTIMATE	21.55		

მრავლობითი რეგრესია და კორელაცია

მარტივი რეგრესიის მოდელი გვეხმარება გარკვევაში, თუ რა გავლენას ახდენს ერთი ცვლადი (დამოუკიდებელი ცვლადი) მეორე ცვლადზე (დამოკიდებულ ცვლადზე). მიუხედავად იმისა, რომ მარტივი რეგრესიის მოდელს იყენებენ მრავალ პრაქტიკულ ამოცანაში, არსებობს ამოცანები, სადაც ერთი დამოუკიდებელი ცვლადით შემოფარგვლა არ არის გამართლებული.

განვიხილოთ მაგალითი.

მაგალითი 13.4. რაზეა დამოკიდებული სახლის გასაყიდი ფასი Y ? ჩამოვთვალოთ შესაძლო დამოუკიდებელი ცვლადები: X_1 სახლის ფართობი (კვ.მ), X_2 – სახლის „ასაკი“, X_3 – საძინებელი ოთახების რაოდენობა, X_4 – სველი წერტილების რაოდენობა, X_5 – ბუხრების რაოდენობა, Y – სახლის გასაყიდი ფასი.

მარტივი რეგრესიის მოდელის განხილვისას ვიხილავდით ერთ დამოუკიდებელ X ცვლადს და ერთ დამოკიდებულ Y ცვლადს. დამოუკიდებელი ცვლადით ვცდილობდით აგვეხსნა დამოკიდებული ცვლადის ცვალებადობა. როგორც წესი, არ არსებობს ზუსტი კავშირი X და Y ცვლადებს შორის, ეს კი ნიშნავს იმას, რომ

თავს იჩენს ϵ , ნაშთი – სხვაობა დაკვირვებულ და რეგრესიით გამოთვლილ სიდიდეებს შორის. დაშვებულ იქნა, რომ ϵ , მნიშვნელობები ნორმალურად არიან განაწილებული ნულის ტოლი საშუალოთი და სტანდარტული გადახრით, რომელსაც ეწოდება შეფასების სტანდარტული შეცდომა.

თუ სტანდარტული შეცდომა საკმაოდ დიდია, უნდა ჩავთვალოთ, რომ რეგრესიის მოდელი არ არის სასარგებლო პროგნოზისათვის.

დამოკიდებულ ცვლადზე ჯერ „არახსნილი“ ცვალებადობის ასახსნელად მრავლობითი რეგრესიის მოდელში დამატებულია დამოუკიდებელი ცვლადები. მრავლობითი რეგრესიის მოდელს ზოგადად აქვს შემდეგი სახე:

$$Y_i = \beta_0 + \beta_1 Y_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik},$$

სადაც Y დამოკიდებული ცვლადია, $X_1, X_2, \dots, X_k$ დამოუკიდებელი ცვლადებია, რომლებსაც ხშირად ამხსნელ ცვლადებს ან პრედიქტორებს, ანდა რეგრესორებს უწოდებენ. β_i არის X_i ცვლადის რეგრესიის კოეფიციენტი, $i = 1, 2, \dots, k$.

იმისათვის, რომ გამოვიყენოთ მრავლობითი რეგრესიის მოდელი, უნდა სრულდებოდეს შემდეგი პირობები:

1. ნაშთებს აქვს ნორმალური განაწილება;
2. ნაშთების საშუალო ნულის ტოლია;
3. დამოუკიდებელი ცვლადების ნებისმიერი კომბინაციისთვის ნაშთების დისპერსია არის ერთი და იგივე და σ^2 -ის ტოლი.

როცა გვაქვს დამოკიდებული ცვლადის და k დამოუკიდებელი ცვლადის n მოცულობის შერჩევები, უმცირეს კვადრატთა მეთოდის გამოყენებით მივიღებთ:

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_k X_{ik},$$

სადაც b_i , $i = 1, 2, \dots, k$, დახრილობის კოეფიციენტია, $\hat{Y}_i$ დამოკიდებული ცვლადის რეგრესიით გამოთვლილი მნიშვნელობაა, X_i – დამოუკიდებელი ცვლადია, b_i – უმცირეს კვადრატთა მეთოდით მიღებული β_i პარამეტრების შეფასებებია, $i = 1, 2, \dots, k$. რეგრესიის კოეფიციენტები გამოთვლილია უმცირეს კვადრატთა მეთოდის გამოყენებით, ანუ $\sum (Y_i - \hat{Y}_i)^2$ მინიმალურობის მოთხოვნით.

შ ე ნ ი შ ვ ნ ა . შერჩევის მოცულობა უნდა იყოს 4-ჯერ მეტი მაინც, ვიდრე დამოუკიდებელი ცვლადების რაოდენობა.

ცხრილი 13.3 (მაგალითი 13.3)

დაკვირვებები	Y	X_1	X_2	X_3	X_4	X_5
1	21400	1410	3.0	3.0	1.0	0
2	37275	37275	5.0	3.0	2.5	1
$n = 531$	47175	2250	1.0	4.0	2.75	2

ცხრილი 13.4

ცვლადი	საშუალო	სტანდარტული გადახრა	მოცულობა
Y	45009.6139	14290.2132	531
X_1	1715.5273	588.1123	531
X_2	4.3936	7.7781	531
X_3	3.4038	0.0555	531
X_4	1.9143	0.6094	531
X_5	0.9208	0.5608	531

კორელაციური მატრიცა

პირველ რიგში გამოსათვლელია კორელაციის კოეფიციენტები დამოკიდებულ ცვლადსა (სახლის გასაყიდ ფასი) და ცალ-ცალკე ყველა დამოუკიდებელ ცვლადს შორის.

კორელაციის კოეფიციენტი ყველა წყვილს შორის გამოითვლება ფორმულით:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

კორელაციურ მატრიცას აქვს სახე:

ცხრილი 13.5

	Y	X_1	X_2	X_3	X_4	X_5
Y	1.000	0.841	-0.068	0.494	0.720	0.599
X_1	0.841	1.000	0.054	0.644	0.680	0.589
X_2	-0.068	0.054	1.000	0.007	-0.149	0.086
X_3	0.494	0.644	0.007	1.000	0.551	0.338
X_4	0.720	0.680	-0.149	0.551	1.000	0.518
X_5	0.599	0.589	0.086	0.338	0.518	1.000

კორელაციის კოეფიციენტი გასაყიდ ფასს და სახლის ფართობს შორის 0.841-ია;

კორელაციის კოეფიციენტი გასაყიდ ფასს და სახლის ასაკს შორის უარყოფითა და უდრის -0.068 -ს, კორელაციის კოეფიციენტი სახლის ფართობსა და სახლის ასაკს შორის 0.054 -ია და ა. შ.

ამონაბეჭდი SPSS პროგრამიდან

ცხრილი 13.6

R SQUARE	0.76686			
ADJUSTED R SQUARE	0.76509			
STANDARD ERROW	6926.16662			
ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
REGRESSION	4	82998243602.67404	20749560900.6651	432.53678
RESIDUAL (UNEXPLAINED)	536	252331158443	47971784.11251	
VARIABLES IN THE EQUATION				
VARIABLE	B	STD.ERROW B	T	
X_1	16.43956	0.84287	19.504	
X_3 BEDROOMS	-2845.89318	616.91611	-4.613	
X_4 BATHS	6599.24009	604.46151	9.367	
X_5 FIREPLACES	2507.93424	681.93968	3.677	
CONSTANT	11549.1426			

რეგრესიის განტოლებას აქვს სახე:

$$Y = 11549.14 + 16.44X_1 - 2845X_3 + 6599.24X_4 + 2507.93X_5.$$

X_2 ცვლადი არ არის შეყვანილი, რადგანაც კორელაცია გასაყიდ ფასსა და სახლის ასაკს შორის უმნიშვნელოა ($r = 0.068$).

გავაკეთოთ პროგნოზი: საშუალოდ რა ფასად გაიყიდება სახლი, თუ $X_1 = 2100$, $X_3 = 4$, $X_4 = 1.75$, $X_5 = 2$?

ფასი = $11549.14 + 16.44(2100) - 2845.89(4) + 6599.24(1.75) + 2507.93(2) = 51253.19$ დოლარი.

13.6 ცხრილზე დაყრდნობით შეგვიძლია გავაკეთოთ დასკვნები რეგრესიის კოეფიციენტების შესახებ. შევამოწმოთ $H_0: \beta = 0$ ჰიპოთეზები $H_A: \beta \neq 0$ ალტერ-

ნატივის საწინააღმდეგოდ. ამისათვის უნდა გავიგოთ რას უდრის T სტატისტიკის მნიშვნელობა. T -ს ყველა (19.504, -4.613, 9.367, 3.677) მნიშვნელობა მეტია 1.96 კრიტიკულ მნიშვნელობაზე. ამრიგად, არსებული მონაცემებით, ნულოვან ჰიპოთეზას უარყოფთ და $\alpha = 0.05$ მნიშვნელოვნების დონით სარწმუნოა მტკიცება, რომ სახლის ფასსა და განხილულ დამოუკიდებელ ცვლადებს შორის წრფივი კავშირი არსებობს.

რამდენად კარგია პროგნოზი?

13.5 ცხრილის მიხედვით, დეტერმინაციის კოეფიციენტი $R^2 = 0.76686$. ფასების თითქმის 77% აიხსნება ოთხი დამოუკიდებელი ცვლადით.

F სტატისტიკაზე დაყრდნობით შეგვიძლია აგრეთვე დავასკვნათ, ხსნის თუ არა მოდელი ფასების ცვალებადობის მნიშვნელოვან ნილს. F -ის მნიშვნელობა უდრის 432.53678. კრიტიკული მნიშვნელობა, როცა $\alpha = 0.01$ და თავისუფლების ხარისხი $531 - 5 = 526$ (შეფასებულია 5 კოეფიციენტი), არის 3.32.

დასკვნა. რეგრესიის მოდელი ხსნის ფასების ცვალებადობის მნიშვნელოვან ნაწილს. ADJUSTED R^2 – შესწორებული დეტერმინაციის კოეფიციენტი (თავისუფლების ხარისხებთან შეთანხმებული). ამ სიდიდის გამოყენების მნიშვნელობა მდგომარეობს იმაში, რომ, თუ დამოუკიდებელ ცვლადთა k რაოდენობა შედარებადია n -თან, R^2 -ის მნიშვნელობა შეიძლება არარეალურად დიდი გამოვიდეს და მიგვიყვანოს მცდარ დასკვნამდე. სწორედ ამის ასაცილებლად შემოღებულია ADJUSTED R^2 .

იმისათვის, რომ ვუპასუხოთ კითხვას, რამდენად კარგია პროგნოზი (რამდენად ვარგისიანია მოდელი), ასევე უნდა ვნახოთ რას უდრის პროგნოზის სტანდარტული შეცდომა (ცხრილი 13.5). მიუხედავად იმისა, რომ დეტერმინაციის კოეფიციენტი დიდია (0.765), სტანდარტული შეცდომაც დიდია და შეადგენს 6926.17 დოლარს.

ალტერნატიული გზა ვარიაციის კოეფიციენტის გამოთვლა. ვარიაციის კოეფიციენტი:

$$CV = \text{სტანდარტული შეცდომა} / Y\text{-ის საშუალო} = \frac{6926.17}{45009.6139} \cdot 100\% = 15.3\%$$

(ითვლება, რომ, თუ მოდელი ვარგისიანია, ის არ უნდა აღემატებოდეს 10%-ს).

როგორ უნდა შევამოწმოთ პროგნოზის შეცდომა?

სანამ ვუპასუხებთ ამ კითხვას, მაგალითზე ვაჩვენოთ რას ნიშნავს რეგრესიის კოეფიციენტები.

β_0 – მოდელში არის თანაკვეთა, β_0 არის Y -ის საშუალო, როცა ყველა $X_i = 0$. ჩვენ მაგალითში ეს იქნება ნაკეთის ფასი.

β_1 – ფასების საშუალო ცვლილება, როცა ფართობი იცვლება ერთი კვადრატული მეტრით. $\beta_1 = 16.44$ ნიშნავს, რომ ფასი საშუალოდ 16.44 დოლარით იზრდება, როცა ფართი იზრდება 1 კვ.მ-ით.

მაგალითი 13.5. რეგრესიის განტოლება სახლის გასაყიდ ფასსა და სახლის ფართობს შორის შემდეგი სახისაა:

$$\text{გასაყიდი ფასი} = 9915.78 + 20.45 (\text{ფართობი}).$$

შემოვიღოთ ახალი ცვლადი:

$X_6 = 1$ – სახლში არის კონდიციონერი,

$X_6 = 0$ – სახლში არ არის კონდიციონერი,

მაშინ რეგრესიის განტოლება მიიღებს სახეს

$$\hat{Y} = b_0 + b_1 X_1 + b_6 X_6 = 8739.21 + 18.11 X_1 + 2455.16 X_6.$$

თუ $X_6 = 0$, მაშინ $\hat{Y} = 8739.21 + 18.11 X_1$.

თუ $X_6 = 1$, მაშინ $\hat{Y} = 11194.37 + 18.11 X_1$.

სახლის ფასი საშუალოდ $11194.37 - 8739.21 = 2455.16$ დოლარით მეტია, თუ სახლში დაყენებულია კონდიციონერები.

შემოვიღოთ კიდევ ერთი ფიქტიური ცვლადი – სახლის ადგილმდებარეობა.

$X_7 = 1$ – პრესტიჟული ადგილი,

$X_7 = 0$ – არაპრესტიჟული ადგილი.

ცხრილი 13.7

R SQUARE	0.8940
ADJUSTED R SQUARE	0.8920
STANDARD ERROW	514.1353

$$\hat{Y} = 1693 X_1 + 4154.21 X_4 + 2010.45 X_5 + 1972.11 X_6 + 9699.80 X_7.$$

ჩვენ ვხედავთ, რომ ცვლადის ამოღებით და ორი ახალი ცვლადის შეყვანით პროგნოზის შეცდომა მნიშვნელოვნად შემცირდა.

შეჯამება

რეგრესია და კორელაცია შეისწავლის ორ ცვლადს შორის სტატისტიკური კავშირის ასპექტებს. აღწერილია უცნობი რეგრესიის კოეფიციენტების შეფასების პროცედურები, ჰიპოთეზების შემოწმება და ნდობის ინტერვალების აგება; უცნობი დისპერსიის შეფასება, კორელაციისა და დეტერმინაციის კოეფიციენტების გამოთვლა; მოდელის გამოსაყენებლად ვარგისიანობის შემოწმება.

სმირ შემთხვევაში რაიმე ცვლადის ვარიაბელობის (ცვალებადობის) ახსნა ვერ ხერხდება მხოლოდ ერთი ფაქტორის გავლენით და ამიტომ განიხილება უფრო რთული მრავლობითი წრფივი რეგრესიული მოდელი.

ასევე განიხილება, როგორ შეიძლება ცვლადთა გარდაქმნით არანრფივი კავშირის შესწავლა.

რამდენიმე პრაქტიკული ამოცანის ამოხსნა

1. მერიის თანამშრომელს დაავალეს გაერკვია, როგორ არის დამოკიდებული წყლის მოხმარება ოჯახის ზომაზე. მან გადანყვიტა წრფივი რეგრესიული მოდელის აგება. წინა თვის 15 ოჯახის მონაცემებია:

წყლის მოხმარება (გალონი):	1100	1425	785	950	1200
ოჯახის ზომა:	3	5	2	3	4

წყლის მოხმარება (გალონი):	1152	973	1225	1600	700
ოჯახის ზომა:	4	3	5	4	3

წყლის მოხმარება (გალონი):	1100	1414	700	953	1063
ოჯახის ზომა:	5	4	2	2	2

კორელაციის კოეფიციენტი მიიღეს კომპიუტერული პროგრამის მეშვეობით.

ცხრილი 13.8

R-SQUARE		0.551986	
CORRELATION COEFFICIENT		0.742958	
F FOR ANALYSIS OF VARIANCE (D.F. = 1,13)		16.1069	
STANDARD ERROR OF ESTIMATE		197.084	
VARIABLE	REG. COEFF.	STD. ERR. COEFF.	T
Y	188.011	46.978	4.00212
INTERCEPT	470.094		

კორელაციის კოეფიციენტი $r = 0.743$.

შევამოწმოთ ჰიპოთეზა: $H_0: \rho = 0$, $H_A: \rho \neq 0$, $\alpha = 0.05$,

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.743}{\sqrt{\frac{1-(0.743)^2}{13}}} = 4.002.$$

$t_{\alpha/2} = t_{n-2, \alpha/2} = t_{13, 0.025} = 2.160$ და რადგან $4.002 > 2.160$ დავასკვნით, რომ ოჯახის ზომასა და მოხმარებულ წყალს შორის დადებითი კავშირი არსებობს.

თუ კომპიუტერით დავთვლით რეგრესიის კოეფიციენტებს, მაშინ მივიღებთ, რომ საპროგნოზო წრფეს აქვს შემდეგი სახე:

$$\hat{Y} = 470.094 + 188.011X.$$

პროგნოზის სტანდარტული შეცდომაა 197.084. დეტერმინაციის კოეფიციენტი უდრის 0.551986, ანუ ოჯახის ზომა ხსნის წყლის ხარჯვის ცვალებადობის მხოლოდ 55%-ს.

ვთქვათ, ამ მოდელზე დაყრდნობით გვინდა გავიგოთ საშუალოდ რამდენ წყალს ხარჯავს ხუთკაციანი ოჯახი და აევაგოთ 90%-იანი ნდობის ინტერვალი. გამოვთვალოთ ნაშთები.

	დაკვირვებული Y	გამოთვლილი $\hat{Y}$	ნაშთი $e = Y - \hat{Y}$
1	1100	1034.13	65.8713
2	1425	1410.15	14.8486
3	785	846.17	-61.1173
4	950	1034.13	-84.1287
5	1200	1222.14	-22.1401
6	1152	1222.14	-70.1401
7	973	1034.13	-61.1287
8	1525	1410.15	114.849
9	1600	1222.14	377.860
10	700	1034.13	-334.129
11	1100	1410.15	-310.151
12	1412	1222.14	191.860
13	700	846.117	-146.117
14	953	846.117	106.883
15	1063	846.117	216.883

ნდობის ინტერვალია

$$\hat{Y} \pm t_{\alpha/2} \sqrt{\frac{SSE}{n-2} \left(1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum X^2 - \frac{(\sum X)^2}{n}} \right)}.$$

$\hat{Y}$ -ის გამოსათვლელად საპროგნოზო განტოლებაში ჩავსვით $X = 5$,

$$\hat{Y} = 470.097 + 188.01 \cdot 5 = 1410.152.$$

ხუთკაციანი ოჯახი საშუალოდ ხარჯავს 1410.152 გალონს.

13.8 ცხრილიდან $\sqrt{\frac{SSE}{n-2}} = 197.84$ (სტანდარტული შეცდომა), $t_{\alpha/2} = 1.771$,

$\alpha = 0.1$.

ნდობის ინტერვალია:

$$1410 \pm 1.771 \cdot 197.084 \sqrt{1 + \frac{1}{n} + \frac{(5-3.4)^2}{191 - \frac{2601}{15}}} = 1410 \pm 384.28.$$

ხუთკაციანი ოჯახი ხარჯავს 1025 გალონიდან 1794.28 გალონამდე. მიუხედავად იმისა, რომ ეს შეიძლება იყოს სასარგებლო ინფორმაცია, შეფასების სიზუსტე არ არის საკმარისი.

კითხვები

1. არსებობს თუ არა კავშირი ადამიანების ასაკსა და საცურო აუზზე დასწრებას შორის?

ასაკი X	16	15	28	16	29	38
კვირაში დასწრება Y	7	3	6	5	1	2

ასაკი X	48	18	24	33	56
კვირაში დასწრება Y	4	7	4	5	1

- ა) ააგეთ გაბნევის დიაგრამა.
 - ბ) გამოთვალეთ კორელაციის კოეფიციენტი და შეამოწმეთ მნიშვნელოვნად განსხვავდება თუ არა ის პოპულაციაში ნულოვანი კორელაციისაგან 0.05 დონეზე.
 - გ) გააკეთეთ დასკვნები მოდელის პარამეტრების შესახებ.
 - დ) შეამოწმეთ მოდელის ვარგისიანობა.
 - ე) კვირაში რამდენჯერ მიდის აუზზე 55 წლის ადამიანი? ააგეთ 95%-იანი ნდობის ინტერვალი ამ ასაკის ადამიანთა დასწრების რაოდენობისათვის. როგორ ფიქრობთ, ეს ინტერვალი 29 წლის ადამიანის შესაბამის ინტერვალთან უფრო ვიწრო იქნება თუ განიერი?
2. ცხრილში მოცემულია სკოლების კურსდამთავრებულთა რაოდენობა და სკოლისათვის შემოწირული ფულის რაოდენობა.

სკოლა	1	2	3	4	5
კურსდამთ.	987	1350	2345	1300	12569
თანხა	234700	769000	1230000	450780	645000

სკოლა	6	7	8	9
კურსდამთ.	8560	3450	1890	23456
თანხა	2650000	143000	230000	4560000

- ა) გამოთვალეთ კორელაციის კოეფიციენტი და შეამოწმეთ მნიშვნელოვნად განსხვავდება თუ არა ის პოპულაციაში ნულოვანი კორელაციისაგან 0.05 დონეზე.
- ბ) საშუალოდ რა თანხა ჩაირიცხება სკოლის ანგარიშზე, თუ 12000 კურსდამთავრებულია?
3. კომპანია ამზადებს კვების დანამატებს მსხვილფეხა რქოსანი პირუტყვისათვის. 335 ხბოს წონაში ყოველდღიურ ზრდასა და დანამატის რაოდენობას შორის კორელაციის კოეფიციენტი $r = 0.104$. ამ სიდიდეზე დაყრდნობით რა დასკვნის გაკეთება შეგიძლიათ ამ ორი სიდიდის რეალურ კავშირზე.
4. კომპანია რეცხავს და წმენდს მანქანებს. მომსახურების გაუმჯობესებისათვის კლიენტებს სთხოვეს შეეფასებინათ სერვისის მუშაობა სარეიტინგო სკალაზე (0 – ძალიან ცუდი, 100 – საუკეთესო მომსახურება). მიღებული იყო შემდეგი მონაცემები:

მომსახურების დრო (საათი), X	1.5	0.2	0.3	0.35	1.0	0.23	0.32
შეფასება, Y	85	60	70	72	80	65	70

მომსახურების დრო (საათი), X	2.5	1.6	0.27	1.6	0.40	0.45	2.0
შეფასება, Y	90	84	70	82	75	77	89

- ა) ააგეთ გაბნევის დიაგრამა და აღწერეთ დამოკიდებულება ამ ორ ცვლადს შორის.
- ბ) ააგეთ წრფივი რეგრესიის მოდელი და შეაფასეთ პროგნოზის შეცდომა.
- გ) მიღებული შედეგების გათვალისწინებით როგორ გარდაქმნილით დამოუკიდებელ ცვლადს?
5. საფეხბურთო კლუბის მენეჯერს სურს მრავლობითი რეგრესიის მოდელის აგება, რომელიც ახსნის კლუბის თამაშებზე დასწრებას. მან შეაგროვა საკუთარ მოედანზე ათი სეზონის 16 თამაშის მონაცემები.

Y – თამაშებზე დასწრება (game attendance),

X_1 – გუნდის მოგება/წაგების პროცენტი,

X_2 – მოწინააღმდეგეების მოგება/წაგების პროცენტი,

X_3 – სეზონში თამაშების რაოდენობა,

X_4 – ჰაერის ტემპერატურა თამაშის დროს.

Y	14502	12459	15600	16780	14600	19300
X_1	33.3	25.0	80.0	75.0	60.0	100.0
X_2	80.0	50.0	66.6	100.0	80.0	60.0
X_3	6	4	5	8	10	10
X_4	47	56	55	60	55	49

Y	14603	15789	17800	19450	13890
X_1	66.6	50.0	80.0	75.0	20.0
X_2	25.0	50.0	40.0	100.0	75.0
X_3	3	6	10	8	5
X_4	67	55	53	48	65

Y	15097	17666	12500	16780	17543
X_1	70.0	83.3	20.0	80.0	80.0
X_2	70.0	66.6	20.0	100.0	70.0
X_3	10	6	5	8	10
X_4	56	60	59	46	50

- ა) გამოთვალეთ კორელაციის მატრიცა;
 - ბ) ააგეთ ოთხი მარტივი რეგრესიის მოდელი თითოეული დამოუკიდებელი ცვლადით. აღნიშნეთ ოთხიდან რომელია საუკეთესო;
 - გ) ააგეთ მრავლობითი რეგრესიის მოდელი ოთხი დამოუკიდებელი ცვლადით;
 - დ) ცვალებადობის რა წილს ხსნის დამოკიდებულ ცვლადში მრავლობითი მოდელი?
 - ე) რომელი დამოუკიდებელი ცვლადია სტატისტიკურად მნიშვნელოვანი $\alpha = 0.05$ დონით?
 - ვ) ააგეთ 95%-იანი ნდობის ინტერვალი თითოეული რეგრესიის კოეფიციენტისათვის;
 - თ) არსებობს თუ არა მულტიკოლინეარულობის პრობლემა ამ ამოცანაში?
6. მენეჯერს სურს რეგრესიულ მოდელში ახალი ცვლადის დამატება, რომელიც იზომება სახელდების სკალაზე კატეგორიებით განათლების დონე: 1 – დაწყებითი; 2 – არასრული საშუალო; 3 – საშუალო ზოგადი; 4 – საშუალო პროფესიული.
- ა) რამდენი ფიქტიური ცვლადის გამოყენება დასჭირდება მენეჯერს?
 - ბ) განსაზღვრეთ ეს ცვლადები.

7. რეგრესიის განტოლება შემდეგი სახისაა:

$$Y = -23200 + 4.2X_1 + 2345X_2 + 4607X_3,$$

სადაც Y ფერმის წლიური შემოსავალია, X_1 დასამუშავებელი ფართობია, $X_2 = 1$ მინის სტრიქონური მორწყვაა, $X_2 = 0$ არაა მორწყვა, $X_3 = 1$ სპრინკლერ მორწყვაა, $X_3 = 0$ არაა მორწყვა. ახსენით თითოეული რეგრესიის კოეფიციენტი.

8. მიღებული მომგებიანობის ასახსნელად მიღებული შემდეგი მონაცემები:

Y	-2345	4200	278	1211	1406	500
X_1	-45	56	26	56	24	23
X_2	-1	2	3	2	2	3

Y	-700	3457	3478	1975	206
X_1	34	45	47	24	32
X_2	3	1	1	2	3

სადაც Y კლიენტისგან მიღებული სუფთა მოგებაა, X_1 კლიენტთან მუშაობის საათების რაოდენობაა, X_2 კლიენტის ტიპია: 1 – თუ კლიენტია წარმოების სფერო, 2 – თუ კლიენტია მომსახურების სფერო, 3 – თუ კლიენტი სამთავრობოა.

ა) ააგეთ სამი გაბნევის დიაგრამა თითოეული დამოუკიდებელი ცვლადით;

ბ) ააგეთ რეგრესიის განტოლება, თუ დამოუკიდებელი ცვლადია X_1 ;

გ) ააგეთ რეგრესიის განტოლება ორივე დამოუკიდებელი ცვლადით. შესაყვანი გაქნებათ ფიქტიური ცვლადები.

9. შეისწავლეს 25 კლიენტის (Y) შესყიდვების დამოკიდებულება (X_1) ასაკზე, (X_2) ოჯახის შემოსავალზე, (X_3) ოჯახის მოცულობაზე. მიიღეს კორელაციის მატრიცა:

	Y	X_1	X_2	X_3
Y	1.00	-0.4057	0.4591	-0.2444
X_1		1.00	0.0512	-0.5037
X_2			1.00	0.2718
X_3				1.00

შეამოწმეთ $\alpha = 0.05$ დონით რომელი კორელაციის კოეფიციენტია მნიშვნელოვანი.

10. რა პირადული მონაცემებით უნდა ხასიათდებოდეს მენეჯერი კარიერის გასაყე-
თებლად? გადაწყვეტილი იყო მრავლობითი რეგრესიის აგება სამუშაო რეიტინ-
გსა და კომუნიკაბელურობის უნარს შორის. კომუნიკაბელურობა შეაფასეს რამ-
დენიმე ფაქტორით. შეარჩიეს 15 საშუალო დონის მენეჯერი. მათ მისცეს გასაანა-
ლიზებლად რამდენიმე სამუშაო სიტუაცია და სთხოვეს წარედგინათ წერილობი-
თი და ვერბალური პრეზენტაცია.

სამუშაო რეიტინგი	შემთხვევათა ანალიზის ქულა	წერილობითი პრეზენტაციის ქულა	ვერბალური პრეზენტაციის ქულა
87	8.4	8.7	9.2
93	8.2	9.4	9.4
91	9.3	9.7	9.5
85	7.9	8.1	8.7
86	8.1	8.3	8.8
97	9.4	9.3	9.6
90	9.1	9.0	9.2
93	8.9	9.2	9.5
88	8.6	8.4	8.5
96	9.7	9.5	9.6
86	8.3	7.9	8.4
89	8.7	8.5	8.7
94	9.2	9.1	9.6
91	8.1	9.5	9.2
95	9.3	9.1	9.7

- გამოიყენეთ კომპიუტერული პროგრამა და შეამოწმეთ, არიან თუ არა ცვლადები კორელირებული ერთმანეთთან;
- შეამოწმეთ მოდელის ვარგისიანობა;
- თუ თქვენ საშუალო დონის მენეჯერი ბრძანდებით, თანახმა იქნებით თუ არა, თქვენი სამუშაო რეიტინგი შეფასდეს ამ მოდელით?
- სამუშაო რეიტინგის ცვალებადობის რამდენი პროცენტია ახსნილი სამი დამოუკიდებელი ცვლადით?
- არის თუ არა ყველა დამოუკიდებელი ცვლადი მნიშვნელოვანი ამ მოდელში?
- ერთ-ერთმა მენეჯერმა მიიღო შემდეგი ქულები: შემთხვევათა ანალიზის ქულა – 9.1, წერილობითი პრეზენტაციის ქულა – 9.4, ვერბალური პრეზენტაციის ქულა – 9.3. ამ მონაცემებზე დაყრდნობით, რას უდრის სამუშაო რეიტინგის ქულის საუკეთესო წერტილოვანი შეფასება?
- ღირს თუ არა ამ მოდელის გამოყენება ახალი თანამშრომლების მისაღებად?

11. გამომცემლობას სურს დაადგინოს მომზადებული ნიგნების საუკეთესო გასაყიდი ფასები. შეაგროვეს შემდეგი მონაცემები 15 გამოშვებულ ნიგნზე:

გაყიდი ნიგნების რაოდენობა, Y	გვერდების რაოდენობა, X_1	კონკურენტული ნიგნების რაოდენობა, X_2	სარეკლამო ბიუჯეტი, X_3	ავტორის ასაკი, X_4
15000	176	5	\$25000	49
140000	296	10	83000	57
75000	483	7	40000	29
100000	811	14	29000	37
26000	302	9	52000	35
33000	411	15	33000	43
59000	333	7	19000	51
103000	602	4	37000	62
88000	504	12	51000	33
10000	204	3	30000	50
9000	376	4	19000	26
77000	600	7	41000	40
59000	400	3	26000	44
183000	597	8	51000	59
16000	126	1	27000	38

- ა) გამოიყენეთ კომპიუტერული პროგრამა და შეამოწმეთ, რომელი დამოუკიდებელი ცვლადი კორელირებს მნიშვნელოვნად დამოუკიდებელ ცვლადთან;
 ბ) შეადგინეთ რეგრესიის განტოლება ოთხივე დამოუკიდებელი ცვლადით;
 გ) ცვალებადობის რამდენი პროცენტი აიხსნება ამ ოთხი ცვლადით?

ხშირად გამოყენებული ფორმულები. შეფასების და პირობებითა შემოწმების ამოცანათა მიმოხილვა

სიხშირეთა განაწილებები

ქულის ფარდობითი სიხშირე - $rf = f/n$, სადაც f - ქულის სიხშირეა, ხოლო n - შერჩევის მოცულობაა.

ქულის პროცენტული სიხშირე - $\%f = rf \cdot 100\%$.

ცენტრალური ტენდენციის საზომები

მედიანა. ვარიაციულ მწკრივში შუაში მდგომი მონაცემი, თუ მონაცემთა რაოდენობა კენტი და ორი შუაში მდგომი მონაცემის საშუალო არითმეტიკული, თუ მონაცემთა რაოდენობა ლუწია.

შერჩევის საშუალო. მონაცემთა საშუალო არითმეტიკული - $\bar{X} = \frac{\sum X}{n}$.

პოპულაციის საშუალო - $\mu = \frac{\sum X}{N}$, სადაც N - პოპულაციის მოცულობაა.

შერჩევის საშუალო - $\bar{X}$ პოპულაციის საშუალოს - μ -ს ნერტილოვანი შეფასებაა.

მოდა - ყველაზე ხშირად განმეორებადი მონაცემი.

ცვალებადობის საზომები

დიაპაზონი. $R = x_{\max} - x_{\min}$.

P -პროცენტილი. $\bar{x}_p = x_{([np]+1)}$, როცა $nP/100 \equiv np$ არ არის მთელი რიცხვი

და $\bar{x}_p = \frac{x_{(np)} + x_{(np+1)}}{2}$, როცა np მთელი რიცხვია.

პირველი კვარტილი - Q_1 . 25-პროცენტილი ანუ იმ მონაცემების მედიანა, რომლებიც არ აღემატება მედიანას.

მესამე კვარტილი - Q_3 . 75-პროცენტილი ანუ იმ მონაცემების მედიანა, რომლებიც მეტია ან ტოლი მედიანაზე.

კვარტილთშორისი გაბნევის დიაპაზონი - $IQR = Q_3 - Q_1$.

რანგი - r . მონაცემის ნომერი ვარიაციულ მწკრივში.

თავი 14. ხშირად გამოყენებული ფორმულები. შეფასების და პიპოთეზათა შემოწმების...

$$\text{პროცენტული რანგი} - \frac{2r-1}{2n} \cdot 100\%.$$

$$\text{შერჩევის დისპერსია} - S^2 = \frac{\sum (X - \bar{X})^2}{n-1} = \frac{1}{n-1} [\sum X^2 - n(\bar{X})^2].$$

$$\text{პოპულაციის დისპერსია} - \sigma^2 = \frac{\sum (X - \mu)^2}{N}.$$

შერჩევის დისპერსია $-S^2$ პოპულაციის დისპერსიის $-\sigma^2$ -ის წერტილოვანი შეფასებაა.

$$\text{შერჩევის სტანდარტული გადახრა} - S = \sqrt{S^2}.$$

$$\text{პოპულაციის სტანდარტული გადახრა} - \sigma = \sqrt{\sigma^2}.$$

შერჩევის სტანდარტული გადახრა $-S$ პოპულაციის სტანდარტული გადახრის $-\sigma$ -ს წერტილოვანი შეფასებაა.

***Z* -ქულები**

$$\text{ერთეული სტანდარტული ქულა} - z = \frac{X - \mu}{\sigma}.$$

z ქულა გვიჩვენებს, რამდენი სტანდარტული გადახრით (რამდენი σ -თი) არის X ქულა დაშორებული μ საშუალოსგან.

$$\text{სტანდარტული ქულა} - z = \frac{X - \bar{X}}{S} = \frac{X - \bar{X}}{\sqrt{\sum (X - \bar{X})^2 / n}}.$$

$$\text{შერჩევის სტანდარტული ქულა} - z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}.$$

დიდი შერჩევები

შერჩევა დიდია, თუ მისი მოცულობა $n \geq 30$.

კრიტერიუმი საშუალოსათვის

დაშვებულია, რომ $\sigma \approx s$, სადაც σ - პოპულაციის სტანდარტული გადახრაა, s - შერჩევის სტანდარტული გადახრა. კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$Z = \frac{\bar{X} - \mu}{s / \sqrt{n}}.$$

Z სტატისტიკა მიახლოებით ნორმალურადაა განაწილებული ნულის ტოლი საშუალოთი და ერთის ტოლი სტანდარტული გადახრით, $N(0,1)$. 95% სანდოობით, პოპულაციის უცნობი μ საშუალო მოთავსებულია ორ საზღვარს შორის:

$$\bar{x} \pm z_{\alpha} \frac{s}{\sqrt{n}}, \text{ სადაც } z_{\alpha} = 1.96.$$

კრიტერიუმი ორი საშუალოსათვის

$\sigma \approx s$ დაშვებულია თითოეული პოპულაციისათვის. თუ შერჩევები დამოუკიდებელია, კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

სადაც n_1 – პირველი შერჩევის მოცულობაა, n_2 – მეორე შერჩევის მოცულობაა. ეს სტატისტიკა მიახლოებით ნორმალურადაა განაწილებული $N(0,1)$.

95%-ანი სანდოობით, საშუალოების სხვაობა $\mu_1 - \mu_2$ მოთავსებულია შემდეგ საზღვრებში:

$$\bar{x}_1 - \bar{x}_2 \mp 1.96 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

კრიტერიუმი პროპორციისათვის

$\hat{p} = \frac{X}{n}$ – შერჩევის პროპორცია (ფარდობითი სიხშირე), სადაც X – წარმატებათა რაოდენობა n დამოუკიდებელ ცდაში. აუცილებელია, რომ შესრულდეს შემდეგი ორი პირობა:

$$0 < \hat{p} \pm 3\sqrt{\hat{p}(1-\hat{p})/n} < 1 \quad \text{და} \quad 0 < p \pm 3\sqrt{p(1-p)/n} < 1.$$

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}.$$

ეს სტატისტიკა მიახლოებით ნორმალურადაა განაწილებული $N(0,1)$.

95% -ანი სანდოობით, პოპულაციის პროპორცია მოთავსებულია შემდეგ საზღვრებში:

$$\hat{p} \mp 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

კრიტერიუმი ორი პროპორციისათვის

საჭიროა შემდეგი პირობების შესრულება

$$0 < \hat{p}_1 \pm 3\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} < 1 \quad \text{და} \quad 0 < \hat{p}_2 \pm 3\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < 1,$$

თავი 14. ხშირად გამოყენებული ფორმულები. შეფასების და პიპოთეზათა შემოწმების...

$$0 < p_1 \pm 3 \sqrt{\frac{p_1(1-p_1)}{n_1}} < 1 \quad \text{და} \quad 0 < p_2 \pm \sqrt{\frac{p_2(1-p_2)}{n_2}} < 1.$$

თუ $H_0: p_1 = p_2$ კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$Z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

სადაც

$$\hat{p} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2}.$$

ეს სტატისტიკა მიახლოებით ნორმალურადაა განაწილებული $N(0,1)$.

95%-ანი სანდოობით, პროპორციების სხვაობა $p_1 - p_2$ მოთავსებულია შემდეგ საზღვრებში

$$\bar{p}_1 - \bar{p}_2 \mp 1.96 \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}.$$

პატარა შერჩევები ნორმალური პოპულაციიდან

იგულისხმება, რომ შერჩევების მოცულობებიდან ერთი მაინც < 30 -ზე და ორივე შერჩევა მიღებულია ნორმალური პოპულაციებიდან. ამის შემოწმება შეიძლება ჰისტოგრამის ან ბოქსპლოტის აგებით.

კრიტერიუმი საშუალოსათვის

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}.$$

ამ სტატისტიკას აქვს t - განაწილება თავისუფლების ხარისხით $n-1$.

95%-იანი სანდოობით, პოპულაციის უცნობი საშუალო μ მოთავსებულია ორ საზღვარს შორის:

$$\bar{x} \mp t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}},$$

სადაც $t_{n-1, \alpha/2}$ მოსაძებნია t - განაწილების ცხრილიდან თავისუფლების $n-1$ ხარისხის გათვალისწინებით.

კრიტერიუმი დისპერსიისათვის

სტატისტიკური ჰიპოთეზის შემოწმება ნორმალური $N(a, \sigma^2)$ განაწილების დისპერსიის შესახებ:

ჰიპოთეზა: $H_0 : \sigma^2 = \sigma_0^2$

მნიშვნელოვნების დონე: α

კრიტერიუმის სტატისტიკა:

$$\begin{cases} \frac{\sum_{i=1}^n (X_i - a)^2}{\sigma_0^2} \equiv \chi^2(n), \text{ თუ } a \text{ ცნობილია,} \\ \frac{(n-1)S^2}{\sigma_0^2} \equiv \chi^2(n-1), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

ალტერნატივა

კრიტიკული არე C.R. (H_0 -ის უარყოფის არე)

$H_A : \sigma^2 \neq \sigma_0^2$

$$\begin{cases} (0, \chi^2_{n-1-\alpha/2}] \cup [\chi^2_{n-1-\alpha/2}, +\infty), \text{ თუ } a \text{ ცნობილია,} \\ (0, \chi^2_{n-1-\alpha/2}] \cup [\chi^2_{n-1-\alpha/2}, +\infty), \text{ თუ } a \text{ უცნობია.} \end{cases}$$

მნიშვნელოვნების დონის მქონე ნდობის ინტერვალის ნორმალური $N(a, \sigma^2)$ პოპულაციის დისპერსიისათვის უცნობი საშუალოს შემთხვევაში:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1, \alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1, 1-\alpha/2}} \right),$$

სადაც $\chi^2_{m, \alpha}$ - თავისუფლების m ხარისხის მქონე ხი-კვადრატ განაწილების $(\chi^2(m))$ ზედა α - კრიტიკული წერტილია.

კრიტერიუმი ორი საშუალოსათვის

იგულისხმება, რომ $\sigma_1^2 = \sigma_2^2 = \sigma^2$ - უცნობია. ამ საერთო დისპერსიის წერტილოვანი შეფასებაა s_p^2 :

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

სადაც n_1 და s_1^2 - შესაბამისად, პირველი შერჩევის მოცულობა და დისპერსია; n_2 და s_2^2 - მეორე შერჩევის მოცულობა და დისპერსიაა.

თუ შერჩევები ერთმანეთისაგან დამოუკიდებელია, მაშინ კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \equiv t(n_1 + n_2 - 2).$$

ამ სტატისტიკას აქვს t - განაწილება თავისუფლების ხარისხით $n_1 + n_2 - 2$.

თავი 14. ხშირად გამოყენებული ფორმულები. შეფასების და პიპოთეზათა შემოწმების...

95%-ანი სანდოობით, საშუალოების სხვაობა მოთავსებულია შემდეგ საზღვრებში:

$$\bar{x}_1 - \bar{x}_2 \mp t_{n_1+n_2-2, \alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

სადაც $t_{n_1+n_2-2, \alpha/2}$ თავისუფლების $n_1 + n_2 - 2$ ხარისხის მქონე სტიუდენტის განაწილების ზედა $\alpha/2$ -კრიტიკული წერტილია.

თუ შერჩევები ერთმანეთზე დამოკიდებულია, განიხილება სხვაობები $d_i = x_i - y_i$, სადაც $x_i - i$ -ური მონაცემია პირველი შერჩევიდან, $y_i - i$ -ური მონაცემია მეორე შერჩევიდან. n სხვაობის საშუალოა $\bar{X}$, პოპულაციის სხვაობების საშუალოა μ_d ; სხვაობების სტანდარტული გადახრაა s , პოპულაციის სტანდარტული გადახრაა σ_d . იგულისხმება, რომ $\sigma_d \approx s$.

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$T = \frac{\bar{X} - \mu_d}{\frac{s}{\sqrt{n}}} = \frac{\bar{X}_1 - \bar{X}_2 - \mu_d}{\sqrt{\frac{\sum d^2 - (\sum d)^2 / n}{n(n-1)}}},$$

სადაც n - წყვილების რაოდენობაა. ამ სტატისტიკას აქვს t -განაწილება თავისუფლების ხარისხით $n - 1$.

95%-იანი სანდოობით, პოპულაციის საშუალო მოთავსებულია შემდეგ საზღვრებში:

$$\bar{x} \mp t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}},$$

სადაც $t_{n-1, \alpha/2}$ თავისუფლების $n - 1$ ხარისხის მქონე სტიუდენტის განაწილების ზედა $\alpha/2$ -კრიტიკული წერტილია.

კრიტერიუმი ორი დისპერსიისათვის

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$F = \frac{s_1^2}{s_2^2},$$

სადაც $s_1^2 > s_2^2$. ამ სტატისტიკას აქვს F -განაწილება $n_1 - 1$ და $n_2 - 1$ თავისუფლების ხარისხებით.

დისპერსიათა σ_1^2 / σ_2^2 შეფარდების $(1 - \alpha) \cdot 100\%$ -იანი ნდობის ინტერვალის მოიკება ფორმულით:

$$\left(\frac{1}{F_{n_2-1, n_1-1, \alpha/2}} \cdot \frac{s_1^2}{s_2^2}, F_{n_1-1, n_2-1, \alpha/2} \cdot \frac{s_1^2}{s_2^2} \right),$$

გამოყენებითი სტატისტიკის საფუძვლები

სადაც $F_{n,m,\alpha}$ თავისუფლების n და m ხარისხების მქონე ფიშერის განაწილების ზედა α -კრიტიკული ნერტილია.

კრიტერიუმი ორზე მეტი საშუალოსათვის

იგულისხმება, რომ k შერჩევის დისპერსია მნიშვნელოვნად არ განსხვავდება ერთმანეთისაგან. კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$F = \frac{MSR}{MSE},$$

სადაც

$$MSR = \frac{\sum \sum (\bar{X}_i - \bar{\bar{X}})^2}{k-1} \quad \text{და} \quad MSE = \frac{\sum \sum (X - \bar{X})^2}{N-k},$$

$\bar{X}_i - i$ -ური შერჩევის საშუალოა, $\bar{\bar{X}}$ - გაერთიანებული შერჩევების საშუალოა, k - შერჩევების რაოდენობა, N - მონაცემთა სრული რაოდენობა.

სტატისტიკას აქვს F -განაწილება $k-1$ და $N-k$ თავისუფლების ხარისხებით.

პატარა შერჩევები, როცა პოპულაციის განაწილება არაა ნორმალური

თუ შერჩევები ერთმანეთისაგან დამოუკიდებელია, მან-უიტნის U კრიტერიუმს აქვს შემდეგი სახე:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} \quad \text{მინუს პირველი შერჩევის რანგების ჯამი,}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} \quad \text{მინუს მეორე შერჩევის რანგების ჯამი,}$$

$$U = \min(U_1, U_2).$$

თუ შერჩევები დამოკიდებულია ერთმანეთზე, უილკოქსონის T კრიტერიუმს აქვს შემდეგი სახე:

$$T = \text{უმცირესი } \sum R_- \text{-სა და } \sum R_+ \text{-ს შორის,}$$

სადაც $\sum R_-$ - უარყოფითი სხვაობების რანგების ჯამია, $\sum R_+$ - დადებითი სხვაობების რანგების ჯამია.

თუ შერჩევების რაოდენობა $k > 2$.

კრუსკალ-უოლისის H კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1).$$

ამ სტატისტიკას აქვს ხი-კვადრატ განაწილება თავისუფლების ხარისხით $k-1$.

კავშირი ორ ცვლადს შორის. კრიტერიუმი ორი კატეგორიული ცვლადისთვის

იგულისხმება, რომ კატეგორიები ურთიერთგამომრიცხავია და მოსალოდნე-
ლი სიხშირეების, სულ ცოტა, 80% 5-ზე მეტია.

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}.$$

ამ სტატისტიკას აქვს ხი-კვადრატ განაწილება თავისუფლების ხარისხით
($r-1$)($c-1$).

თუ $r = c = 2$, გამოსაყენებელია მამრავლი უწყვეტობაზე და კრიტერიუმის
სტატისტიკას აქვს შემდეგი სახე:

$$\chi^2 = \sum \frac{(|o_{ij} - e_{ij}| - 0.5)^2}{e_{ij}}.$$

ამ სტატისტიკას აქვს ხი-კვადრატ განაწილება 1-ის ტოლი თავისუფლების ხარი-
სხით.

ორი რაოდენობრივი ცვლადის წრფივი კავშირის კრიტერიუმი

კორელაციური კვლევის მიზანია ორ ან მეტ ცვლადს შორის კავშირის ხარის-
ხის დადგენა. დადებითი კავშირის შემთხვევაში, X ცვლადის მნიშვნელობის
ზრდასთან ერთად იზრდება Y ცვლადის მნიშვნელობაც. უარყოფითი კავშირის
შემთხვევაში, X ცვლადის მნიშვნელობის ზრდასთან ერთად Y ცვლადის მნიშ-
ვნელობა მცირდება.

თუ ორივე ცვლადი გაზომილია ინტერვალების სკალაზე, კავშირის მიმართუ-
ლების და ხარისხის დასადგენად გამოიყენება პირსონის კორელაციის კოეფიცი-
ენტი:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}.$$

პოპულაციაში X და Y ცვლადებს შორის კავშირი წარმოდგენილია ρ -თი. თუ
(X, Y) სიდიდეების განაწილება ნორმალურია, მაშინ H_0 -ის სამართლიანობისას,
 $H_0: \rho = 0$, სტატისტიკას

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

აქვს t - განაწილება თავისუფლების ხარისხით $n-2$.

დეტერმინაციის კოეფიციენტი r^2 გამოხატავს X ცვლადის ცვალებადობის იმ პროპორციას, რომელიც დაკავშირებულია Y ცვლადის ცვალებადობასთან. წრფივი კავშირი ორ ცვლადს შორის მოცემულია განტოლებით:

$$Y = a + bX,$$

სადაც

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}, \quad a = \bar{Y} - b\bar{X}.$$

კრიტერიუმის სტატისტიკას აქვს შემდეგი სახე:

$$T = \frac{b - \beta}{\frac{\hat{\sigma}}{\sqrt{\sum (X - \bar{X})^2}}},$$

სადაც

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (Y - \bar{Y})^2 - b \sum (X - \bar{X})(Y - \bar{Y})}{n-2}.$$

T სტატისტიკას აქვს t - განაწილება თავისუფლების ხარისხით $n-2$.

EXCEL-ის სტატისტიკური ფუნქციების აღწერა

ალბათობა

ფაქტორიალი – $FACT(n) = n!$;

წყობა – $PERMUT(n,k) = n!/(n-k)!$;

ჯუფდება – $COMBIN(n,k) = n!/[k!(n-k)!]$;

EXCEL-ში დისკრეტული განაწილებების ჩანერისას გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
x	განაწილების სიმკვრივის ან განაწილების ფუნქციის არგუმენტის რიცხვითი მნიშვნელობა
X range	არე, სადაც მითითებულია შემთხვევითი სიდიდის შესაძლო მნიშვნელობები
Prob-range	არე, სადაც მითითებულია შემთხვევითი სიდიდის შესაძლო მნიშვნელობების ალბათობები
Lower-limit	ქვედა საზღვარი
Upper-limit	ზედა საზღვარი
Mean	განაწილების მათემატიკური ლოდინი
Numbers	რაოდენობა (ნარმატებათა)
Trials	ცდათა სერიის სიგრძე
Probability-s	ალბათობა
Sample-s	1 სახის ობიექტთა რაოდენობა შერჩევაში
Number-sample	შემთხვევით (დაბრუნების გარეშე) ამორჩეულ ობიექტთა რაოდენობა – შერჩევის მოცულობა
Population-s	1 სახის ობიექტთა რაოდენობა პოპულაციაში
Number-pop.	ობიექტების რაოდენობა პოპულაციაში
Cumulative	თუ Cumulative პარამეტრის რიცხვითი მნიშვნელობა ნულის ტოლია, მაშინ გამოიანგარიშება $P\{X = k\}$ ალბათობა; თუ Cumulative პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, მაშინ გამოიანგარიშება $P\{X \leq k\}$ ალბათობა

PROB(x-range,prob-range,lower-limit,upper-limit) – ფუნქციით გამოითვლება ალბათობა: $P\{Lower\ limit \leq X \leq Upper\ limit\}$; მაგალითად, თუ X შემთხვევითი სიდიდის კანონი EXCEL-ში არის:

	A	B	C	D	E	F
1	-1	0	2	4	6	9
2	0.05	0.15	0.18	0.22	0.3	0.1

მაშინ $\text{PROB}(A1:F1, A2:F2, 0, 7)$ – გამოითვლის $P(0 \leq X \leq 5) = 0.45$ -ს.

$\text{BINOMDIST}(\text{number}-s, \text{trials}, \text{probability}, \text{cumulative})$ – ფუნქციით გამოითვლება ბინომურად განაწილებული შემთხვევითი სიდიდისათვის $P\{X = k\}$ ალბათობა, თუ $\text{cumulative} = 0$ -ს და $P\{X \leq k\}$ ალბათობა, თუ $\text{cumulative} = 1$ -ს.

$\text{CRITBINOM}(\text{trials}, \text{probability}-s, \alpha)$ – ფუნქციით, მოცემული α -სათვის, გამოითვლება ის მინიმალური k , რომლისთვისაც: $\sum_{m=0}^k P_n(m) \geq \alpha$.

$\text{HYPGEOMDIST}(\text{sample}-s, \text{number}-\text{sample}, \text{population}-s, \text{number}-\text{pop.})$ – ფუნქციით გამოითვლება ჰიპერგეომეტრიული განაწილებისათვის $P\{X = k\}$ ალბათობა. მაგალითად, თუ ყუთში 100 ბურთია, რომელთა შორის 30 თეთრია და შემთხვევით, დაბრუნების გარეშე, ვიღებთ 45 ბურთს, მაშინ ალბათობა იმისა, რომ მათ შორის აღმოჩნდება 10 თეთრი ბურთი, გამოითვლება შემდეგნაირად:

$$\text{HYPGEOMDIST}(10, 40, 30, 100) = \frac{C_{30}^{10} \cdot C_{70}^{35}}{C_{100}^{45}}.$$

$\text{POISSON}(x, \text{mean}, \text{cumulative})$ – ფუნქციით გამოითვლება პუასონის კანონით განაწილებული შემთხვევითი სიდიდისათვის $P\{X = k\}$ ალბათობა, თუ $\text{cumulative} = 0$ -ს და $P\{X \leq k\}$ ალბათობა, თუ $\text{cumulative} = 1$ -ს.

EXCEL-ში უწყვეტი განაწილებების ჩანერგვის გამოყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
x	განაწილების სიმკვრივის ან განაწილების ფუნქციის არგუმენტის რიცხვითი მნიშვნელობა
Mean	განაწილების მათემატიკური ლოდინი
Standard-dev	განაწილების საშუალო კვადრატული გადახრა
Cumulative	თუ Cumulative პარამეტრის რიცხვითი მნიშვნელობა ნულის ტოლია, მაშინ გამოითვლება განაწილების სიმკვრივის რიცხვითი მნიშვნელობა; თუ Cumulative პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, მაშინ გამოითვლება განაწილების ფუნქციის რიცხვითი მნიშვნელობა
Lambda	λ პარამეტრი
Probability	ალბათობის P დონე, რომლისთვისაც გამოითვლება კვანტილი
Deg-freedom	განაწილების თავისუფლების ხარისხი
Probability	$P\{X > k\}$ ალბათობის მნიშვნელობა
Tails	თუ Tails პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, მაშინ TDIST ფუნქცია გვაძლევს ცალმხრივ $T(u) > x$ არეში მოხვედრის ალბათობას, $P\{T(u) > x\}$, სტიუდენტის განაწილებისათვის; თუ Tails პარამეტრის რიცხვითი მნიშვნელობა ორის ტოლია, მაშინ TDIST ფუნქცია გვაძლევს ორმხრივ $ T(u) > x$ არეში მოხვედრის ალბათობას, $P\{ T(u) > x\}$ -ს, სტიუდენტის განაწილებისათვის

NORMSDIST(x) – ფუნქციით გამოითვლება სტანდარტული ნორმალური $N(0,1)$ შემთხვევითი სიდიდისათვის $P\{N(0,1) \leq x\}$ ალბათობა.

NORMDIST(x,mean,standard_dev,cumulative) – ფუნქციით, თუ **cumulative** პარამეტრის რიცხვითი მნიშვნელობა ნულის ტოლია, გამოითვლება $N(\mu, \sigma^2)$ -ის (არასტანდარტული ნორმალური) განაწილების სიმკვრივის მნიშვნელობა; ხოლო, თუ **cumulative** პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, გამოითვლება $N(\mu, \sigma^2)$ -ის განაწილების ფუნქციის მნიშვნელობა.

NORMSINV(probability) – ფუნქციით გამოითვლება სტანდარტული ნორმალური განაწილების p დონის კვანტილი. მაგალითად, **NORMSINV(0.65)** = $x_{0.65} = 0.39$.

NORMINV(probability,mean,standard_dev) – ფუნქციით გამოითვლება არასტანდარტული ნორმალური განაწილების p დონის კვანტილი.

EXPONDIST(x,lambda,cumulative) – ითვლის $\lambda e^{-\lambda x}$ ფუნქციის მნიშვნელობას, როცა **Cumulative**=0-ს და $1 - e^{-\lambda x}$ ფუნქციის მნიშვნელობას, როცა **Cumulative** = 1-ს.

CHIDIST(x,deg_freedom) – ითვლის $P\{\chi^2(k) > x\}$ ალბათობას ხი-კვადრატ განაწილებისათვის.

CHIINV(probability,deg_freedom) – ითვლის ხი-კვადრატ განაწილების ზედა კრიტიკულ წერტილს, ანუ ისეთ x რიცხვს, რომლისთვისაც $P\{\chi^2(Deg\ freedom) > x\} = Probability$.

TDIST(x,deg_freedom,tails): TDIST(x,k,1) – ითვლის $P\{T(k) > x\}$ ალბათობას სტიუდენტის $T(k)$ შემთხვევითი სიდიდისათვის, ხოლო **TDIST(x,k,2)** – კი ითვლის $P\{|T(k)| > x\}$ ალბათობას სტიუდენტის $T(k)$ შემთხვევითი სიდიდისათვის.

TINV(probability,deg_freedom): TINV(a,k) – პოულობს ისეთ x -ს, რომლისთვისაც $P\{|T(k)| > x\} = a$ (ფაქტობრივად x იქნება სტიუდენტის $T(k)$ შემთხვევითი სიდიდის ზედა $a/2$ -კრიტიკული წერტილი).

FDIST(x,deg_freedom1,deg_freedom2): FDIST(x,n,m) – ითვლის $P\{F(n,m) > x\}$ ალბათობას ფიშერის $F(n,m)$ შემთხვევითი სიდიდისათვის.

FINV(probability,deg_freedom1,deg_freedom2): FINV(a,n,m) – პოულობს ფიშერის $F(n,m)$ განაწილების ზედა a -კრიტიკულ წერტილს, ანუ ისეთ x -ს, რომლისთვისაც $P\{F(n,m) > x\} = a$.

სტატისტიკა

EXCEL-ის ქვეპროგრამა **Tools/Data Analysis/Random Number Generation** საშუალებას იძლევა მივიღოთ ფსევდოშემთხვევითი რიცხვების რეალიზაციები. ამ დროს გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Number of variables	დაკვირვებულ სერიათა რაოდენობა
Number of Random Numbers	მონაცემთა რაოდენობა სერიაში
Distribution	განაწილების კანონი
Discrete	დისკრეტული
Value and Probability Input Range	ეთითება ვერტიკალურად ჩანერგილი დისკრეტული განაწილების კანონი, რომლის მიხედვითაც უნდა გათამაშდეს ფსევდოშემთხვევითი სიდიდეები
Uniform	თანაბარი
Parameters	ეთითება თანაბარი განაწილების a და b პარამეტრები
Normal	ნორმალური
Mean	საშუალო
Standard Deviation	სტანდარტული გადახრა
Bernouli	ბერნული
Binomial	ბინომური
p-Value	p პარამეტრის რიცხვითი მნიშვნელობა
Number of Trials	n პარამეტრის რიცხვითი მნიშვნელობა ბერნულის სქემაში
Poisson	პუასონი
Lambda	λ პარამეტრი

CONFIDENCE(alpha,standard_dev,size): CONFIDENCE(α,σ,n) – ნორმალური პოპულაციის საშუალოსათვის ითვლის $(1-\alpha)$ საიმედოობის ნდობის ინტერვალის სიგრძის ნახევარს – $z_{\alpha/2} \cdot \sigma / \sqrt{n}$ -ს.

***p*-მნიშვნელობა**

ZTEST ფუნქცია გამოიყენება Z სტატისტიკის p -მნიშვნელობის გამოსათვლელად. ამ დროს გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Array	დაკვირვებულ მონაცემთა არე
x	პოპულაციის საშუალო
Sigma	პოპულაციის სტანდარტული გადახრა

ZTEST(array,x,sigma) – Z სტატისტიკის დაკვირვებული $z = \frac{\bar{x} - a_0}{\sigma / \sqrt{n}}$ მნიშვნელობისათვის

(მარჯვენა ცალმხრივი კრიტერიუმის შემთხვევაში) ითვლის p -მნიშვნელობას, ანუ

$$P\left\{Z \geq \frac{\bar{x} - a_0}{\sigma / \sqrt{n}}\right\} \text{ ალბათობას.}$$

CHITEST ფუნქცია გამოიყენება ბი-კვადრატ განაწილების p -მნიშვნელობის გამოსათვლელად. ამ დროს გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Actual-range	დაკვირვებული სიხშირეები
Expected-range	მოსალოდნელი სიხშირეები

CHITEST(actual_range, expected_range) – ითვლის $(R-1)(C-1)$ თავისუფლების ხარისხის მქონე ბი-კვადრატ განაწილების p -მნიშვნელობას.

ორამოკრეფიანი ამოცანები

TTEST(array1, array2, tails, type) – ითა და **FTEST(array1, array2)** – ით სარგებლობის დროს გამოიყენება აღნიშვნები:

აღნიშვნა	შინაარსი
Array	არე
Tails	კრიტიკული არე
Type	ტიპი

და ქვეპროგრამები:

- ❖ **z-Test: Two Sample for Means** – Z-ტესტი (ორამოკრეფიან ამოცანაში) საშუალოების შესახებ;
- ❖ **t-Test: Paired Two Sample for Mean** – T-ტესტი დანყვილებულ მონაცემთა საშუალოებისათვის;
- ❖ **t-Test: Two Sample Assuming Equal Variances** – T-ტესტი უცნობი ტოლი დისპერსიების შემთხვევაში;
- ❖ **t-Test: Two Sample Assuming Unequal Variances** – T-ტესტი უცნობი არატოლი დისპერსიების შემთხვევაში;
- ❖ **F-Test: Two Sample for Variances** – F-ტესტი დისპერსიებისათვის.

ამ ქვეპროგრამებით სარგებლობისას გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Input	შემავალი მონაცემები
Variable 1(2) Range	დაკვირვებათა არე
Variance (known)	პოპულაციითა დისპერსიების მნიშვნელობები
Labels	მონიშვნა
Hypothesized Mean Difference	ძირითადი ჰიპოთეზით განსაზღვრული პოპულაციითა საშუალოების სხვაობა (ანუ ნული)
Alpha	მნიშვნელოვნების დონე (0.05)

z-Test-ის გამოყენების მაგალითი. ორი დამოუკიდებელი ნორმალური პოპულაციიდან დისპერსიებით, შესაბამისად, 5 და 9, მიღებულია შემდეგი შერჩევები: I) 2.4, -2.3, 5.2, 10.3, 9.9, 12.6, -6.9, 2.8, 9.4, -1.4 და II) -8.2, 17.4, 16.4, 4, 21.7, -3.2, -2.9. 0.05 მნიშვნელოვნების დონით შევამოწმოთ ჰიპოთეზა საშუალოთა ტოლობის შესახებ.

ამოსხნა. მოწმდება ძირითადი ჰიპოთეზა $H_0 : \mu_1 - \mu_2 = 0$ ცალმხრივი $H_1 : \mu_1 - \mu_2 < 0$ და ორმხრივი $H_1 : \mu_1 - \mu_2 \neq 0$ ალტერნატივების წინააღმდეგ. ქვეპროგრამა **z-Test: Two Sample for Means** გვაძლევს:

	I	II
Mean/ საშუალო	4.2	6.5
known Variance/ ცნობილი დისპერსიები	5	9
Observations / დაკვირვებები	10	7
Hypothesized Mean Difference / ძირითადი ჰიპოთეზით განსაზღვრული საშუალოთა სხვაობა	0	
z Stat / Z-სტატისტიკის დაკვირვებული მნიშვნელობა	-1.689	
P(Z ≤ z) one-tail / p-მნიშვნელობა ცალმხრივი კრიტიკული არისათვის	0.046	
z Critical one-tail / კრიტიკული ნერტილი ცალმხრივი კრიტიკული არისათვის	1.645	
P(Z ≤ z) two-tail / p-მნიშვნელობა ორმხრივი კრიტიკული არისათვის	0.091	
z Critical two-tail / კრიტიკული ნერტილი ორმხრივი კრიტიკული არისათვის	1.96	

დასკვნა:

- ❖ ცალმხრივი $H_1 : \mu_1 - \mu_2 < 0$ ალტერნატივისათვის p -მნიშვნელობა $= 0.046 < \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით ძირითად ჰიპოთეზას უარყოფთ;
- ❖ ორმხრივი $H_1 : \mu_1 - \mu_2 \neq 0$ ალტერნატივისათვის p -მნიშვნელობა $= 0.091 > \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.

t-Test-ის გამოყენების მაგალითი. კომპანიის მენეჯერს სურს შეადაროს ორი *A* და *B* შეთანხმებიდან შემოსული თვიური შემოსავლები. *A* შეთანხმება გულისხმობს დაბალ საარენდო გადასახადს, მაგრამ ავალდებულებს არენდატორს დაზიანებული საარენდო ქონების აღდგენის ხარჯების დაფარვას. *B* შეთანხმება ადგენს მაღალ საარენდო გადასახადს, მაგრამ კომპანია თავის თავზე იღებს დაზიანებული საარენდო ქონების აღდგენის ხარჯებს. მენეჯერმა შეაგროვა თითოეული შეთანხმების თვიური შემოსავლები და $\alpha = 0.05$ მნიშვნელოვნების დონით სურს შეამოწმოს ჰიპოთეზა, რომ ამ შეთანხმებების საშუალო შემოსავლები არ განსხვავდება ერთმანეთისაგან. ეს მონაცემებია:

<i>A</i> შეთანხმება	84.5	102.6	79.1	92.5	101.7	78.7	88.5	93	72.2	96.8	92.6	93.7
<i>B</i> შეთანხმება	106	102.1	105.1	91.5	82.3	88.4	101.1	97.5	75.7	93.6		

ამოხსნა 1. ვიგულისხმობთ, რომ ორივე პოპულაცია ნორმალურადაა განაწილებული და უცნობი დისპერსიები ტოლია. უნდა შემოწმდეს ძირითადი ჰიპოთეზა $H_0: \mu_1 - \mu_2 = 0$ ცალმხრივი $H_1: \mu_1 - \mu_2 < 0$ და ორმხრივი $H_1: \mu_1 - \mu_2 \neq 0$ ალტერნატივების წინააღმდეგ. ქვეპროგრამა **t-Test: Two Assuming Equal Variances** გვაძლევს:

	A	B
Mean/ საშუალო	89.658	94.33
Variance / დისპერსია	88.366	100.438
Observations / დაკვირვებები	12	10
Pooled Variance/ დანეწილებული მონაცემების დისპერსია	93.799	
Hypothesized Mean Difference / ძირითადი ჰიპოთეზით განსაზღვრული საშუალოთა სხვაობა	0	
df / თავისუფლების ხარისხი	20	
t Stat /T-სტატისტიკის დაკვირვებული მნიშვნელობა	-1.1266	
P{T≤t} one-tail / p-მნიშვნელობა ცალმხრივი კრიტიკული არისათვის	0.1366	
Critical one-tail / კრიტიკული ნერტილი ცალმხრივი კრიტიკული არისათვის	1.7247	
P{T≤t} two-tail / p-მნიშვნელობა ორმხრივი კრიტიკული არისათვის	0.2733	
Critical two-tail / კრიტიკული ნერტილი ორმხრივი კრიტიკული არისათვის	2.086	

დასკვნა:

- ❖ ცალმხრივი $H_1: \mu_1 - \mu_2 < 0$ ალტერნატივისათვის p -მნიშვნელობა = 0.1366 > $\alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს;
- ❖ ორმხრივი $H_1: \mu_1 - \mu_2 \neq 0$ ალტერნატივისათვის p -მნიშვნელობა = 0.2733 > $\alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.

შენიშვნა. სტატისტიკური ფუნქციით **TTEST(array1,array2,1,2)** - გამოითვლება სტიუდენტის განაწილების p -მნიშვნელობა ცალმხრივი ალტერნატივისათვის, ხოლო **TTEST(array1,array2,2,2)**-ით კი p -მნიშვნელობა ორმხრივი ალტერნატივისათვის.

ამოხსნა 2. ვიგულისხმობთ, რომ ორივე პოპულაცია ნორმალურადაა განაწილებული და უცნობი დისპერსიები არაა ტოლი. ვამოწმებთ ძირითად ჰიპოთეზას $H_0: \mu_1 - \mu_2 = 0$ ცალმხრივი $H_1: \mu_1 - \mu_2 < 0$ და ორმხრივი $H_1: \mu_1 - \mu_2 \neq 0$ ალტერნატივების წინააღმდეგ. ქვეპროგრამა **t-Test: Two Assuming Unequal Variances** გვაძლევს:

	A	B
Mean/ საშუალო	89.658	94.33
Variance/ დისპერსია	88.366	100.438
Observations / დაკვირვებები	12	10
Hypothesized Mean Difference / ძირითადი ჰიპოთეზით განსაზღვრული საშუალოთა სხვაობა	0	
df / თავისუფლების ხარისხი	19	
t Stat / T-სტატისტიკის დაკვირვებული მნიშვნელობა	-1.1197	
P(T≤t) one-tail / p-მნიშვნელობა ცალმხრივი კრიტიკული არისათვის	0.1384	
Critical one-tail / კრიტიკული ნერტილი ცალმხრივი კრიტიკული არისათვის	1.7291	
P(T≤t) two-tail / p-მნიშვნელობა ორმხრივი კრიტიკული არისათვის	0.2768	
Critical two-tail / კრიტიკული ნერტილი ორმხრივი კრიტიკული არისათვის	2.093	

დასკვნა:

- ❖ ცალმხრივი $H_1: \mu_1 - \mu_2 < 0$ ალტერნატივისათვის p -მნიშვნელობა $= 0.1384 > \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს;
- ❖ ორმხრივი $H_1: \mu_1 - \mu_2 \neq 0$ ალტერნატივისათვის p -მნიშვნელობა $= 0.2768 > \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.

შენიშვნა 1. როგორც ხედავთ, ამოხსნა 1-საგან განსხვავებით, აქ ამოიბეჭდა სხვა თავისუფლების ხარისხი და T სტატისტიკის სხვა დაკვირვებული მნიშვნელობა. ეს აიხსნება იმით, რომ უკანასკნელ შემთხვევაში სარგებლობენ ე. წ. სატერტვაიტის მეთოდით.

შენიშვნა 2. სტატისტიკური ფუნქციით $TTEST(array1,array2,1,3)$ – გამოითვლება სტიუდენტის განაწილების p -მნიშვნელობა ცალმხრივი ალტერნატივისათვის, ხოლო $TTEST(array1,array2,2,3)$ -ით კი p -მნიშვნელობა ორმხრივი ალტერნატივისათვის.

t -Test-ი დაწყვილებული მონაცემებისათვის. ორპექტარიანი 10 ნაკვეთი დაიყო ორ-ორ ტოლ ნაწილად, ერთ ნაწილზე გამოიყენეს I ტიპის სასუქი, ხოლო მეორეზე კი – II ტიპის სასუქი. მოსავლიანობამ, შესაბამისად, შეადგინა:

	მოსავლიანობა									
ნაკვეთის #	1	2	3	4	5	6	7	8	9	10
I სასუქი	1.2	1.4	1.6	1.1	1.4	1.1	0.7	1.5	1.3	1
II სასუქი	1.4	1.2	1.5	1.1	1.6	1.2	1	1.4	1.3	1.5

0.05 მნიშვნელოვნების დონით შეამოწმეთ, ზრდის თუ არა II ტიპის სასუქის გამოყენება მოსავლიანობას I ტიპის სასუქის გამოყენებასთან შედარებით.

ამოხსნა. ვამოწმებთ ძირითად ჰიპოთეზას $H_0: \mu_D = \mu_1 - \mu_2 = 0$ ცალმხრივი $H_1: \mu_D = \mu_1 - \mu_2 < 0$ და ორმხრივი $H_1: \mu_D = \mu_1 - \mu_2 \neq 0$ ალტერნატივების წინააღმდეგ. ქვეპროგრამა t -Test: Paired Two Sample for Means გვაძლევს:

	Variable 1	Variable 2
Mean / საშუალო	1.23	1.32
Variance/ დისპერსია	0.0712	0.0373
Observations/ დაკვირვებები	10	10
Pearson Correlation / პირსონის კორელაცია	0.612	
Hypothesized Mean Difference / ძირითადი ჰიპოთეზით განსაზღვრული საშუალოთა სხვაობა	0	
df / თავისუფლების ხარისხი	9	
t Stat /T-სტატისტიკის დაკვირვებული მნიშვნელობა	-1.3351	
P{T ≤ t} one-tail / p-მნიშვნელობა ცალმხრივი კრიტიკული არისათვის	0.1073	
/ Critical one-tail / კრიტიკული ნერტილი ცალმხრივი კრიტიკული არისათვის	1.8331	
P{T ≤ t} two-tail / p-მნიშვნელობა ორმხრივი კრიტიკული არისათვის	0.2146	
/ Critical two-tail / კრიტიკული ნერტილი ორმხრივი კრიტიკული არისათვის	2.2622	

დასკვნა:

- ❖ ცალმხრივი $H_1: a_b = a_1 - a_2 < 0$ ალტერნატივისათვის p -მნიშვნელობა = 0.1073 $> \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს;
- ❖ ორმხრივი $H_1: a_b = a_1 - a_2 \neq 0$ ალტერნატივისათვის p -მნიშვნელობა = 0.2146 $> \alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითადი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.

შენიშვნა. სტატისტიკური ფუნქციით TTEST(array1,array2,1,1) – გამოითვლება სტიუდენტის განაწილების p -მნიშვნელობა ცალმხრივი ალტერნატივისათვის, ხოლო TTEST(array1,array2, 2,1)-ით კი p -მნიშვნელობა ორმხრივი ალტერნატივისათვის.

F-Test-ის გამოყენების მაგალითი. 0.05 მნიშვნელოვნების დონით შეამოწმეთ ჰიპოთეზა ორი დამოუკიდებელი ნორმალური პოპულაციის დისპერსიების ტოლობის შესახებ, თუ მათგან მიღებული შერჩევებია:

I შერჩევა	12.1	4.1	4.6	15.2	-0.9	3.2	2.5	10.1	7.2	13.5
II შერჩევა	7	11.6	11.6	9.8	7.3	9	6.3	13.7		

ამოხსნა. მოწმდება ძირითადი ჰიპოთეზა $H_0: \sigma_1^2 = \sigma_2^2$ ცალმხრივი $H_1: \sigma_1^2 > \sigma_2^2$ ალტერნატივის წინააღმდეგ. ქვეპროგრამა-ტესტ: **Two Sample for Variances**

	I	II
Mean / საშუალო	7.16	9.54
Variance / დისპერსია	28.463	6.874
Observations / დაკვირვებები	10	8
df / თავისუფლების ხარისხი	9	7
F Stat : F--სტატისტიკის დაკვირვებული მნიშვნელობა	4.141	
P(F ≤ f) one-tail / p-მნიშვნელობა ცალმხრივი კრიტიკული არისათვის	0.037	
F Critical one-tail / კრიტიკული ნერტილი ცალმხრივი კრიტიკული არისათვის	3.677	

დასკვნა: ცალმხრივი $H_1: \sigma_1^2 > \sigma_2^2$ ალტერნატივისათვის p -მნიშვნელობა=0.037 < $\alpha = 0.05$. შესაბამისად, 0.05 მნიშვნელოვნების დონით, ძირითად ჰიპოთეზას უარყოფთ.

შენიშვნა. სტატისტიკური ფუნქციით **FTEST(array1,array2)** – გამოითვლება ფიშერის განაწილების p -მნიშვნელობა ცალმხრივი ალტერნატივისათვის.

დისპერსიული ანალიზი

ქვეპროგრამა **Anova: Single Factor** სარგებლობისას გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Input range	შემავალი მონაცემები
Grouped by	დაჯგუფებული
Labels	მონიშვნა

მაგალითი. მენეჯერის მიზანია, 0.05 მნიშვნელოვნების დონით შეამოწმოს შემთხვევით შერჩეული 3 ჩარხის ერთგვაროვნება მათზე გამოჩარხული ერთი და იმავე დეტალების ზომების მონაცემების ცვალებადობის მიხედვით:

I ჩარხი	6.783	6.781	6.776	6.788	6.785	6.789
II ჩარხი	6.775	6.77	6.772	6.771	6.779	6.773
III ჩარხი	6.778	6.776	6.769	6.772	6.78	6.777

Anova: Single Factor ქვეპროგრამის გამოყენება გეძლეხს:
SUMMARX

Groups	Caunt	Sum	Average	Variance
Column 1	6	502	83.67	23.07
Column 2	6	440	73.33	10.67
Column 3	6	452	75.33	16.67

ANOVA

Source of Variation SS df MS F

P-value F crit

ვარიაციის წყარო

Between Groups 360.44 2 180.22 10.73 0.00 3.68

Within Groups 252 15 16.8

Total 612.44 17

დასკვნა: ვინაიდან $f = 10.73 > F_{2,15,0.05} = 3.68$ ამიტომ უარეყოფთ ჰიპოთეზას ჩარხების ერთგვაროვნების შესახებ - 0.05 მნიშვნელოვნების დონით ჩარხები არაერთგვაროვანია.

კორელაცია

ფუნქცია **COVAR(array1,array2)** – გამოიყენება კოვარიაციის გამოსათვლელად.

ფუნქცია **CORREL(array1,array2)** – გამოიყენება კორელაციის კოეფიციენტის გამოსათვლელად (**CORREL**-ის იდენტური აღმუქცია)

CovarianceqveprogramiT გამოითვლება ერთნაირი მოცულობის მონაცემთა რამდენიმე მასივის კოვარიაციული მატრიცა.

CorrelationqveprogramiT გამოითვლება ერთნაირი მოცულობის მონაცემთა რამდენიმე მასივის კორელაციური მატრიცა.

რეგრესია

ნრფივი რეგრესიული ანალიზის ჩასატარებლად გამოიყენება ფუნქცია **Tools/Data Analysis/Regression**. თუაღნიშნული ფუნქცია მიუწეოდმელია, მაშინ უნდა შესრულდეს შემდეგი მოქმედებები: **Tools/ add-ins-SI** უნდა გააქტიურდეს **Analysis ToolPak, Analysis ToolPak-VBA**.

მარტივ ნრფივ რეგრესიულ მოდელთან დაკავშირებით გამოიყენება შემდეგი აღნიშვნები:

აღნიშვნა	შინაარსი
Known-y's	ცნობილი დამოკიდებული ცვლადის მნიშვნელობები
Known-x's	ცნობილი დამოუკიდებელი ცვლადის მნიშვნელობები
x	დამოუკიდებელი ცვლადის მნიშვნელობა
Const	თუ Const= 0 , მაშინ რეგრესიის წრფე გადის სათავეზე

ფუნქცია **DEVSQ(number1,number2,p)** – ითვლის დაკვირვებული მონაცემების საშუალოდან გადახრების კვადრატების ჯამს.

ფუნქცია **INTERCEPT(known-y's,known-x's)** – ითვლის B_0 კოეფიციენტის b_0 შეფასებას.

ფუნქცია **LINEST(known-y's,known-x's,const,stats)** – ითვლის B_1 კოეფიციენტის b_1 შეფასებას (თუ $\text{Const} = 0$, მაშინ გამოითვლება B_1 კოეფიციენტის შეფასება $Y = B_1 \cdot x + \varepsilon$ მოდელისათვის, ხოლო თუ $\text{Const} = 1$, მაშინ გამოითვლება B_1 -ის შეფასება $Y = B_0 + B_1 \cdot x + \varepsilon$ მოდელისათვის). **LINEST**-ის იდენტურია **ფუნქცია**.

ფუნქცია **FORECAST(x,known-y's,known-x's)** – ითვლის საპროგნოზო მნიშვნელობას.

ფუნქცია **RSQ(known-y's,known-x's)** – ითვლის პირსონის კორელაციის კოეფიციენტის კვადრატს (ანუ r^2). **დეტერმინაციის კოეფიციენტი**).

ფუნქცია **TREND(known-y's,known-x's,new-x's,const)** – ითვლის საპროგნოზო მნიშვნელობას (თუ $\text{Const} = 0$, მაშინ გამოითვლება B_1 კოეფიციენტის შეფასება $Y = b_0 + b_1 \cdot x$ მოდელისათვის, ხოლო – თუ $\text{Const} = 1$, მაშინ გამოითვლება B_1 -ის შეფასება $Y = b_0 + b_1 \cdot x$ მოდელისათვის).

ფუნქცია **GROWTH(known-y's,known-x's,new-x's,const)** – ითვლის საპროგნოზო მნიშვნელობას (თუ $\text{Const} = 0$, მაშინ გამოითვლება Y -ის შეფასება $Y = (\beta_1)^x$ მოდელისათვის, ხოლო, თუ $\text{Const} = 1$, მაშინ Y გამოითვლება $Y = \beta_0 \cdot (\beta_1)^x$ მოდელისათვის).

მაგალითი XVIII. 1-ის მონაცემების ანალიზი. Regression ფუნქციის გამოყენება გვაძლევს:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8047
R Square	0.6476
Adjusted R Square	0.6204
Standard Error	12.9965
Observations	15

შესაბამისად, ამონაბეჭდი გვაძლევს კორელაციისა (**Multiple R**) და დეტერმინაციის (**R Square**) კოეფიციენტების, შეთანხმებული დეტერმინაციის კოეფიციენტის (**Adjusted R Square**) და სტანდარტული შეცდომის (**Standard Error**) რიცხვით მნიშვნელობებს. დეტერმინაციის კოეფიციენტი 0.6476 გვიჩვენებს, რომ მარტივი რეგრესიით აიხსნება სახლის გასაყიდი ფასის ვარიაციის მხოლოდ 64.76%, დანარჩენი 35.24% კი აუხსნელი რჩება. საჭიროა შეირჩეს სხვა მოდელი.

(სტატისტიკური ცხრილები)

თანაბრად განაწილებული შემთხვევითი რიცხვები

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	44482	66558	37649	08882	90870	12462	41810	01806	02977
36792	26236	33266	66583	60881	97395	20461	36742	02852	50564
73944	04773	12032	51414	82384	38370	00249	80709	72605	67497
49563	12872	14063	93104	78483	72717	68714	18048	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	72907	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
67642	05204	30697	44806	96989	68403	85621	45556	35434	09532
64041	99011	14610	40273	09482	62864	01573	82274	81446	32477
17048	94523	97444	59904	16936	39384	97551	09620	63932	03091
93039	89416	52795	10631	09728	68202	20963	02477	55494	39563
82244	34392	96607	17220	51984	10753	76272	50985	97593	34320
96990	55244	70693	25255	40029	23289	48819	07159	60172	81697
09119	74803	97303	88701	51380	73143	98251	78635	27556	20712
57666	41204	47589	78364	38266	94393	70713	53388	79865	92069
46492	61594	26729	58272	81754	14648	77210	12923	53712	87771
08433	19172	08320	20839	13715	10597	17234	39355	74816	03363

გამოყენებითი სტატისტიკის საფუძვლები

10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
07428	58863	96023	88936	51343	70958	96768	74317	27176	29600
35379	27922	28906	55013	26937	48174	04197	36074	65315	12537
10982	22807	10920	26299	23593	64629	57801	10437	43965	15344
90127	33341	77806	12446	15444	49244	47277	11346	15884	28131
63002	12990	23510	68774	48983	20481	59815	67248	17076	78910
40779	86382	48454	65269	91239	45989	45389	54847	77919	41105
43216	12608	18167	84631	94058	82458	15139	76856	86019	47928
96167	64375	74108	93643	09204	98855	59051	56492	11933	64958
70975	62693	35684	72607	23026	37004	32989	24843	01128	74658
85812	61875	23570	75754	29090	40264	80399	47254	40135	69916

კუმულატიური ბინომური განაწილების ცხრილი

$$P\{Bin(n, p) \leq x\} = \sum_{i=0}^x C_n^i p^i (1-p)^{n-i}$$

		P											
n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	
2	0	0.9801	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500	
	1	0.9999	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500	
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
3	0	0.97030	0.85738	0.729	0.61413	0.512	0.42187	0.343	0.27463	0.216	0.16638	0.125	
	1	0.99970	0.99275	0.972	0.93925	0.896	0.84375	0.784	0.71825	0.648	0.57475	0.500	
	2	1.00000	0.99988	0.999	0.99663	0.992	0.98437	0.973	0.95713	0.936	0.90887	0.875	
	3	1.00000	1.00000	1.000	1.00000	1.000	1.00000	1.000	1.00000	1.000	1.00000	1.000	
4	0	0.96060	0.81451	0.6561	0.52201	0.4096	0.31641	0.2401	0.17851	0.1296	0.09151	0.0625	
	1	0.99941	0.98598	0.9477	0.89048	0.8192	0.73828	0.6517	0.56298	0.4752	0.39098	0.3125	
	2	1.00000	0.99952	0.9963	0.98802	0.9728	0.94922	0.9163	0.87352	0.8208	0.75852	0.6875	
	3	1.00000	0.99999	0.9999	0.99949	0.9984	0.99609	0.9919	0.98499	0.9744	0.95899	0.9375	
	4	1.00000	1.00000	1.0000	1.00000	1.0000	1.00000	1.0000	1.00000	1.0000	1.00000	1.0000	
5	0	0.95099	0.77378	0.59049	0.44371	0.32768	0.23730	0.16807	0.11603	0.07776	0.05033	0.03125	
	1	0.99902	0.97741	0.91854	0.83521	0.73728	0.63281	0.52822	0.42842	0.33696	0.25622	0.18750	
	2	0.99999	0.99884	0.99144	0.97339	0.94208	0.89648	0.83692	0.76483	0.68256	0.59313	0.50000	
	3	1.00000	0.99997	0.99954	0.99777	0.99328	0.98437	0.96922	0.94598	0.91296	0.86878	0.81250	
	4	1.00000	1.00000	0.99999	0.99992	0.99968	0.99902	0.99757	0.99475	0.98976	0.98155	0.96875	
	5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
6	0	0.94148	0.73509	0.53144	0.37715	0.26214	0.17798	0.11765	0.07542	0.04666	0.02768	0.01563	
	1	0.99854	0.96723	0.88573	0.77648	0.65536	0.53394	0.42017	0.31908	0.23328	0.16357	0.10938	
	2	0.99998	0.99777	0.98415	0.95266	0.90112	0.83057	0.74431	0.64709	0.54432	0.44152	0.34375	
	3	1.00000	0.99991	0.99873	0.99411	0.98304	0.96240	0.92953	0.88258	0.82080	0.74474	0.65625	
	4	1.00000	1.00000	0.99994	0.99960	0.99840	0.99536	0.98906	0.97768	0.95904	0.93080	0.89062	
	5	1.00000	1.00000	1.00000	0.99999	0.99994	0.99976	0.99927	0.99816	0.99590	0.99170	0.98437	
	6	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
7	0	0.93207	0.69834	0.47830	0.32058	0.20972	0.13348	0.08235	0.04902	0.02799	0.01522	0.00781	
	1	0.99797	0.95562	0.85031	0.71658	0.57672	0.44495	0.32942	0.23380	0.15863	0.10242	0.06250	
	2	0.99997	0.99624	0.97431	0.92623	0.85197	0.75641	0.64707	0.53228	0.41990	0.31644	0.22656	
	3	1.00000	0.99981	0.99727	0.98790	0.96666	0.92944	0.87396	0.80015	0.71021	0.60829	0.50000	
	4	1.00000	0.99999	0.99982	0.99878	0.99533	0.98712	0.97120	0.94439	0.90374	0.84707	0.77344	
	5	1.00000	1.00000	0.99999	0.99993	0.99963	0.99866	0.99621	0.99099	0.98116	0.96429	0.93750	
	6	1.00000	1.00000	1.00000	1.00000	0.99999	0.99994	0.99978	0.99936	0.99836	0.99626	0.99219	
	7	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	
8	0	0.92274	0.66342	0.43047	0.27249	0.16777	0.10011	0.05765	0.03186	0.01680	0.00837	0.00391	
	1	0.99731	0.94276	0.81310	0.65718	0.50332	0.36708	0.25530	0.16913	0.10638	0.06318	0.03516	
	2	0.99995	0.99421	0.96191	0.89479	0.79692	0.67854	0.55177	0.42781	0.31539	0.22013	0.14453	
	3	1.00000	0.99963	0.99498	0.97865	0.94372	0.88618	0.80590	0.70640	0.59409	0.47696	0.36328	
	4	1.00000	0.99998	0.99957	0.99715	0.98959	0.97270	0.94203	0.89391	0.82633	0.73962	0.63672	
	5	1.00000	1.00000	0.99998	0.99976	0.99877	0.99577	0.98871	0.97468	0.95019	0.91154	0.85547	
	6	1.00000	1.00000	1.00000	0.99999	0.99992	0.99962	0.99871	0.99643	0.99148	0.98188	0.96484	
	7	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99993	0.99977	0.99934	0.99832	0.99609	

გამოყენებითი სტატისტიკის საფუძვლები

8	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
9	0	0.91352	0.63025	0.38742	0.23162	0.13422	0.07508	0.04035	0.02071	0.01008	0.00461	0.00195
	1	0.99656	0.92879	0.77484	0.59948	0.43621	0.30034	0.19600	0.12109	0.07054	0.03852	0.01953
	2	0.99992	0.99164	0.94703	0.85915	0.73820	0.60068	0.46283	0.33727	0.23179	0.14950	0.08984
	3	1.00000	0.99936	0.99167	0.96607	0.91436	0.83427	0.72966	0.60889	0.48261	0.36138	0.25391
	4	1.00000	0.99997	0.99911	0.99437	0.98042	0.95107	0.90119	0.82828	0.73343	0.62142	0.50000
	5	1.00000	1.00000	0.99994	0.99937	0.99693	0.99001	0.97471	0.94641	0.90065	0.83418	0.74609
	6	1.00000	1.00000	1.00000	0.99995	0.99969	0.99866	0.99571	0.98882	0.97497	0.95023	0.91016
	7	1.00000	1.00000	1.00000	1.00000	0.99998	0.99989	0.99957	0.99860	0.99620	0.99092	0.98047
	8	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99992	0.99974	0.99924	0.99805
	9	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
10	0	0.90438	0.59874	0.34868	0.19687	0.10737	0.05631	0.02825	0.01346	0.00605	0.00253	0.00098
	1	0.99573	0.91386	0.73610	0.54430	0.37581	0.24403	0.14931	0.08595	0.04636	0.02326	0.01074
	2	0.99989	0.98850	0.92981	0.82020	0.67780	0.52559	0.38278	0.26161	0.16729	0.09956	0.05469
	3	1.00000	0.99897	0.98720	0.95003	0.87913	0.77588	0.64961	0.51383	0.38228	0.26604	0.17188
	4	1.00000	0.99994	0.99837	0.99013	0.96721	0.92187	0.84973	0.75150	0.63310	0.50440	0.37695
	5	1.00000	1.00000	0.99985	0.99862	0.99363	0.98027	0.95265	0.90507	0.83376	0.73844	0.62305
	6	1.00000	1.00000	0.99999	0.99987	0.99914	0.99649	0.98941	0.97398	0.94524	0.89801	0.82812
	7	1.00000	1.00000	1.00000	0.99999	0.99992	0.99958	0.99841	0.99518	0.98771	0.97261	0.94531
	8	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99986	0.99946	0.99832	0.99550	0.98926
	9	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99997	0.99990	0.99966	0.99902
	10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
11	0	0.89534	0.56880	0.31381	0.16734	0.08590	0.04224	0.01977	0.00875	0.00363	0.00139	0.00049
	1	0.99482	0.89811	0.69736	0.49219	0.32212	0.19710	0.11299	0.06058	0.03023	0.01393	0.00586
	2	0.99984	0.98476	0.91044	0.77881	0.61740	0.45520	0.31274	0.20013	0.11892	0.06522	0.03271
	3	1.00000	0.99845	0.98147	0.93056	0.83886	0.71330	0.56956	0.42555	0.29628	0.19112	0.11328
	4	1.00000	0.99989	0.99725	0.98411	0.94959	0.88537	0.78970	0.66831	0.53277	0.39714	0.27441
	5	1.00000	0.99999	0.99970	0.99734	0.98835	0.96567	0.92178	0.85132	0.75350	0.63312	0.50000
	6	1.00000	1.00000	0.99998	0.99968	0.99803	0.99244	0.97838	0.94986	0.90065	0.82620	0.72559
	7	1.00000	1.00000	1.00000	0.99997	0.99976	0.99881	0.99571	0.98776	0.97072	0.93904	0.88672
	8	1.00000	1.00000	1.00000	1.00000	0.99998	0.99987	0.99942	0.99796	0.99408	0.98520	0.96729
	9	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99995	0.99979	0.99927	0.99779	0.99414
	10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99996	0.99985	0.99951
	11	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
12	0	0.88638	0.54036	0.28243	0.14224	0.06872	0.03168	0.01384	0.00569	0.00218	0.00077	0.00024
	1	0.99383	0.88164	0.65900	0.44346	0.27488	0.15838	0.08503	0.04244	0.01959	0.00829	0.00317
	2	0.99979	0.98043	0.88913	0.73582	0.55835	0.39068	0.25282	0.15129	0.08344	0.04214	0.01929
	3	1.00000	0.99776	0.97436	0.90779	0.79457	0.64878	0.49252	0.34665	0.22534	0.13447	0.07300
	4	1.00000	0.99982	0.99567	0.97608	0.92744	0.84236	0.72366	0.58335	0.43818	0.30443	0.19385
	5	1.00000	0.99999	0.99946	0.99536	0.98059	0.94560	0.88215	0.78726	0.66521	0.52693	0.38721
	6	1.00000	1.00000	0.99995	0.99933	0.99610	0.98575	0.96140	0.91537	0.84179	0.73931	0.61279
	7	1.00000	1.00000	1.00000	0.99993	0.99942	0.99722	0.99051	0.97449	0.94269	0.88826	0.80615
	8	1.00000	1.00000	1.00000	0.99999	0.99994	0.99961	0.99831	0.99439	0.98473	0.96443	0.92700
	9	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99979	0.99915	0.99719	0.99212	0.98071
	10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99992	0.99968	0.99892	0.99683
	11	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99993	0.99976	0.99976
	12	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
13	0	0.87752	0.51334	0.25419	0.12091	0.05498	0.02376	0.00969	0.00370	0.00131	0.00042	0.00012
	1	0.99275	0.86458	0.62134	0.39828	0.23365	0.12671	0.06367	0.02958	0.01263	0.00490	0.00171

2	0.99973	0.97549	0.86612	0.69196	0.50165	0.33260	0.20248	0.11319	0.05790	0.02691	0.01123
3	0.99999	0.99690	0.96584	0.88200	0.74732	0.58425	0.42061	0.27827	0.16858	0.09292	0.04614
4	1.00000	0.99971	0.99354	0.96584	0.90087	0.79396	0.65431	0.50050	0.35304	0.22795	0.13342
5	1.00000	0.99998	0.99908	0.99247	0.96996	0.91979	0.83460	0.71589	0.57440	0.42681	0.29053
6	1.00000	1.00000	0.99990	0.99873	0.99300	0.97571	0.93762	0.87053	0.77116	0.64374	0.50000
7	1.00000	1.00000	0.99999	0.99984	0.99875	0.99435	0.98178	0.95380	0.90233	0.82123	0.70947
8	1.00000	1.00000	1.00000	0.99998	0.99983	0.99901	0.99597	0.98743	0.96792	0.93015	0.86658
9	1.00000	1.00000	1.00000	1.00000	0.99998	0.99987	0.99935	0.99749	0.99221	0.97966	0.95386
10	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99993	0.99965	0.99868	0.99586	0.98877
11	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99997	0.99986	0.99948	0.99829
12	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99997	0.99997	0.99988
13	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
14	0	0.86875	0.48767	0.22877	0.10277	0.04398	0.01782	0.00678	0.00240	0.00078	0.00023
1	0.99160	0.84701	0.58463	0.35667	0.19791	0.10097	0.04748	0.02052	0.00810	0.00289	0.00092
2	0.99966	0.96995	0.84164	0.64791	0.44805	0.28113	0.16084	0.08393	0.03979	0.01701	0.00647
3	0.99999	0.99583	0.95587	0.85349	0.69819	0.52134	0.35517	0.22050	0.12431	0.06322	0.02869
4	1.00000	0.99957	0.99077	0.95326	0.87016	0.74153	0.58420	0.42272	0.27926	0.16719	0.08978
5	1.00000	0.99997	0.99853	0.98847	0.95615	0.88833	0.78052	0.64051	0.48585	0.33732	0.21198
6	1.00000	1.00000	0.99982	0.99779	0.98839	0.96173	0.90672	0.81641	0.69245	0.54612	0.39526
7	1.00000	1.00000	0.99998	0.99967	0.99760	0.98969	0.96853	0.92466	0.84986	0.74136	0.60474
8	1.00000	1.00000	1.00000	0.99996	0.99962	0.99785	0.99171	0.97566	0.94168	0.88114	0.78802
9	1.00000	1.00000	1.00000	1.00000	0.99995	0.99966	0.99833	0.99396	0.98249	0.95738	0.91022
10	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99975	0.99889	0.99609	0.98857	0.97131
11	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99986	0.99939	0.99785	0.99353
12	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99994	0.99975	0.99908
13	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99994
14	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
15	0	0.86006	0.46329	0.20589	0.08735	0.03518	0.01336	0.00475	0.00156	0.00047	0.00013
1	0.99037	0.82905	0.54904	0.31859	0.16713	0.08018	0.03527	0.01418	0.00517	0.00169	0.00049
2	0.99958	0.96380	0.81594	0.60423	0.39802	0.23609	0.12683	0.06173	0.02711	0.01065	0.00369
3	0.99999	0.99453	0.94444	0.82266	0.64816	0.46129	0.29687	0.17270	0.09050	0.04242	0.01758
4	1.00000	0.99939	0.98728	0.93829	0.83577	0.68649	0.51549	0.35194	0.21728	0.12040	0.05923
5	1.00000	0.99995	0.99775	0.98319	0.93895	0.85163	0.72162	0.56428	0.40322	0.26076	0.15088
6	1.00000	1.00000	0.99969	0.99639	0.98194	0.94338	0.86886	0.75484	0.60981	0.45216	0.30362
7	1.00000	1.00000	0.99997	0.99939	0.99576	0.98270	0.94999	0.88677	0.78690	0.65350	0.50000
8	1.00000	1.00000	1.00000	0.99992	0.99922	0.99581	0.98476	0.95781	0.90495	0.81824	0.69638
9	1.00000	1.00000	1.00000	0.99999	0.99989	0.99921	0.99635	0.98756	0.96617	0.92307	0.84912
10	1.00000	1.00000	1.00000	1.00000	0.99999	0.99988	0.99933	0.99717	0.99065	0.97453	0.94077
11	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99991	0.99952	0.99807	0.99367	0.98242
12	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99994	0.99972	0.99889	0.99631
13	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99988	0.99951
14	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99997
15	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
20	0	0.81791	0.35849	0.12158	0.03876	0.01153	0.00317	0.00080	0.00018	0.00004	0.00001
1	0.98314	0.73584	0.39175	0.17556	0.06918	0.02431	0.00764	0.00213	0.00052	0.00011	0.00002
2	0.99900	0.92452	0.67693	0.40490	0.20608	0.09126	0.03548	0.01212	0.00361	0.00093	0.00020
3	0.99996	0.98410	0.86705	0.64773	0.41145	0.22516	0.10709	0.04438	0.01596	0.00493	0.00129
4	1.00000	0.99743	0.95683	0.82985	0.62965	0.41484	0.23751	0.11820	0.05095	0.01886	0.00591
5	1.00000	0.99967	0.98875	0.93269	0.80421	0.61717	0.41637	0.24540	0.12560	0.05533	0.02069
6	1.00000	0.99997	0.99761	0.97806	0.91331	0.78578	0.60801	0.41663	0.25001	0.12993	0.05766
7	1.00000	1.00000	0.99958	0.99408	0.96786	0.89819	0.77227	0.60103	0.41589	0.25201	0.13159

გამოყენებითი სტატისტიკის საფუძვლები

8	1.00000	1.00000	0.99994	0.99867	0.99002	0.95907	0.88667	0.76238	0.59560	0.41431	0.25172
9	1.00000	1.00000	0.99999	0.99975	0.99741	0.98614	0.95204	0.87822	0.75534	0.59136	0.41190
10	1.00000	1.00000	1.00000	0.99996	0.99944	0.99606	0.98286	0.94683	0.87248	0.75071	0.58810
11	1.00000	1.00000	1.00000	0.99999	0.99990	0.99906	0.99486	0.98042	0.94347	0.86923	0.74828
12	1.00000	1.00000	1.00000	1.00000	0.99998	0.99982	0.99872	0.99398	0.97897	0.94197	0.86841
13	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99974	0.99848	0.99353	0.97859	0.94234
14	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99969	0.99839	0.99357	0.97931
15	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99995	0.99968	0.99847	0.99409
16	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99995	0.99972	0.99871
17	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99996	0.99980
18	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998
19	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
20	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
25	0	0.77782	0.27739	0.07179	0.01720	0.00378	0.00075	0.00013	0.00002	0.00000	0.00000
1	0.97424	0.64238	0.27121	0.09307	0.02739	0.00702	0.00157	0.00030	0.00005	0.00001	0.00000
2	0.99805	0.87289	0.53709	0.25374	0.09823	0.03211	0.00896	0.00213	0.00043	0.00007	0.00001
3	0.99989	0.96591	0.76359	0.47112	0.23399	0.09621	0.03324	0.00968	0.00237	0.00048	0.00008
4	1.00000	0.99284	0.90201	0.68211	0.42067	0.21374	0.09047	0.03205	0.00947	0.00231	0.00046
5	1.00000	0.99879	0.96660	0.83848	0.61669	0.37828	0.19349	0.08262	0.02936	0.00860	0.00204
6	1.00000	0.99983	0.99052	0.93047	0.78004	0.56110	0.34065	0.17340	0.07357	0.02575	0.00732
7	1.00000	0.99998	0.99774	0.97453	0.89088	0.72651	0.51185	0.30608	0.15355	0.06385	0.02164
8	1.00000	1.00000	0.99954	0.99203	0.95323	0.85056	0.67693	0.46682	0.27353	0.13398	0.05388
9	1.00000	1.00000	0.99992	0.99786	0.98267	0.92867	0.81056	0.63031	0.42462	0.24237	0.11476
10	1.00000	1.00000	0.99999	0.99951	0.99445	0.97033	0.90220	0.77116	0.58577	0.38426	0.21218
11	1.00000	1.00000	1.00000	0.99990	0.99846	0.98927	0.95575	0.87458	0.73228	0.54257	0.34502
12	1.00000	1.00000	1.00000	0.99998	0.99963	0.99663	0.98253	0.93956	0.84623	0.69368	0.50000
13	1.00000	1.00000	1.00000	1.00000	0.99992	0.99908	0.99401	0.97454	0.92220	0.81731	0.65498
14	1.00000	1.00000	1.00000	1.00000	0.99999	0.99979	0.99822	0.99069	0.96561	0.90402	0.78782
15	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99955	0.99706	0.98683	0.95604	0.88524
16	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99990	0.99921	0.99567	0.98264	0.94612
17	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99982	0.99879	0.99417	0.97836
18	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99972	0.99836	0.99268
19	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99995	0.99962	0.99796
20	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99999	0.99954
21	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99992
22	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999
23	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
24	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
25	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
50	0	0.60501	0.07694	0.00515	0.00030	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.91056	0.27943	0.03379	0.00291	0.00019	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.98618	0.54053	0.11173	0.01419	0.00129	0.00009	0.00000	0.00000	0.00000	0.00000	0.00000
3	0.99840	0.76041	0.25029	0.04605	0.00566	0.00050	0.00003	0.00000	0.00000	0.00000	0.00000
4	0.99985	0.89638	0.43120	0.11211	0.01850	0.00211	0.00017	0.00001	0.00000	0.00000	0.00000
5	0.99999	0.96222	0.61612	0.21935	0.04803	0.00705	0.00072	0.00005	0.00000	0.00000	0.00000
6	1.00000	0.98821	0.77023	0.36130	0.10340	0.01939	0.00249	0.00022	0.00001	0.00000	0.00000
7	1.00000	0.99681	0.87785	0.51875	0.19041	0.04526	0.00726	0.00080	0.00006	0.00000	0.00000
8	1.00000	0.99924	0.94213	0.66810	0.30733	0.09160	0.01825	0.00248	0.00023	0.00001	0.00000
9	1.00000	0.99984	0.97546	0.79109	0.44374	0.16368	0.04023	0.00670	0.00076	0.00006	0.00000
10	1.00000	0.99997	0.99065	0.88008	0.58356	0.26220	0.07885	0.01601	0.00220	0.00020	0.00001
11	1.00000	1.00000	0.99678	0.93719	0.71067	0.38162	0.13904	0.03423	0.00569	0.00063	0.00005
12	1.00000	1.00000	0.99900	0.96994	0.81394	0.51099	0.22287	0.06613	0.01325	0.00177	0.00015

13	1.00000	1.00000	0.99971	0.98683	0.88941	0.63704	0.32788	0.11633	0.02799	0.00449	0.00047
14	1.00000	1.00000	0.99993	0.99471	0.93928	0.74808	0.44683	0.18778	0.05395	0.01038	0.00130
15	1.00000	1.00000	0.99998	0.99805	0.96920	0.83692	0.56918	0.28010	0.09550	0.02195	0.00330
16	1.00000	1.00000	1.00000	0.99934	0.98556	0.90169	0.68388	0.38886	0.15609	0.04265	0.00767
17	1.00000	1.00000	1.00000	0.99979	0.99374	0.94488	0.78219	0.50597	0.23688	0.07653	0.01642
18	1.00000	1.00000	1.00000	0.99994	0.99749	0.97127	0.85944	0.62159	0.33561	0.12734	0.03245
19	1.00000	1.00000	1.00000	0.99998	0.99907	0.98608	0.91520	0.72644	0.44648	0.19737	0.05946
20	1.00000	1.00000	1.00000	1.00000	0.99968	0.99374	0.95224	0.81394	0.56103	0.28617	0.10132
21	1.00000	1.00000	1.00000	1.00000	0.99990	0.99738	0.97491	0.88126	0.67014	0.38996	0.16112
22	1.00000	1.00000	1.00000	1.00000	0.99997	0.99898	0.98772	0.92904	0.76602	0.50191	0.23994
23	1.00000	1.00000	1.00000	1.00000	0.99999	0.99963	0.99441	0.96036	0.84383	0.61341	0.33591
24	1.00000	1.00000	1.00000	1.00000	1.00000	0.99988	0.99763	0.97933	0.90219	0.71604	0.44386
25	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99907	0.98996	0.94266	0.80337	0.55614
26	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99966	0.99546	0.96859	0.87207	0.66409
27	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99988	0.99809	0.98397	0.92204	0.76006
28	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99925	0.99238	0.95562	0.83888
29	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99973	0.99664	0.97646	0.89868
30	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99991	0.99863	0.98840	0.94054
31	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99948	0.99470	0.96755
32	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99982	0.99776	0.98358
33	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99994	0.99913	0.99233
34	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99969	0.99670
35	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99990	0.99987	0.99870
36	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99953
37	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99985
38	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99995
39	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999
40	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
41	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
42	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
43	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
44	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
45	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
46	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
47	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
48	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
49	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
50	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
100	0	0.36603	0.00592	0.00003	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.73576	0.03708	0.00032	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.92063	0.11826	0.00194	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	0.98163	0.25784	0.00784	0.00009	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4	0.99657	0.43598	0.02371	0.00043	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
5	0.99947	0.61600	0.05758	0.00155	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
6	0.99993	0.76601	0.11716	0.00470	0.00008	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
7	0.99999	0.87204	0.20605	0.01217	0.00028	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
8	1.00000	0.93691	0.32087	0.02748	0.00086	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
9	1.00000	0.97181	0.45129	0.05509	0.00233	0.00004	0.00000	0.00000	0.00000	0.00000	0.00000
10	1.00000	0.98853	0.58316	0.09945	0.00570	0.00014	0.00000	0.00000	0.00000	0.00000	0.00000
11	1.00000	0.99573	0.70303	0.16349	0.01257	0.00039	0.00001	0.00000	0.00000	0.00000	0.00000
12	1.00000	0.99854	0.80182	0.24730	0.02533	0.00103	0.00002	0.00000	0.00000	0.00000	0.00000
13	1.00000	0.99954	0.87612	0.34742	0.04691	0.00246	0.00006	0.00000	0.00000	0.00000	0.00000
14	1.00000	0.99986	0.92743	0.45722	0.08044	0.00542	0.00016	0.00000	0.00000	0.00000	0.00000

გამოყენებითი სტატისტიკის საფუძვლები

15	1.00000	0.99996	0.96011	0.56832	0.12851	0.01108	0.00040	0.00001	0.00000	0.00000	0.00000
16	1.00000	0.99999	0.97940	0.67246	0.19234	0.02111	0.00097	0.00002	0.00000	0.00000	0.00000
17	1.00000	1.00000	0.98999	0.76328	0.27119	0.03763	0.00216	0.00005	0.00000	0.00000	0.00000
18	1.00000	1.00000	0.99542	0.83717	0.36209	0.06301	0.00452	0.00014	0.00000	0.00000	0.00000
19	1.00000	1.00000	0.99802	0.89346	0.46016	0.09953	0.00889	0.00034	0.00001	0.00000	0.00000
20	1.00000	1.00000	0.99919	0.93368	0.55946	0.14883	0.01646	0.00078	0.00002	0.00000	0.00000
21	1.00000	1.00000	0.99969	0.96072	0.65403	0.21144	0.02883	0.00169	0.00004	0.00000	0.00000
22	1.00000	1.00000	0.99989	0.97786	0.73893	0.28637	0.04787	0.00343	0.00011	0.00000	0.00000
23	1.00000	1.00000	0.99996	0.98811	0.81091	0.37108	0.07553	0.00662	0.00025	0.00000	0.00000
24	1.00000	1.00000	0.99999	0.99392	0.86865	0.46167	0.11357	0.01213	0.00056	0.00001	0.00000
25	1.00000	1.00000	1.00000	0.99703	0.91252	0.55347	0.16313	0.02114	0.00119	0.00003	0.00000
26	1.00000	1.00000	1.00000	0.99862	0.94417	0.64174	0.22440	0.03514	0.00240	0.00007	0.00000
27	1.00000	1.00000	1.00000	0.99939	0.96585	0.72238	0.29637	0.05581	0.00460	0.00016	0.00000
28	1.00000	1.00000	1.00000	0.99974	0.97998	0.79246	0.37678	0.08482	0.00843	0.00036	0.00001
29	1.00000	1.00000	1.00000	0.99989	0.98875	0.85046	0.46234	0.12360	0.01478	0.00076	0.00002
30	1.00000	1.00000	1.00000	0.99996	0.99394	0.89621	0.54912	0.17302	0.02478	0.00154	0.00004
31	1.00000	1.00000	1.00000	0.99998	0.99687	0.93065	0.63311	0.23311	0.03985	0.00297	0.00009
33	1.00000	1.00000	1.00000	1.00000	0.99926	0.97241	0.77926	0.38029	0.09125	0.00976	0.00044
34	1.00000	1.00000	1.00000	1.00000	0.99966	0.98357	0.83714	0.46243	0.13034	0.01663	0.00089
35	1.00000	1.00000	1.00000	1.00000	0.99985	0.99059	0.88392	0.54584	0.17947	0.02724	0.00176
36	1.00000	1.00000	1.00000	1.00000	0.99994	0.99482	0.92012	0.62692	0.23861	0.04290	0.00332
37	1.00000	1.00000	1.00000	1.00000	0.99998	0.99725	0.94695	0.70245	0.30681	0.06507	0.00602
38	1.00000	1.00000	1.00000	1.00000	0.99999	0.99860	0.96602	0.76987	0.38219	0.09514	0.01049
39	1.00000	1.00000	1.00000	1.00000	1.00000	0.99931	0.97901	0.82758	0.46208	0.13425	0.01760
40	1.00000	1.00000	1.00000	1.00000	1.00000	0.99968	0.98750	0.87498	0.54329	0.18306	0.02844
41	1.00000	1.00000	1.00000	1.00000	1.00000	0.99985	0.99283	0.91232	0.62253	0.24149	0.04131
42	1.00000	1.00000	1.00000	1.00000	1.00000	0.99994	0.99603	0.94057	0.69674	0.30865	0.06661
43	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99789	0.96109	0.76347	0.38277	0.09667
44	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99891	0.97540	0.82110	0.46133	0.13563
45	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99946	0.98499	0.86891	0.54132	0.18410
46	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99974	0.99116	0.90702	0.61956	0.24206
47	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99988	0.99498	0.93621	0.69312	0.30865
48	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99995	0.99725	0.95770	0.75957	0.38218
49	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99855	0.97290	0.81727	0.46020
50	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99926	0.98324	0.86542	0.53979
51	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99964	0.98999	0.90405	0.61782
52	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99983	0.99420	0.93383	0.69135
53	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99992	0.99680	0.95589	0.75794
54	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99997	0.99829	0.97161	0.81590
55	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99912	0.98236	0.86437
56	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99956	0.98943	0.90333
57	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99979	0.99389	0.93339
58	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99990	0.99660	0.95569
59	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99996	0.99818
60	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99906	0.98240
61	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99953	0.98951
62	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99978	0.99398
63	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99990	0.99668
64	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996	0.99824
65	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99998	0.99911
66	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999	0.99956
67	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99980
68	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99991
69	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.99996

კუმულატური პუასონის განაწილების ცხრილი

$$P\{Po(\lambda) \leq x\} = \sum_{k=0}^x \frac{\lambda^k}{k!} e^{-\lambda}$$

λ

x	0.01	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.9950	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
1	1.0000	1.0000	0.9998	0.9996	0.9992	0.9988	0.9983	0.9977	0.9970	0.9962
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.6990	0.6526	0.6268	0.5918	0.5578	0.5249	0.4932	0.4628	0.4337	0.4060
2	0.9004	0.8795	0.8571	0.8335	0.8088	0.7834	0.7572	0.7306	0.7037	0.6767
3	0.9743	0.9662	0.9569	0.9463	0.9344	0.9212	0.9068	0.8913	0.8747	0.8571
4	0.9946	0.9923	0.9893	0.9857	0.9814	0.9763	0.9704	0.9636	0.9559	0.9473
5	0.9990	0.9985	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868	0.9834
6	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966	0.9955
7	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992	0.9989
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.3796	0.3546	0.3309	0.3084	0.2873	0.2674	0.2487	0.2311	0.2146	0.1991
2	0.6496	0.6227	0.5960	0.5697	0.5438	0.5184	0.4936	0.4695	0.4460	0.4232
3	0.8386	0.8194	0.7993	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9379	0.9275	0.9162	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
5	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9941	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9985	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
8	0.9997	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
9	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1847	0.1712	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916
2	0.4012	0.3799	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531	0.2381
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4532	0.4335
4	0.7982	0.7806	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6484	0.6288
5	0.9057	0.8946	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851
6	0.9612	0.9554	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991
12	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5
0	0.0166	0.0150	0.0136	0.0123	0.0111	0.0101	0.0091	0.0082	0.0074	0.0067
1	0.0845	0.0780	0.0719	0.0663	0.0611	0.0563	0.0518	0.0477	0.0439	0.0404
2	0.2238	0.2102	0.1974	0.1851	0.1736	0.1626	0.1523	0.1425	0.1333	0.1247
3	0.4142	0.3954	0.3772	0.3594	0.3423	0.3257	0.3097	0.2942	0.2793	0.2650
4	0.6093	0.5896	0.5704	0.5512	0.5321	0.5132	0.4946	0.4763	0.4582	0.4405
5	0.7693	0.7531	0.7367	0.7199	0.7029	0.6858	0.6684	0.6510	0.6335	0.6160
6	0.8786	0.8675	0.8558	0.8436	0.8311	0.8180	0.8046	0.7908	0.7767	0.7622
7	0.9427	0.9361	0.9290	0.9214	0.9134	0.9049	0.8960	0.8867	0.8769	0.8666
8	0.9755	0.9721	0.9683	0.9642	0.9597	0.9549	0.9497	0.9442	0.9382	0.9319
9	0.9905	0.9889	0.9871	0.9851	0.9829	0.9805	0.9778	0.9749	0.9717	0.9682
10	0.9966	0.9959	0.9952	0.9943	0.9933	0.9922	0.9910	0.9896	0.9880	0.9863
11	0.9989	0.9986	0.9983	0.9980	0.9976	0.9971	0.9966	0.9960	0.9953	0.9945
12	0.9997	0.9996	0.9995	0.9993	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980
13	0.9999	0.9999	0.9998	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993
14	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6
0	0.0061	0.0055	0.0050	0.0045	0.0041	0.0037	0.0033	0.0030	0.0027	0.0025
1	0.0372	0.0342	0.0314	0.0289	0.0266	0.0244	0.0224	0.0206	0.0189	0.0174
2	0.1165	0.1088	0.1016	0.0948	0.0884	0.0824	0.0768	0.0715	0.0666	0.0620
3	0.2513	0.2381	0.2254	0.2133	0.2017	0.1906	0.1800	0.1700	0.1604	0.1512
4	0.4231	0.4061	0.3895	0.3733	0.3575	0.3422	0.3272	0.3127	0.2987	0.2851
5	0.5984	0.5809	0.5635	0.5461	0.5289	0.5119	0.4950	0.4783	0.4619	0.4457
6	0.7474	0.7324	0.7171	0.7017	0.6860	0.6703	0.6544	0.6384	0.6224	0.6063
7	0.8560	0.8449	0.8335	0.8217	0.8095	0.7970	0.7841	0.7710	0.7576	0.7440
8	0.9252	0.9181	0.9106	0.9027	0.8944	0.8857	0.8766	0.8672	0.8574	0.8472
9	0.9644	0.9603	0.9559	0.9512	0.9462	0.9409	0.9352	0.9292	0.9228	0.9161

გამოყენებითი სტატისტიკის საფუძვლები

10	0.9844	0.9823	0.9800	0.9775	0.9747	0.9718	0.9686	0.9651	0.9614	0.9574
11	0.9937	0.9927	0.9916	0.9904	0.9890	0.9875	0.9859	0.9841	0.9821	0.9799
12	0.9976	0.9972	0.9967	0.9962	0.9955	0.9949	0.9941	0.9932	0.9922	0.9912
13	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980	0.9977	0.9973	0.9969	0.9964
14	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991	0.9990	0.9988	0.9986
15	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9996	0.9996	0.9995
16	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0159	0.0146	0.0134	0.0123	0.0113	0.0103	0.0095	0.0087	0.0080	0.0073
2	0.0577	0.0536	0.0498	0.0463	0.0430	0.0400	0.0371	0.0344	0.0320	0.0296
3	0.1425	0.1342	0.1264	0.1189	0.1118	0.1052	0.0988	0.0928	0.0871	0.0818
4	0.2719	0.2592	0.2469	0.2351	0.2237	0.2127	0.2022	0.1920	0.1823	0.1730
5	0.4298	0.4141	0.3988	0.3837	0.3690	0.3547	0.3406	0.3270	0.3137	0.3007
6	0.5902	0.5742	0.5582	0.5423	0.5265	0.5108	0.4953	0.4799	0.4647	0.4497
7	0.7301	0.7160	0.7017	0.6873	0.6728	0.6581	0.6433	0.6285	0.6136	0.5987
8	0.8367	0.8259	0.8148	0.8033	0.7916	0.7796	0.7673	0.7548	0.7420	0.7291
9	0.9090	0.9016	0.8939	0.8858	0.8774	0.8686	0.8596	0.8502	0.8405	0.8305
10	0.9531	0.9486	0.9437	0.9386	0.9332	0.9274	0.9214	0.9151	0.9084	0.9015
11	0.9776	0.9750	0.9723	0.9693	0.9661	0.9627	0.9591	0.9552	0.9510	0.9467
12	0.9900	0.9887	0.9873	0.9857	0.9840	0.9821	0.9801	0.9779	0.9755	0.9730
13	0.9958	0.9952	0.9945	0.9937	0.9929	0.9920	0.9909	0.9898	0.9885	0.9872
14	0.9984	0.9981	0.9978	0.9974	0.9970	0.9966	0.9961	0.9956	0.9950	0.9943
15	0.9994	0.9993	0.9992	0.9990	0.9988	0.9986	0.9984	0.9982	0.9979	0.9976
16	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993	0.9992	0.9990
17	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996
18	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003
1	0.0067	0.0061	0.0056	0.0051	0.0047	0.0043	0.0039	0.0036	0.0033	0.0030
2	0.0275	0.0255	0.0236	0.0219	0.0203	0.0188	0.0174	0.0161	0.0149	0.0138
3	0.0767	0.0719	0.0674	0.0632	0.0591	0.0554	0.0518	0.0485	0.0453	0.0424
4	0.1641	0.1555	0.1473	0.1395	0.1321	0.1249	0.1181	0.1117	0.1055	0.0996
5	0.2881	0.2759	0.2640	0.2526	0.2414	0.2307	0.2203	0.2103	0.2006	0.1912
6	0.4349	0.4204	0.4060	0.3920	0.3782	0.3646	0.3514	0.3384	0.3257	0.3134
7	0.5838	0.5689	0.5541	0.5393	0.5246	0.5100	0.4956	0.4812	0.4670	0.4530

8	0.7160	0.7027	0.6892	0.6757	0.6620	0.6482	0.6343	0.6204	0.6065	0.5925
9	0.8202	0.8096	0.7988	0.7877	0.7764	0.7649	0.7531	0.7411	0.7290	0.7166
10	0.8942	0.8867	0.8788	0.8707	0.8622	0.8535	0.8445	0.8352	0.8257	0.8159
11	0.9420	0.9371	0.9319	0.9265	0.9208	0.9148	0.9085	0.9020	0.8952	0.8881
12	0.9703	0.9673	0.9642	0.9609	0.9573	0.9536	0.9496	0.9454	0.9409	0.9362
13	0.9857	0.9841	0.9824	0.9805	0.9784	0.9762	0.9739	0.9714	0.9687	0.9658
14	0.9935	0.9927	0.9918	0.9908	0.9897	0.9886	0.9873	0.9859	0.9844	0.9827
15	0.9972	0.9969	0.9964	0.9959	0.9954	0.9948	0.9941	0.9934	0.9926	0.9918
16	0.9989	0.9987	0.9985	0.9983	0.9980	0.9978	0.9974	0.9971	0.9967	0.9963
17	0.9996	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989	0.9988	0.9986	0.9984
18	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993
19	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9
0	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1	0.0028	0.0025	0.0023	0.0021	0.0019	0.0018	0.0016	0.0015	0.0014	0.0012
2	0.0127	0.0118	0.0109	0.0100	0.0093	0.0086	0.0079	0.0073	0.0068	0.0062
3	0.0396	0.0370	0.0346	0.0323	0.0301	0.0281	0.0262	0.0244	0.0228	0.0212
4	0.0940	0.0887	0.0837	0.0789	0.0744	0.0701	0.0660	0.0621	0.0584	0.0550
5	0.1822	0.1736	0.1653	0.1573	0.1496	0.1422	0.1352	0.1284	0.1219	0.1157
6	0.3013	0.2896	0.2781	0.2670	0.2562	0.2457	0.2355	0.2256	0.2160	0.2068
7	0.4391	0.4254	0.4119	0.3987	0.3856	0.3728	0.3602	0.3478	0.3357	0.3239
8	0.5786	0.5647	0.5507	0.5369	0.5231	0.5094	0.4958	0.4823	0.4689	0.4557
9	0.7041	0.6915	0.6788	0.6659	0.6530	0.6400	0.6269	0.6137	0.6006	0.5874
10	0.8058	0.7955	0.7850	0.7743	0.7634	0.7522	0.7409	0.7294	0.7178	0.7060
11	0.8807	0.8731	0.8652	0.8571	0.8487	0.8400	0.8311	0.8220	0.8126	0.8030
12	0.9313	0.9261	0.9207	0.9150	0.9091	0.9029	0.8965	0.8898	0.8829	0.8758
13	0.9628	0.9595	0.9561	0.9524	0.9486	0.9445	0.9403	0.9358	0.9311	0.9261
14	0.9810	0.9791	0.9771	0.9749	0.9726	0.9701	0.9675	0.9647	0.9617	0.9585
15	0.9908	0.9898	0.9887	0.9875	0.9862	0.9848	0.9832	0.9816	0.9798	0.9780
16	0.9958	0.9953	0.9947	0.9941	0.9934	0.9926	0.9918	0.9909	0.9899	0.9889
17	0.9982	0.9979	0.9977	0.9973	0.9970	0.9966	0.9962	0.9957	0.9952	0.9947
18	0.9992	0.9991	0.9990	0.9989	0.9987	0.9985	0.9983	0.9981	0.9978	0.9976
19	0.9997	0.9997	0.9996	0.9995	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989
20	0.9999	0.9999	0.9998	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996
21	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

გამოყენებითი სტატისტიკის საფუძვლები

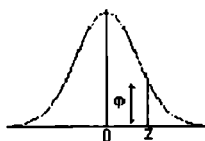
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
1	0.0011	0.0010	0.0009	0.0009	0.0008	0.0007	0.0007	0.0006	0.0005	0.0005
2	0.0058	0.0053	0.0049	0.0045	0.0042	0.0038	0.0035	0.0033	0.0030	0.0028
3	0.0198	0.0184	0.0172	0.0160	0.0149	0.0138	0.0129	0.0120	0.0111	0.0103
4	0.0517	0.0486	0.0456	0.0429	0.0403	0.0378	0.0355	0.0333	0.0312	0.0293
5	0.1098	0.1041	0.0986	0.0935	0.0885	0.0838	0.0793	0.0750	0.0710	0.0671
6	0.1978	0.1892	0.1808	0.1727	0.1649	0.1574	0.1502	0.1433	0.1366	0.1301
7	0.3123	0.3010	0.2900	0.2792	0.2687	0.2584	0.2485	0.2388	0.2294	0.2202
8	0.4426	0.4296	0.4168	0.4042	0.3918	0.3796	0.3676	0.3558	0.3442	0.3328
9	0.5742	0.5611	0.5479	0.5349	0.5218	0.5089	0.4960	0.4832	0.4705	0.4579
10	0.6941	0.6820	0.6699	0.6576	0.6453	0.6329	0.6205	0.6080	0.5955	0.5830
11	0.7932	0.7832	0.7730	0.7626	0.7520	0.7412	0.7303	0.7193	0.7081	0.6968
12	0.8684	0.8607	0.8529	0.8448	0.8364	0.8279	0.8191	0.8101	0.8009	0.7916
13	0.9210	0.9156	0.9100	0.9042	0.8981	0.8919	0.8853	0.8786	0.8716	0.8645
14	0.9552	0.9517	0.9480	0.9441	0.9400	0.9357	0.9312	0.9265	0.9216	0.9165
15	0.9760	0.9738	0.9715	0.9691	0.9665	0.9638	0.9609	0.9579	0.9546	0.9513
16	0.9878	0.9865	0.9852	0.9838	0.9823	0.9806	0.9789	0.9770	0.9751	0.9730
17	0.9941	0.9934	0.9927	0.9919	0.9911	0.9902	0.9892	0.9881	0.9870	0.9857
18	0.9973	0.9969	0.9966	0.9962	0.9957	0.9952	0.9947	0.9941	0.9935	0.9928
19	0.9988	0.9986	0.9985	0.9983	0.9980	0.9978	0.9975	0.9972	0.9969	0.9965
20	0.9995	0.9994	0.9993	0.9992	0.9991	0.9990	0.9989	0.9987	0.9986	0.9984
21	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9995	0.9994	0.9993
22	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997
23	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	11	12	13	14	15	16	17	18	19	20
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
5	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003	0.0002	0.0001
6	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010	0.0005	0.0003
7	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029	0.0015	0.0008
8	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071	0.0039	0.0021
9	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089	0.0050
10	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304	0.0183	0.0108

11	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549	0.0347	0.0214
12	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917	0.0606	0.0390
13	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426	0.0984	0.0661
14	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081	0.1497	0.1049
15	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867	0.2148	0.1565
16	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751	0.2920	0.2211
17	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686	0.3784	0.2970
18	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622	0.4695	0.3814
19	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509	0.5606	0.4703
20	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307	0.6472	0.5591
21	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991	0.7255	0.6437
22	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551	0.7931	0.7206
23	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989	0.8490	0.7875
24	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317	0.8933	0.8432
25	0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554	0.9269	0.8878
26	1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718	0.9514	0.9221
27	1.0000	0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827	0.9687	0.9475
28	1.0000	1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897	0.9805	0.9657
29	1.0000	1.0000	1.0000	0.9999	0.9996	0.9989	0.9973	0.9941	0.9882	0.9782
30	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9967	0.9930	0.9865
31	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9982	0.9960	0.9919
32	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9990	0.9978	0.9953
33	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9988	0.9973
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9985
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
39	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

შენიშვნა: პუასონის განაწილების მისაღებად იხელმძღვანელოთ თანაფარდობით:

$$P\{P_0(\lambda) = x\} = P\{P_0(\lambda) \leq x\} - P\{P_0(\lambda) \leq x-1\}$$

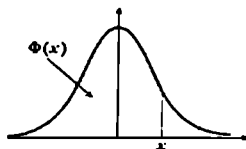
$N(0,1)$ -ის სიმკვრივის $(\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2})$ მნიშვნელობები



Z	0	1	2	3	4	5	6	7	8	9
0.0	.398942	.398922	.398862	.398763	.398623	.398444	.398225	.397966	.397668	.397330
0.1	.396953	.396536	.396080	.395585	.395052	.394479	.393868	.393219	.392531	.391806
0.2	.391043	.390242	.389404	.388529	.387617	.386668	.385683	.384663	.383606	.382515
0.3	.381388	.380226	.379031	.377801	.376537	.375240	.373911	.372548	.371154	.369728
0.4	.368270	.366782	.365263	.363714	.362135	.360527	.358890	.357225	.355533	.353812
0.5	.352065	.350292	.348493	.346668	.344818	.342944	.341046	.339124	.337180	.335213
0.6	.333225	.331215	.329184	.327133	.325062	.322972	.320864	.318737	.316593	.314432
0.7	.312254	.310060	.307851	.305627	.303389	.301137	.298872	.296595	.294305	.292004
0.8	.289692	.287369	.285036	.282694	.280344	.277985	.275618	.273244	.270864	.268477
0.9	.266085	.263688	.261286	.258881	.256471	.254059	.251644	.249228	.246809	.244390
Z	0	1	2	3	4	5	6	7	8	9
1.0	.241971	.239551	.237132	.234714	.232297	.229882	.227470	.225060	.222653	.220251
1.1	.217852	.215458	.213069	.210686	.208308	.205936	.203571	.201214	.198863	.196520
1.2	.194186	.191860	.189543	.187235	.184937	.182649	.180371	.178104	.175847	.173602
1.3	.171369	.169147	.166937	.164740	.162555	.160383	.158225	.156080	.153948	.151831
1.4	.149727	.147639	.145564	.143505	.141460	.139431	.137417	.135418	.133435	.131468
1.5	.129518	.127583	.125665	.123763	.121878	.120009	.118157	.116323	.114505	.112704
1.6	.110921	.109155	.107406	.105675	.103961	.102265	.100586	.098925	.097282	.095657
1.7	.094049	.092459	.090887	.089333	.087796	.086277	.084776	.083293	.081828	.080380
1.8	.078950	.077538	.076143	.074766	.073407	.072065	.070740	.069433	.068144	.066871
1.9	.065616	.064378	.063157	.061952	.060765	.059595	.058441	.057304	.056183	.055079
Z	0	1	2	3	4	5	6	7	8	9
2.0	.053991	.052919	.051864	.050824	.049800	.048792	.047800	.046823	.045861	.044915
2.1	.043984	.043067	.042166	.041280	.040408	.039550	.038707	.037878	.037063	.036262
2.2	.035475	.034701	.033941	.033194	.032460	.031740	.031032	.030337	.029655	.028985
2.3	.028327	.027682	.027048	.026426	.025817	.025218	.024631	.024056	.023491	.022937
2.4	.022395	.021862	.021341	.020829	.020328	.019837	.019356	.018885	.018423	.017971
2.5	.017528	.017095	.016670	.016254	.015848	.015449	.015060	.014678	.014305	.013940
2.6	.013583	.013234	.012892	.012558	.012232	.011912	.011600	.011295	.010997	.010706
2.7	.010421	.010143	.0098712	.0096058	.0093466	.0090936	.0088465	.0086052	.0083697	.0081398
2.8	.0079155	.0076965	.0074829	.0072744	.0070711	.0068728	.0066793	.0064907	.0063067	.0061274
2.9	.0059525	.0057821	.0056160	.0054541	.0052963	.0051426	.0049929	.0048470	.0047050	.0045666

სტანდარტული ნორმალური განაწილების ფუნქციის

$$\Phi(x) = P(N(0,1) \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad \text{მნიშვნელობები } (x \geq 0)$$

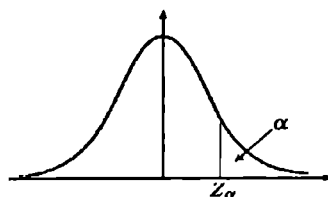


x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

**სტანდარტული ნორმალური განაწილების ფუნქციის
მნიშვნელობები ($-x \leq 0$)**

$-x$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45621	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42466
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10384	0.10204	0.10027	0.09853
1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08692	0.08534	0.08379	0.08226
1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.8	0.03593	0.03515	0.03438	0.03363	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00509	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00403	0.00391	0.00379	0.00368	0.00357
2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.9	0.00187	0.00181	0.00175	0.00170	0.00164	0.00159	0.00154	0.00149	0.00144	0.00140
3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
3.1	0.00097	0.00094	0.00090	0.00087	0.00085	0.00082	0.00079	0.00076	0.00074	0.00071
3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
3.4	0.00034	0.00033	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017

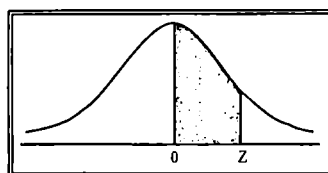
სტანდარტული ნორმალური განაწილების ზედა α -კრიტიკული
ნერტილები (z_α)



α	0.1	0.05	0.025	0.125	0.01	0.005	0.0025	0.001
z_α	1.28	1.64	1.96	2.24	2.33	2.57	2.81	3.08

შენიშვნა: $\Phi(z_\alpha) = 1 - \alpha$

$$\Phi_0(z) = P(0 \leq N(0,1) \leq z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt \quad \text{ფუნქციის მნიშვნელობები (} z \geq 0 \text{)}$$



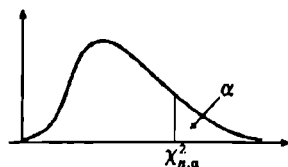
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177

გამოყენებითი სტატისტიკის საფუძვლები

1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49897	.49900
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.2	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.3	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.4	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.5	.49977	.49978	.49978	.49979	.49980	.49981	.49981	.49982	.49983	.49983
3.6	.49984	.49985	.49985	.49986	.49986	.49987	.49987	.49988	.49988	.49989
3.7	.49989	.49990	.49990	.49990	.49991	.49991	.49992	.49992	.49992	.49992
3.8	.49993	.49993	.49993	.49994	.49994	.49994	.49994	.49995	.49995	.49995
3.9	.49995	.49995	.49996	.49996	.49996	.49996	.49996	.49996	.49997	.49997

შენიშვნა: თუ $\alpha \leq 0.5$, მაშინ $\Phi_0(z_\alpha) = 0.5 - \alpha$;

თუ $\alpha \geq 0.5$, მაშინ $\Phi_0(-z_\alpha) = \alpha - 0.5$.



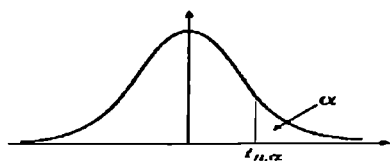
$\chi^2(n)$ (ხი-კვადრატ) განაწილების ზედა

α -კრიტიკული წერტილები ($\chi^2_{n, \alpha}$)

თავისუფლების ხარისხი

α	1	2	3	4	5	6	7	8	9
0.005	7.87946	10.5966	12.8382	14.8603	16.7496	18.5476	20.2778	21.9550	23.5893
0.010	6.63491	9.2103	11.3449	13.2767	15.0863	16.8119	18.4753	20.0902	21.6660
0.025	5.02389	7.3778	9.3484	11.1433	12.8325	14.4494	16.0128	17.5346	19.0228
0.050	3.84146	5.9915	7.8147	9.4877	11.0705	12.5916	14.0671	15.5073	16.9190
0.100	2.70554	4.6052	6.2514	7.7794	9.2364	10.6446	12.0170	13.3616	14.6837
0.900	0.01579	0.2107	0.5844	1.0636	1.6103	2.2041	2.8331	3.4895	4.1682
0.950	0.00393	0.1026	0.3518	0.7107	1.1455	1.6354	2.1674	2.7326	3.3251
0.975	0.00098	0.0506	0.2158	0.4844	0.8312	1.2373	1.6899	2.1797	2.7004
0.990	0.00016	0.0201	0.1148	0.2971	0.5543	0.8721	1.2390	1.6465	2.0879
0.995	0.00004	0.0100	0.0717	0.2070	0.4117	0.6757	0.9893	1.344	1.7349
α	10	11	12	13	14	15	16	17	18
0.005	25.1882	26.7569	28.2996	29.8194	31.3194	32.8013	34.2674	35.7182	37.1564
0.010	23.2093	24.7250	26.2170	27.6882	29.1413	30.5779	32.0001	33.4085	34.8053
0.025	20.4832	21.9201	23.3367	24.7356	26.1190	27.4884	28.8454	30.1910	31.5264
0.050	18.3070	19.6751	21.0261	22.3620	23.6848	24.9958	26.2963	27.5871	28.8693
0.100	15.9872	17.2750	18.5493	19.8119	21.0641	22.3071	23.5418	24.7690	25.9894
0.900	4.8652	5.5778	6.3038	7.0415	7.7895	8.5468	9.3122	10.0852	10.8649
0.950	3.9403	4.5748	5.2260	5.8919	6.5706	7.2609	7.9616	8.6718	9.3905
0.975	3.2470	3.8157	4.4038	5.0088	5.6287	6.2621	6.9077	7.5642	8.2307
0.990	2.5582	3.0535	3.5706	4.1069	4.6604	5.2293	5.8122	6.4078	7.0149
0.995	2.1559	2.6032	3.0738	3.5650	4.0747	4.6009	5.1422	5.6972	6.2648
α	19	20	21	22	23	24	25	26	27
0.005	38.5820	39.9968	41.4011	42.7960	44.1808	45.5586	46.9285	48.2903	49.6452
0.010	36.1907	37.5662	38.9322	40.2895	41.6382	42.9798	44.3144	45.6419	46.9631
0.025	32.8523	34.1696	35.4789	36.7808	38.0755	39.3641	40.6466	41.9233	43.1946
0.050	30.1435	31.4104	32.6706	33.9245	35.1724	36.4150	37.6525	38.8852	40.1133
0.100	27.2036	28.4120	29.6151	30.8133	32.0069	33.1962	34.3816	35.5632	36.7412
0.900	11.6509	12.4426	13.2396	14.0415	14.8480	15.6587	16.4734	17.2919	18.1139
0.950	10.1170	10.8508	11.5913	12.3380	13.0905	13.8484	14.6114	15.3792	16.1514
0.975	8.9065	9.5908	10.2829	10.9823	11.6886	12.4012	13.1197	13.8439	14.5734
0.990	7.6327	8.2604	8.8972	9.5425	10.1957	10.8564	11.5240	12.1981	12.8785
0.995	6.8440	7.4338	8.0337	8.6427	9.2604	9.8862	10.5197	11.1602	11.8076
α	28	29	30	40	50	60	100		
0.005	50.9936	52.3360	53.6720	66.7673	79.4896	91.9547	140.177		
0.010	48.2783	49.5881	50.8922	63.6914	76.1537	88.3810	135.811		
0.025	44.4608	45.7224	46.9793	59.3420	71.4201	83.2984	129.563		
0.050	41.3372	42.5570	43.7730	55.7586	67.5048	79.0823	124.343		
0.100	37.9159	39.0875	40.2560	51.8051	63.1671	74.3972	118.499		
0.900	18.9392	19.7677	20.5992	29.0505	37.6886	46.4589	82.358		
0.950	16.9279	17.7084	18.4927	26.5093	34.7643	43.1880	77.930		
0.975	15.3079	16.0471	16.7908	24.4330	32.3574	40.4818	74.222		
0.990	13.5647	14.2565	14.9535	22.1643	29.7067	37.4849	70.065		
0.995	12.4613	13.1211	13.7867	20.7065	27.9907	35.5345	67.328		

$t(n)$ (სტიუდენტის) განაწილების ზედა α -კრიტიკული
წერტილები ($t_{n,\alpha}$)

მნიშვნელობების დონე (α)

df	.45	.40	.35	.30	.25	.20	.15	.10	.05	.025	.01	.005	.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.3
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.33
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.21
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.611
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
31	0.127	0.256	0.389	0.530	0.682	0.853	1.054	1.309	1.696	2.040	2.453	2.744	3.375
32	0.127	0.255	0.389	0.530	0.682	0.853	1.054	1.309	1.694	2.037	2.449	2.738	3.365
33	0.127	0.255	0.389	0.530	0.682	0.853	1.053	1.308	1.692	2.035	2.445	2.733	3.356

34	0.127	0.255	0.389	0.529	0.682	0.852	1.052	1.307	1.691	2.032	2.441	2.728	3.348
35	0.127	0.255	0.388	0.529	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.340
36	0.127	0.255	0.388	0.529	0.681	0.852	1.052	1.306	1.688	2.028	2.434	2.719	3.333
37	0.127	0.255	0.388	0.529	0.681	0.851	1.051	1.305	1.687	2.026	2.431	2.715	3.326
38	0.127	0.255	0.388	0.529	0.681	0.851	1.051	1.304	1.686	2.024	2.429	2.712	3.319
39	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.304	1.685	2.023	2.426	2.708	3.313
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
41	0.126	0.255	0.388	0.529	0.681	0.850	1.050	1.303	1.683	2.020	2.421	2.701	3.301
42	0.126	0.255	0.388	0.528	0.680	0.850	1.049	1.302	1.682	2.018	2.418	2.698	3.296
43	0.126	0.255	0.388	0.528	0.680	0.850	1.049	1.302	1.681	2.017	2.416	2.695	3.291
44	0.126	0.255	0.388	0.528	0.680	0.850	1.049	1.301	1.680	2.015	2.414	2.692	3.286
45	0.126	0.255	0.388	0.528	0.680	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.281
46	0.126	0.255	0.388	0.528	0.680	0.850	1.048	1.300	1.679	2.013	2.410	2.687	3.277
47	0.126	0.255	0.388	0.528	0.680	0.849	1.048	1.300	1.678	2.012	2.408	2.685	3.273
48	0.126	0.255	0.388	0.528	0.680	0.849	1.048	1.299	1.677	2.011	2.407	2.682	3.269
49	0.126	0.255	0.388	0.528	0.680	0.849	1.048	1.299	1.677	2.010	2.405	2.680	3.265
50	0.126	0.255	0.388	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261

მნიშვნელოვნების დონე (α)

df	.45	.40	.35	.30	.25	.20	.15	.10	.05	.025	.01	.005	.001
51	0.126	0.255	0.387	0.528	0.679	0.849	1.047	1.298	1.675	2.008	2.402	2.676	3.258
52	0.126	0.255	0.387	0.528	0.679	0.849	1.047	1.298	1.675	2.007	2.400	2.674	3.255
53	0.126	0.255	0.387	0.528	0.679	0.848	1.047	1.298	1.674	2.006	2.399	2.672	3.251
54	0.126	0.255	0.387	0.528	0.679	0.848	1.046	1.297	1.674	2.005	2.397	2.670	3.248
55	0.126	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.673	2.004	2.396	2.668	3.245
56	0.126	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.673	2.003	2.395	2.667	3.242
57	0.126	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.672	2.002	2.394	2.665	3.239
58	0.126	0.255	0.387	0.527	0.679	0.848	1.046	1.296	1.672	2.002	2.392	2.663	3.237
59	0.126	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.001	2.391	2.662	3.234
60	0.126	0.254	0.387	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
61	0.126	0.254	0.387	0.527	0.679	0.848	1.045	1.296	1.670	2.000	2.389	2.659	3.229
62	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.670	1.999	2.388	2.658	3.227
63	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.998	2.387	2.656	3.225
64	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.998	2.386	2.655	3.223
65	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.997	2.385	2.654	3.221
66	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.668	1.997	2.384	2.652	3.218
67	0.126	0.254	0.387	0.527	0.678	0.847	1.045	1.294	1.668	1.996	2.383	2.651	3.217
68	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.668	1.995	2.382	2.650	3.215
69	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.995	2.382	2.649	3.213
70	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.211

გამოყენებითი სტატისტიკის საფუძვლები

71	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.380	2.647	3.209
72	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.293	1.666	1.993	2.379	2.646	3.207
73	0.126	0.254	0.387	0.527	0.678	0.847	1.044	1.293	1.666	1.993	2.379	2.645	3.206
74	0.126	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.666	1.993	2.378	2.644	3.204
75	0.126	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.665	1.992	2.377	2.643	3.203
76	0.126	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.665	1.992	2.376	2.642	3.201
77	0.126	0.254	0.387	0.527	0.678	0.846	1.043	1.293	1.665	1.991	2.376	2.641	3.200
78	0.126	0.254	0.387	0.527	0.678	0.846	1.043	1.292	1.665	1.991	2.375	2.640	3.198
79	0.126	0.254	0.387	0.527	0.678	0.846	1.043	1.292	1.664	1.990	2.375	2.640	3.197
80	0.126	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
81	0.126	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.373	2.638	3.194
82	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.664	1.989	2.373	2.637	3.193
83	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.989	2.372	2.636	3.191
84	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.989	2.372	2.636	3.190
85	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.988	2.371	2.635	3.189
86	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.663	1.988	2.371	2.634	3.188
87	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.663	1.988	2.370	2.634	3.187
88	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.662	1.987	2.369	2.633	3.186
89	0.126	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.662	1.987	2.369	2.632	3.184
90	0.126	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.987	2.369	2.632	3.183
91	0.126	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.986	2.368	2.631	3.182
92	0.126	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.986	2.368	2.630	3.181
93	0.126	0.254	0.387	0.526	0.677	0.845	1.042	1.291	1.661	1.986	2.367	2.630	3.180
94	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.291	1.661	1.986	2.367	2.629	3.179
95	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.291	1.661	1.985	2.366	2.629	3.178
96	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.985	2.366	2.628	3.177
97	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.985	2.365	2.627	3.176
98	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.984	2.365	2.627	3.176
99	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.365	2.626	3.175
100	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
140	0.126	0.254	0.386	0.526	0.676	0.844	1.040	1.288	1.656	1.977	2.353	2.611	3.149
160	0.126	0.254	0.386	0.525	0.676	0.844	1.040	1.287	1.654	1.975	2.350	2.607	3.142
180	0.126	0.254	0.386	0.525	0.676	0.844	1.039	1.286	1.653	1.973	2.347	2.603	3.136
200	0.126	0.254	0.386	0.525	0.676	0.843	1.039	1.286	1.653	1.972	2.345	2.601	3.131
∞	0.126	0.253	0.385	0.524	0.675	0.842	1.036	1.282	1.645	1.960	2.327	2.576	3.091

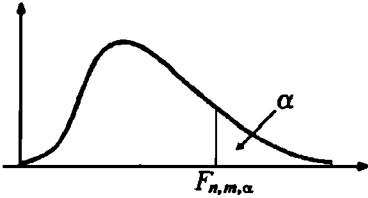
შენიშვნით, რომ, თუ $\mu = \infty$, მაშინ სტიუდენტის კრიტიკული მნიშვნელობები ემთხვევა სტანდარტული ნორმალური განაწილების კრიტიკულ მნიშვნელობებს ($t_{\alpha, \infty} = z_{\alpha}$).

$t(n)$ განაწილების ზედა $\alpha/2$ -კრიტიკული
ნერტილები $t_{n, \alpha/2}$ (ორკუდიანი)



n	α	0.80 0.20	0.90 0.10	0.95 0.05	0.98 0.02	0.99 0.01	0.995 0.005	0.998 0.002	0.999 0.001
1		3.078	6.314	12.706	31.820	63.657	127.321	318.309	636.619
2		1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3		1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5		1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		1.397	1.860	2.306	2.897	3.355	3.833	4.501	5.041
9		1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10		1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11		1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12		1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13		1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14		1.345	1.761	2.145	2.625	2.977	3.326	3.787	4.140
15		1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16		1.337	1.746	2.120	2.584	2.921	3.252	3.686	4.015
17		1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18		1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19		1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20		1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21		1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22		1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23		1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24		1.318	1.711	2.064	2.492	2.797	3.090	3.467	3.745
25		1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26		1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27		1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28		1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29		1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30		1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
∞		1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

$F(n, m)$ (ფიშერის) განაწილებისზედა α -კრიტიკული
წერტილები ($F_{n, m, \alpha}$)



α	n											
	1	2	3	4	5	6	7	8	9	10	11	12
$m = 1$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	39.8	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
0.050	161	200	216	225	230	234	237	239	241	242	243	244
0.025	648	800	864	900	922	937	948	957	963	968	973	977
0.010	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106
$m = 2$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41
0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.41
0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42
$m = 3$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34
0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05
$m = 4$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75
0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37
$m = 5$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27
0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68
0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52
0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89
$m = 6$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37
0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
$m = 7$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67
0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47

$m = 8$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20
0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
$m = 9$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87
0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11
$m = 10$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62
0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
$m = 11$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
0.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43
0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40
$m = 12$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28
0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16
$m = 13$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15
0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
$m = 14$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.07	2.05
0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05
0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
$m = 15$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
0.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96
0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
$m = 16$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89
0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
$m = 17$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
0.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.87	2.82
0.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46

გამოყენებითი სტატისტიკის საფუძვლები

$m = 18$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.95	1.93
0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77
0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
$m = 19$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.93	1.91
0.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
0.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.76	2.72
0.010	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
$m = 20$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89
0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68
0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23
$m = 21$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.90	1.87
0.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
0.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.68	2.64
0.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17
$m = 22$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
0.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
0.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.65	2.60
0.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
$m = 23$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.87	1.84
0.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20
0.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.62	2.57
0.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07
$m = 24$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
0.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18
0.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54
0.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
$m = 25$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.84	1.82
0.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16
0.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.56	2.51
0.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99
$m = 26$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.83	1.81
0.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
0.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.54	2.49
0.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
$m = 27$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.82	1.80
0.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13
0.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.51	2.47
0.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93

$m = 28$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
0.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
0.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.49	2.45
0.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
$m = 29$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.80	1.78
0.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10
0.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.48	2.43
0.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87
$m = 30$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41
0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
$m = 31$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.87	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84	1.81	1.79	1.77
0.050	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.11	2.08
0.025	5.55	4.16	3.57	3.23	3.01	2.85	2.73	2.64	2.56	2.50	2.44	2.40
0.010	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04	2.96	2.88	2.82
$m = 32$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.87	2.48	2.26	2.13	2.04	1.97	1.91	1.87	1.83	1.81	1.78	1.76
0.050	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07
0.025	5.53	4.15	3.56	3.22	3.00	2.84	2.71	2.62	2.54	2.48	2.43	2.38
0.010	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.86	2.80
$m = 34$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.86	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82	1.79	1.77	1.75
0.050	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05
0.025	5.50	4.12	3.53	3.19	2.97	2.81	2.69	2.59	2.52	2.45	2.40	2.35
0.010	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.82	2.76
$m = 36$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.85	2.46	2.24	2.11	2.01	1.94	1.89	1.85	1.81	1.78	1.76	1.73
0.050	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03
0.025	5.47	4.09	3.50	3.17	2.94	2.78	2.66	2.57	2.49	2.43	2.37	2.33
0.010	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72
$m = 38$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.84	2.45	2.23	2.10	2.01	1.94	1.88	1.84	1.80	1.77	1.75	1.72
0.050	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
0.025	5.45	4.07	3.48	3.15	2.92	2.76	2.64	2.55	2.47	2.41	2.35	2.31
0.010	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.69
$m = 40$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71
0.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29
0.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
$m = 42$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.83	2.43	2.22	2.08	1.99	1.92	1.86	1.82	1.78	1.75	1.73	1.71
0.050	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99
0.025	5.40	4.03	3.45	3.11	2.89	2.73	2.61	2.51	2.43	2.37	2.32	2.27
0.010	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64

გამოყენებითი სტატისტიკის საფუძვლები

m = 44	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.82	2.43	2.21	2.08	1.98	1.91	1.86	1.81	1.78	1.75	1.72	1.70
0.050	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98
0.025	5.39	4.02	3.43	3.09	2.87	2.71	2.59	2.50	2.42	2.36	2.30	2.26
0.010	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62
m = 46	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77	1.74	1.71	1.69
0.050	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97
0.025	5.37	4.00	3.42	3.08	2.86	2.70	2.58	2.48	2.41	2.34	2.29	2.24
0.010	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60
m = 48	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.81	2.42	2.20	2.07	1.97	1.90	1.85	1.80	1.77	1.73	1.71	1.69
0.050	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96
0.025	5.35	3.99	3.40	3.07	2.84	2.69	2.56	2.47	2.39	2.33	2.27	2.23
0.010	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71	2.64	2.58
m = 50	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.70	1.68
0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95
0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.26	2.22
0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56
m = 55	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.80	2.40	2.19	2.05	1.95	1.88	1.83	1.78	1.75	1.72	1.69	1.67
0.050	4.02	3.17	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93
0.025	5.31	3.95	3.36	3.03	2.81	2.65	2.53	2.43	2.36	2.29	2.24	2.19
0.010	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53
m = 60	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
0.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
0.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17
0.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
m = 65	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.78	2.39	2.17	2.03	1.94	1.87	1.81	1.77	1.73	1.70	1.67	1.65
0.050	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90
0.025	5.26	3.91	3.32	2.99	2.77	2.61	2.49	2.39	2.32	2.25	2.20	2.15
0.010	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.53	2.47
m = 70	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72	1.69	1.66	1.64
0.050	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89
0.025	5.25	3.89	3.31	2.97	2.75	2.59	2.47	2.38	2.30	2.24	2.18	2.14
0.010	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45
m = 80	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68	1.65	1.63
0.050	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88
0.025	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28	2.21	2.16	2.11
0.010	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42
m = 100	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.64	1.61
0.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85
0.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	2.12	2.08
0.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37

$m = 125$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.75	2.35	2.13	1.99	1.89	1.82	1.77	1.72	1.68	1.65	1.62	1.60
0.050	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83
0.025	5.15	3.80	3.22	2.89	2.67	2.51	2.39	2.30	2.22	2.15	2.10	2.05
0.010	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.39	2.33
$m = 200$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.58
0.050	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
0.025	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18	2.11	2.06	2.01
0.010	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
$m = 400$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.72	2.32	2.10	1.96	1.86	1.79	1.73	1.69	1.65	1.61	1.59	1.56
0.050	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.81	1.78
0.025	5.06	3.72	3.15	2.82	2.60	2.44	2.32	2.22	2.15	2.08	2.03	1.98
0.010	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.29	2.23
$m = 1000$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64	1.61	1.58	1.55
0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76
0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	2.13	2.06	2.01	1.96
0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20
$m = 500$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.71	2.30	2.08	1.95	1.85	1.78	1.72	1.67	1.63	1.60	1.57	1.55
0.050	3.84	3.00	2.61	2.37	2.22	2.10	2.01	1.94	1.88	1.83	1.79	1.75
0.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.12	2.05	2.00	1.95
0.010	6.64	4.61	3.79	3.32	3.02	2.81	2.64	2.51	2.41	2.32	2.25	2.19

	n											
α	14	16	20	24	30	40	50	75	100	200	500	10000
$m = 1$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	61.1	61.4	61.7	62.0	62.3	62.5	62.7	62.9	63.0	63.2	63.3	63.7
0.050	245	246	248	249	250	251	252	253	253	254	254	256
0.025	983	987	993	997	1001	1006	1008	1011	1014	1016	1017	1025
0.010	6143	6170	6209	6235	6261	6287	6303	6324	6334	6350	6360	6366
$m = 2$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	9.42	9.43	9.44	9.45	9.46	9.47	9.47	9.48	9.48	9.49	9.49	9.51
0.050	19.42	19.43	19.45	19.45	19.46	19.47	19.48	19.48	19.49	19.49	19.49	19.54
0.025	39.42	39.44	39.45	39.46	39.46	39.47	39.48	39.48	39.49	39.49	39.50	39.59
0.010	99.43	99.44	99.45	99.46	99.46	99.47	99.48	99.49	99.49	99.49	99.50	99.73
$m = 3$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	5.20	5.20	5.18	5.18	5.17	5.16	5.15	5.15	5.14	5.14	5.14	5.14
0.050	8.71	8.69	8.66	8.64	8.62	8.59	8.58	8.56	8.55	8.54	8.53	8.54
0.025	14.28	14.23	14.17	14.12	14.08	14.04	14.01	13.97	13.96	13.93	13.91	13.92
0.010	26.92	26.83	26.69	26.60	26.50	26.41	26.35	26.28	26.24	26.18	26.15	26.15
$m = 4$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.88	3.86	3.84	3.83	3.82	3.80	3.80	3.78	3.78	3.77	3.76	3.76
0.050	5.87	5.84	5.80	5.77	5.75	5.72	5.70	5.68	5.66	5.65	5.64	5.63
0.025	8.68	8.63	8.56	8.51	8.46	8.41	8.38	8.34	8.32	8.29	8.27	8.26
0.010	14.25	14.15	14.02	13.93	13.84	13.75	13.69	13.61	13.58	13.52	13.49	13.47

გამოყენებითი სტატისტიკის საფუძვლები

$m = 5$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	3.25	3.23	3.21	3.19	3.17	3.16	3.15	3.13	3.13	3.12	3.11	3.10
0.050	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.41	4.39	4.37	4.36
0.025	6.46	6.40	6.33	6.28	6.23	6.18	6.14	6.10	6.08	6.05	6.03	6.01
0.010	9.77	9.68	9.55	9.47	9.38	9.29	9.24	9.17	9.13	9.08	9.04	9.02
$m = 6$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.88	2.86	2.84	2.82	2.80	2.78	2.77	2.75	2.75	2.73	2.73	2.73
0.050	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.73	3.71	3.69	3.68	3.67
0.025	5.30	5.24	5.17	5.12	5.07	5.01	4.98	4.94	4.92	4.88	4.86	4.85
0.010	7.60	7.52	7.40	7.31	7.23	7.14	7.09	7.02	6.99	6.93	6.90	6.89
$m = 7$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.64	2.62	2.59	2.58	2.56	2.54	2.52	2.51	2.50	2.48	2.48	2.47
0.050	3.53	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.27	3.25	3.24	3.23
0.025	4.60	4.54	4.47	4.42	4.36	4.31	4.28	4.23	4.21	4.18	4.16	4.15
0.010	6.36	6.28	6.16	6.07	5.99	5.91	5.86	5.79	5.75	5.70	5.67	5.65
$m = 8$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.48	2.45	2.42	2.40	2.38	2.36	2.35	2.33	2.32	2.31	2.30	2.29
0.050	3.24	3.20	3.15	3.12	3.08	3.04	3.02	2.99	2.97	2.95	2.94	2.93
0.025	4.13	4.08	4.00	3.95	3.89	3.84	3.81	3.76	3.74	3.70	3.68	3.67
0.010	5.56	5.48	5.36	5.28	5.20	5.12	5.07	5.00	4.96	4.91	4.88	4.86
$m = 9$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.35	2.33	2.30	2.28	2.25	2.23	2.22	2.20	2.19	2.17	2.17	2.16
0.050	3.03	2.99	2.94	2.90	2.86	2.83	2.80	2.77	2.76	2.73	2.72	2.71
0.025	3.80	3.74	3.67	3.61	3.56	3.51	3.47	3.43	3.40	3.37	3.35	3.34
0.010	5.01	4.92	4.81	4.73	4.65	4.57	4.52	4.45	4.41	4.36	4.33	4.32
$m = 10$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.26	2.23	2.20	2.18	2.16	2.13	2.12	2.10	2.09	2.07	2.06	2.06
0.050	2.86	2.83	2.77	2.74	2.70	2.66	2.64	2.60	2.59	2.56	2.55	2.54
0.025	3.55	3.50	3.42	3.37	3.31	3.26	3.22	3.18	3.15	3.12	3.09	3.08
0.010	4.60	4.52	4.41	4.33	4.25	4.17	4.12	4.05	4.01	3.96	3.93	3.91
$m = 11$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.18	2.16	2.12	2.10	2.08	2.05	2.04	2.02	2.01	1.99	1.98	1.97
0.050	2.74	2.70	2.65	2.61	2.57	2.53	2.51	2.47	2.46	2.43	2.42	2.41
0.025	3.36	3.30	3.23	3.17	3.12	3.06	3.03	2.98	2.96	2.92	2.90	2.88
0.010	4.29	4.21	4.10	4.02	3.94	3.86	3.81	3.74	3.71	3.66	3.62	3.60
$m = 12$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.12	2.09	2.06	2.04	2.01	1.99	1.97	1.95	1.94	1.92	1.91	1.91
0.050	2.64	2.60	2.54	2.51	2.47	2.43	2.40	2.37	2.35	2.32	2.31	2.30
0.025	3.21	3.15	3.07	3.02	2.96	2.91	2.87	2.82	2.80	2.76	2.74	2.73
0.010	4.05	3.97	3.86	3.78	3.70	3.62	3.57	3.50	3.47	3.41	3.38	3.36
$m = 13$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.07	2.04	2.01	1.98	1.96	1.93	1.92	1.89	1.88	1.86	1.85	1.85
0.050	2.55	2.51	2.46	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.22	2.21
0.025	3.08	3.03	2.95	2.89	2.84	2.78	2.74	2.70	2.67	2.63	2.61	2.60
0.010	3.86	3.78	3.66	3.59	3.51	3.43	3.38	3.31	3.27	3.22	3.19	3.17
$m = 14$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	2.02	2.00	1.96	1.94	1.91	1.89	1.87	1.85	1.83	1.82	1.80	1.80
0.050	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
0.025	2.98	2.92	2.84	2.79	2.73	2.67	2.64	2.59	2.56	2.53	2.50	2.49
0.010	3.70	3.62	3.51	3.43	3.35	3.27	3.22	3.15	3.11	3.06	3.03	3.01

$m = 15$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.99	1.96	1.92	1.90	1.87	1.85	1.83	1.80	1.79	1.77	1.76	1.76
0.050	2.42	2.38	2.33	2.29	2.25	2.20	2.18	2.14	2.12	2.10	2.08	2.07
0.025	2.89	2.84	2.76	2.70	2.64	2.59	2.55	2.50	2.47	2.44	2.41	2.40
0.010	3.56	3.49	3.37	3.29	3.21	3.13	3.08	3.01	2.98	2.92	2.89	2.87
$m = 16$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.95	1.93	1.89	1.87	1.84	1.81	1.79	1.77	1.76	1.74	1.73	1.72
0.050	2.37	2.33	2.28	2.24	2.19	2.15	2.12	2.09	2.07	2.04	2.02	2.01
0.025	2.82	2.76	2.68	2.63	2.57	2.51	2.47	2.42	2.40	2.36	2.33	2.32
0.010	3.45	3.37	3.26	3.18	3.10	3.02	2.97	2.90	2.86	2.81	2.78	2.75
$m = 17$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.93	1.90	1.86	1.84	1.81	1.78	1.76	1.74	1.73	1.71	1.69	1.69
0.050	2.33	2.29	2.23	2.19	2.15	2.10	2.08	2.04	2.02	1.99	1.97	1.96
0.025	2.75	2.70	2.62	2.56	2.50	2.44	2.41	2.35	2.33	2.29	2.26	2.25
0.010	3.35	3.27	3.16	3.08	3.00	2.92	2.87	2.80	2.76	2.71	2.68	2.65
$m = 18$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.90	1.87	1.84	1.81	1.78	1.75	1.74	1.71	1.70	1.68	1.67	1.66
0.050	2.29	2.25	2.19	2.15	2.11	2.06	2.04	2.00	1.98	1.95	1.93	1.92
0.025	2.70	2.64	2.56	2.50	2.44	2.38	2.35	2.30	2.27	2.23	2.20	2.19
0.010	3.27	3.19	3.08	3.00	2.92	2.84	2.78	2.71	2.68	2.62	2.59	2.57
$m = 19$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.88	1.85	1.81	1.79	1.76	1.73	1.71	1.69	1.67	1.65	1.64	1.63
0.050	2.26	2.21	2.16	2.11	2.07	2.03	2.00	1.96	1.94	1.91	1.89	1.88
0.025	2.65	2.59	2.51	2.45	2.39	2.33	2.30	2.24	2.22	2.18	2.15	2.13
0.010	3.19	3.12	3.00	2.92	2.84	2.76	2.71	2.64	2.60	2.55	2.51	2.49
$m = 20$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.86	1.83	1.79	1.77	1.74	1.71	1.69	1.66	1.65	1.63	1.62	1.61
0.050	2.22	2.18	2.12	2.08	2.04	1.99	1.97	1.93	1.91	1.88	1.86	1.84
0.025	2.60	2.55	2.46	2.41	2.35	2.29	2.25	2.20	2.17	2.13	2.10	2.09
0.010	3.13	3.05	2.94	2.86	2.78	2.69	2.64	2.57	2.54	2.48	2.44	2.42
$m = 21$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.84	1.81	1.78	1.75	1.72	1.69	1.67	1.64	1.63	1.61	1.60	1.59
0.050	2.20	2.16	2.10	2.05	2.01	1.96	1.94	1.90	1.88	1.84	1.83	1.81
0.025	2.56	2.51	2.42	2.37	2.31	2.25	2.21	2.16	2.13	2.09	2.06	2.04
0.010	3.07	2.99	2.88	2.80	2.72	2.64	2.58	2.51	2.48	2.42	2.38	2.36
$m = 22$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.83	1.80	1.76	1.73	1.70	1.67	1.65	1.63	1.61	1.59	1.58	1.57
0.050	2.17	2.13	2.07	2.03	1.98	1.94	1.91	1.87	1.85	1.82	1.80	1.78
0.025	2.53	2.47	2.39	2.33	2.27	2.21	2.17	2.12	2.09	2.05	2.02	2.01
0.010	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.36	2.33	2.31
$m = 23$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.81	1.78	1.74	1.72	1.69	1.66	1.64	1.61	1.59	1.57	1.56	1.55
0.050	2.15	2.11	2.05	2.01	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
0.025	2.50	2.44	2.36	2.30	2.24	2.18	2.14	2.08	2.06	2.01	1.99	1.97
0.010	2.97	2.89	2.78	2.70	2.62	2.54	2.48	2.41	2.37	2.32	2.28	2.26
$m = 24$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.80	1.77	1.73	1.70	1.67	1.64	1.62	1.59	1.58	1.56	1.54	1.53
0.050	2.13	2.09	2.03	1.98	1.94	1.89	1.86	1.82	1.80	1.77	1.75	1.73
0.025	2.47	2.41	2.33	2.27	2.21	2.15	2.11	2.05	2.02	1.98	1.95	1.94
0.010	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.37	2.33	2.27	2.24	2.21

გამოყენებითი სტატისტიკის საფუძვლები

$m = 25$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.79	1.76	1.72	1.69	1.66	1.63	1.61	1.58	1.56	1.54	1.53	1.52
0.050	2.11	2.07	2.01	1.96	1.92	1.87	1.84	1.80	1.78	1.75	1.73	1.71
0.025	2.44	2.38	2.30	2.24	2.18	2.12	2.08	2.02	2.00	1.95	1.92	1.91
0.010	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.33	2.29	2.23	2.19	2.17
$m = 26$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.77	1.75	1.71	1.68	1.65	1.61	1.59	1.57	1.55	1.53	1.51	1.50
0.050	2.09	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.73	1.71	1.69
0.025	2.42	2.36	2.28	2.22	2.16	2.09	2.05	2.00	1.97	1.92	1.90	1.88
0.010	2.86	2.78	2.66	2.58	2.50	2.42	2.36	2.29	2.25	2.19	2.16	2.13
$m = 27$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.76	1.74	1.70	1.67	1.64	1.60	1.58	1.55	1.54	1.52	1.50	1.49
0.050	2.08	2.04	1.97	1.93	1.88	1.84	1.81	1.76	1.74	1.71	1.69	1.67
0.025	2.39	2.34	2.25	2.19	2.13	2.07	2.03	1.97	1.94	1.90	1.87	1.85
0.010	2.82	2.75	2.63	2.55	2.47	2.38	2.33	2.26	2.22	2.16	2.12	2.10
$m = 28$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.75	1.73	1.69	1.66	1.63	1.59	1.57	1.54	1.53	1.50	1.49	1.48
0.050	2.06	2.02	1.96	1.91	1.87	1.82	1.79	1.75	1.73	1.69	1.67	1.66
0.025	2.37	2.32	2.23	2.17	2.11	2.05	2.01	1.95	1.92	1.88	1.85	1.83
0.010	2.79	2.72	2.60	2.52	2.44	2.35	2.30	2.23	2.19	2.13	2.09	2.07
$m = 29$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.75	1.72	1.68	1.65	1.62	1.58	1.56	1.53	1.52	1.49	1.48	1.47
0.050	2.05	2.01	1.94	1.90	1.85	1.81	1.77	1.73	1.71	1.67	1.65	1.64
0.025	2.36	2.30	2.21	2.15	2.09	2.03	1.99	1.93	1.90	1.86	1.83	1.81
0.010	2.77	2.69	2.57	2.49	2.41	2.33	2.27	2.20	2.16	2.10	2.06	2.04
$m = 30$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.74	1.71	1.67	1.64	1.61	1.57	1.55	1.52	1.51	1.48	1.47	1.46
0.050	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.70	1.66	1.64	1.62
0.025	2.34	2.28	2.20	2.14	2.07	2.01	1.97	1.91	1.88	1.84	1.81	1.79
0.010	2.74	2.66	2.55	2.47	2.39	2.30	2.25	2.17	2.13	2.07	2.03	2.01
$m = 31$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.73	1.70	1.66	1.63	1.60	1.56	1.54	1.51	1.50	1.47	1.46	1.45
0.050	2.03	1.98	1.92	1.88	1.83	1.78	1.75	1.70	1.69	1.65	1.62	1.61
0.025	2.32	2.26	2.18	2.12	2.06	1.99	1.95	1.89	1.86	1.82	1.79	1.77
0.010	2.72	2.64	2.52	2.45	2.36	2.27	2.22	2.14	2.11	2.04	2.01	1.98
$m = 32$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.72	1.69	1.65	1.62	1.59	1.56	1.53	1.50	1.49	1.46	1.45	1.44
0.050	2.01	1.97	1.91	1.86	1.82	1.77	1.74	1.69	1.67	1.63	1.61	1.60
0.025	2.31	2.25	2.16	2.10	2.04	1.98	1.93	1.88	1.85	1.80	1.77	1.75
0.010	2.70	2.62	2.50	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
$m = 34$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.71	1.68	1.64	1.61	1.58	1.54	1.52	1.49	1.47	1.45	1.43	1.42
0.050	1.99	1.95	1.89	1.84	1.80	1.75	1.71	1.67	1.65	1.61	1.59	1.57
0.025	2.28	2.22	2.13	2.07	2.01	1.95	1.90	1.85	1.82	1.77	1.74	1.72
0.010	2.66	2.58	2.46	2.38	2.30	2.21	2.16	2.08	2.04	1.98	1.94	1.91
$m = 36$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.70	1.67	1.63	1.60	1.56	1.53	1.51	1.47	1.46	1.43	1.42	1.40
0.050	1.98	1.93	1.87	1.82	1.78	1.73	1.69	1.65	1.62	1.59	1.56	1.55
0.025	2.25	2.20	2.11	2.05	1.99	1.92	1.88	1.82	1.79	1.74	1.71	1.69
0.010	2.62	2.54	2.43	2.35	2.26	2.18	2.12	2.04	2.00	1.94	1.90	1.87

m = 38	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.69	1.66	1.61	1.58	1.55	1.52	1.49	1.46	1.45	1.42	1.40	1.39
0.050	1.96	1.92	1.85	1.81	1.76	1.71	1.68	1.63	1.61	1.57	1.54	1.53
0.025	2.23	2.17	2.09	2.03	1.96	1.90	1.85	1.79	1.76	1.71	1.68	1.66
0.010	2.59	2.51	2.40	2.32	2.23	2.14	2.09	2.01	1.97	1.90	1.86	1.84
m = 40	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.68	1.65	1.61	1.57	1.54	1.51	1.48	1.45	1.43	1.41	1.39	1.38
0.050	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
0.025	2.21	2.15	2.07	2.01	1.94	1.88	1.83	1.77	1.74	1.69	1.66	1.64
0.010	2.56	2.48	2.37	2.29	2.20	2.11	2.06	1.98	1.94	1.87	1.83	1.81
m = 42	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.67	1.64	1.60	1.57	1.53	1.50	1.47	1.44	1.42	1.40	1.38	1.37
0.050	1.94	1.89	1.83	1.78	1.73	1.68	1.65	1.60	1.57	1.53	1.51	1.49
0.025	2.20	2.14	2.05	1.99	1.92	1.86	1.81	1.75	1.72	1.67	1.64	1.62
0.010	2.54	2.46	2.34	2.26	2.18	2.09	2.03	1.95	1.91	1.85	1.80	1.78
m = 44	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.66	1.63	1.59	1.56	1.52	1.49	1.46	1.43	1.41	1.39	1.37	1.36
0.050	1.92	1.88	1.81	1.77	1.72	1.67	1.63	1.59	1.56	1.52	1.49	1.48
0.025	2.18	2.12	2.03	1.97	1.91	1.84	1.80	1.73	1.70	1.65	1.62	1.60
0.010	2.52	2.44	2.32	2.24	2.15	2.07	2.01	1.93	1.89	1.82	1.78	1.75
m = 46	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.65	1.63	1.58	1.55	1.52	1.48	1.46	1.42	1.40	1.38	1.36	1.35
0.050	1.91	1.87	1.80	1.76	1.71	1.65	1.62	1.57	1.55	1.51	1.48	1.46
0.025	2.17	2.11	2.02	1.96	1.89	1.82	1.78	1.72	1.69	1.63	1.60	1.58
0.010	2.50	2.42	2.30	2.22	2.13	2.04	1.99	1.91	1.86	1.80	1.76	1.73
m = 48	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.65	1.62	1.57	1.54	1.51	1.47	1.45	1.41	1.40	1.37	1.35	1.34
0.050	1.90	1.86	1.79	1.75	1.70	1.64	1.61	1.56	1.54	1.49	1.47	1.45
0.025	2.15	2.09	2.01	1.94	1.88	1.81	1.77	1.70	1.67	1.62	1.58	1.56
0.010	2.48	2.40	2.28	2.20	2.12	2.02	1.97	1.89	1.84	1.78	1.73	1.71
m = 50	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.64	1.61	1.57	1.54	1.50	1.46	1.44	1.41	1.39	1.36	1.34	1.33
0.050	1.89	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
0.025	2.14	2.08	1.99	1.93	1.87	1.80	1.75	1.69	1.66	1.60	1.57	1.55
0.010	2.46	2.38	2.27	2.18	2.10	2.01	1.95	1.87	1.82	1.76	1.71	1.69
m = 55	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.63	1.60	1.55	1.52	1.49	1.45	1.43	1.39	1.37	1.34	1.32	1.31
0.050	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.53	1.50	1.46	1.43	1.41
0.025	2.11	2.05	1.97	1.90	1.84	1.77	1.72	1.66	1.62	1.57	1.54	1.51
0.010	2.42	2.34	2.23	2.15	2.06	1.97	1.91	1.83	1.78	1.71	1.67	1.64
m = 60	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.62	1.59	1.54	1.51	1.48	1.44	1.41	1.38	1.36	1.33	1.31	1.29
0.050	1.86	1.82	1.75	1.70	1.65	1.59	1.56	1.51	1.48	1.44	1.41	1.39
0.025	2.09	2.03	1.94	1.88	1.82	1.74	1.70	1.63	1.60	1.54	1.51	1.48
0.010	2.39	2.31	2.20	2.12	2.03	1.94	1.88	1.79	1.75	1.68	1.63	1.60
m = 65	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.61	1.58	1.53	1.50	1.47	1.43	1.40	1.37	1.35	1.31	1.29	1.28
0.050	1.85	1.80	1.73	1.69	1.63	1.58	1.54	1.49	1.46	1.42	1.39	1.37
0.025	2.07	2.01	1.93	1.86	1.80	1.72	1.68	1.61	1.58	1.52	1.48	1.46
0.010	2.37	2.29	2.17	2.09	2.00	1.91	1.85	1.77	1.72	1.65	1.60	1.57

გამოყენებითი სტატისტიკის საფუძვლები

$m = 70$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.60	1.57	1.53	1.49	1.46	1.42	1.39	1.36	1.34	1.30	1.28	1.27
0.050	1.84	1.79	1.72	1.67	1.62	1.57	1.53	1.48	1.45	1.40	1.37	1.35
0.025	2.06	2.00	1.91	1.85	1.78	1.71	1.66	1.59	1.56	1.50	1.46	1.44
0.010	2.35	2.27	2.15	2.07	1.98	1.89	1.83	1.74	1.70	1.62	1.57	1.54
$m = 80$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.59	1.56	1.51	1.48	1.44	1.40	1.38	1.34	1.32	1.28	1.26	1.25
0.050	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.43	1.38	1.35	1.33
0.025	2.03	1.97	1.88	1.82	1.75	1.68	1.63	1.56	1.53	1.47	1.43	1.40
0.010	2.31	2.23	2.12	2.03	1.94	1.85	1.79	1.70	1.65	1.58	1.53	1.50
$m = 100$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.57	1.54	1.49	1.46	1.42	1.38	1.35	1.32	1.29	1.26	1.23	1.22
0.050	1.79	1.75	1.68	1.63	1.57	1.52	1.48	1.42	1.39	1.34	1.31	1.28
0.025	2.00	1.94	1.85	1.78	1.71	1.64	1.59	1.52	1.48	1.42	1.38	1.35
0.010	2.27	2.19	2.07	1.98	1.89	1.80	1.74	1.65	1.60	1.52	1.47	1.43
$m = 125$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.56	1.53	1.48	1.44	1.41	1.36	1.34	1.30	1.27	1.23	1.21	1.19
0.050	1.77	1.73	1.66	1.60	1.55	1.49	1.45	1.40	1.36	1.31	1.27	1.25
0.025	1.97	1.91	1.82	1.75	1.68	1.61	1.56	1.49	1.45	1.38	1.34	1.30
0.010	2.23	2.15	2.03	1.94	1.85	1.76	1.69	1.60	1.55	1.47	1.41	1.37
$m = 200$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.54	1.51	1.46	1.42	1.38	1.34	1.31	1.27	1.24	1.20	1.17	1.15
0.050	1.74	1.69	1.62	1.57	1.52	1.46	1.41	1.35	1.32	1.26	1.22	1.19
0.025	1.93	1.87	1.78	1.71	1.64	1.56	1.51	1.44	1.39	1.32	1.27	1.23
0.010	2.17	2.09	1.97	1.89	1.79	1.69	1.63	1.53	1.48	1.39	1.33	1.28
$m = 400$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.52	1.49	1.44	1.40	1.36	1.32	1.29	1.24	1.21	1.17	1.13	1.10
0.050	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.17	1.13
0.025	1.90	1.84	1.74	1.68	1.60	1.52	1.47	1.39	1.35	1.27	1.21	1.16
0.010	2.13	2.05	1.92	1.84	1.75	1.64	1.58	1.48	1.42	1.32	1.25	1.19
$m = 1000$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.51	1.48	1.43	1.39	1.35	1.30	1.27	1.22	1.20	1.15	1.10	1.06
0.050	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
0.025	1.88	1.82	1.72	1.65	1.58	1.50	1.45	1.36	1.32	1.23	1.16	1.10
0.010	2.10	2.02	1.90	1.81	1.72	1.61	1.54	1.44	1.38	1.28	1.19	1.12
$m = 5000$	*	*	*	*	*	*	*	*	*	*	*	*
0.100	1.51	1.47	1.42	1.38	1.34	1.30	1.27	1.22	1.19	1.13	1.09	1.03
0.050	1.69	1.65	1.57	1.52	1.46	1.40	1.35	1.29	1.25	1.17	1.11	1.04
0.025	1.87	1.81	1.71	1.64	1.57	1.49	1.43	1.35	1.30	1.21	1.14	1.05
0.010	2.09	2.00	1.88	1.79	1.70	1.60	1.53	1.42	1.36	1.25	1.16	1.06

სტიუდენტის რანგობრივი სტატისტიკის q -ს მნიშვნელობები

	შესადარებელი საშუალოების რაოდენობა								
α	2	3	4	5	6	7	8	9	10
df = 1	*	*	*	*	*	*	*	*	*
0.05	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
0.01	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
df = 2	*	*	*	*	*	*	*	*	*
0.05	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
0.01	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
df = 3	*	*	*	*	*	*	*	*	*
0.05	4.50	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
0.01	8.26	10.62	12.17	13.33	14.24	15.00	11.55	11.93	12.27
df = 4	*	*	*	*	*	*	*	*	*
0.05	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
0.01	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27
df = 5	*	*	*	*	*	*	*	*	*
0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
0.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
df = 6	*	*	*	*	*	*	*	*	*
0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
0.01	5.25	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
df = 7	*	*	*	*	*	*	*	*	*
0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
0.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
df = 8	*	*	*	*	*	*	*	*	*
0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
0.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
df = 9	*	*	*	*	*	*	*	*	*
0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
df = 10	*	*	*	*	*	*	*	*	*
0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
df = 11	*	*	*	*	*	*	*	*	*
0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
df = 12	*	*	*	*	*	*	*	*	*
0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
0.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
df = 13	*	*	*	*	*	*	*	*	*
0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67

გამოყენებითი სტატისტიკის საფუძვლები

df = 14	*	*	*	*	*	*	*	*	*
0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.94	5.08	5.20
0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
df = 15	*	*	*	*	*	*	*	*	*
0.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
0.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
df = 16	*	*	*	*	*	*	*	*	*
0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
0.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35
df = 17	*	*	*	*	*	*	*	*	*
0.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
df = 18	*	*	*	*	*	*	*	*	*
0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.89	6.02	6.14
df = 19	*	*	*	*	*	*	*	*	*
0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14

შესადარებელი საშუალოების რაოდენობა

α	2	3	4	5	6	7	8	9	10
df = 20	*	*	*	*	*	*	*	*	*
0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
df = 24	*	*	*	*	*	*	*	*	*
0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
0.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
df = 30	*	*	*	*	*	*	*	*	*
0.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.63	4.73
0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
df = 40	*	*	*	*	*	*	*	*	*
0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
0.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
df = 60	*	*	*	*	*	*	*	*	*
0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
0.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
df = 120	*	*	*	*	*	*	*	*	*
0.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
0.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
df = ∞	*	*	*	*	*	*	*	*	*
0.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47
0.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

	შესადარებელი საშუალოების რაოდენობა									
α	11	12	13	14	15	16	17	18	19	20
$df = 1$	*	*	*	*	*	*	*	*	*	*
0.05	50.59	51.96	53.20	54.33	55.30	56.32	57.22	58.04	58.83	59.56
0.01	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
$df = 2$	*	*	*	*	*	*	*	*	*	*
0.05	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
0.01	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
$df = 3$	*	*	*	*	*	*	*	*	*	*
0.05	9.72	9.95	10.15	10.35	10.52	10.69	10.84	10.98	11.11	11.24
0.01	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
$df = 4$	*	*	*	*	*	*	*	*	*	*
0.05	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
0.01	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
$df = 5$	*	*	*	*	*	*	*	*	*	*
0.05	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
0.01	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
$df = 6$	*	*	*	*	*	*	*	*	*	*
0.05	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
0.01	9.30	9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
$df = 7$	*	*	*	*	*	*	*	*	*	*
0.05	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
0.01	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
$df = 8$	*	*	*	*	*	*	*	*	*	*
0.05	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
0.01	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
$df = 9$	*	*	*	*	*	*	*	*	*	*
0.05	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
0.01	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
$df = 10$	*	*	*	*	*	*	*	*	*	*
0.05	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
0.01	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
$df = 11$	*	*	*	*	*	*	*	*	*	*
0.05	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
0.01	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95

გამოყენებითი სტატისტიკის საფუძვლები

შესადარებელი საშუალოების რაოდენობა										
α	11	12	13	14	15	16	17	18	19	20
$df = 12$	*	*	*	*	*	*	*	*	*	*
0.05	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
0.01	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
$df = 13$	*	*	*	*	*	*	*	*	*	*
0.05	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
0.01	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
$df = 14$	*	*	*	*	*	*	*	*	*	*
0.05	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
0.01	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39
$df = 15$	*	*	*	*	*	*	*	*	*	*
0.05	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
0.01	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
$df = 16$	*	*	*	*	*	*	*	*	*	*
0.05	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
0.01	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
$df = 17$	*	*	*	*	*	*	*	*	*	*
0.05	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
0.01	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
$df = 18$	*	*	*	*	*	*	*	*	*	*
0.05	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
0.01	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
$df = 19$	*	*	*	*	*	*	*	*	*	*
0.05	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
0.01	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
$df = 20$	*	*	*	*	*	*	*	*	*	*
0.05	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
0.01	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
$df = 24$	*	*	*	*	*	*	*	*	*	*
0.05	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
0.01	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
$df = 30$	*	*	*	*	*	*	*	*	*	*
0.05	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
0.01	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
$df = 40$	*	*	*	*	*	*	*	*	*	*
0.05	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
0.01	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
$df = 60$	*	*	*	*	*	*	*	*	*	*
0.05	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
0.01	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
$df = 120$	*	*	*	*	*	*	*	*	*	*
0.05	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
0.01	5.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
$df = \infty$	*	*	*	*	*	*	*	*	*	*
0.05	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01
0.01	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

ფაქტორულ სქემაში უბრალე ეფექტების კრიტერიუმის CD-სთვის q-ს მოსაძებნად საჭირო სვეტის დასადგენად გამოიყენება შემდეგი გარდაქმნა:

სქემის ტიპი	სვეტის ნომერი
2X2	3
2X3	5
3X2	5
3X3	7
3X4	8
4X3	8
4X4	10

პირსონის კორელაციის კოეფიციენტის (r) კრიტიკული მნიშვნელობები

α (ცალმხრივი) α (ორმხრივი)					α (ცალმხრივი) α (ორმხრივი)				
n	5%	2.5%	1%	0.5%	n	5%	2.5%	1%	0.5%
3	0.9877	0.9969	0.9995	0.9999	40	0.2638	0.3120	0.3665	0.4026
4	0.9000	0.9500	0.9800	0.9900	41	0.2605	0.3081	0.3621	0.3978
5	0.8054	0.8783	0.9343	0.9587	42	0.2573	0.3044	0.3578	0.3932
6	0.7293	0.8114	0.8822	0.9172	43	0.2542	0.3008	0.3536	0.3887
7	0.6694	0.7545	0.8329	0.8745	44	0.2512	0.2973	0.3496	0.3843
8	0.6215	0.7067	0.7887	0.8343	45	0.2483	0.2940	0.3457	0.3801
9	0.5822	0.6664	0.7498	0.7977	46	0.2455	0.2907	0.3420	0.3761
10	0.5494	0.6319	0.7155	0.7646	47	0.2429	0.2876	0.3384	0.3721
11	0.5214	0.6021	0.6851	0.7348	48	0.2403	0.2845	0.3348	0.3683
12	0.4973	0.5760	0.6581	0.7079	49	0.2377	0.2816	0.3314	0.3646
13	0.4762	0.5529	0.6339	0.6835	50	0.2353	0.2787	0.3281	0.3610
14	0.4575	0.5324	0.6120	0.6614	55	0.2241	0.2656	0.3129	0.3445
15	0.4409	0.5140	0.5923	0.6411	60	0.2144	0.2542	0.2997	0.3301
16	0.4259	0.4973	0.5742	0.6226	65	0.2058	0.2441	0.2880	0.3173
17	0.4124	0.4821	0.5577	0.6055	70	0.1982	0.2352	0.2776	0.3060
18	0.4000	0.4683	0.5425	0.5897	75	0.1914	0.2272	0.2682	0.2957
19	0.3887	0.4555	0.5285	0.5751	80	0.1852	0.2199	0.2597	0.2864
20	0.3783	0.4438	0.5155	0.5614	85	0.1796	0.2133	0.2520	0.2780
21	0.3687	0.4329	0.5034	0.5487	90	0.1745	0.2072	0.2449	0.2702
22	0.3598	0.4227	0.4921	0.5368	95	0.1698	0.2017	0.2384	0.2631
23	0.3515	0.4132	0.4815	0.5256	100	0.1654	0.1966	0.2324	0.2565
24	0.3438	0.4044	0.4716	0.5151	110	0.1576	0.1874	0.2216	0.2446
25	0.3365	0.3961	0.4622	0.5052	120	0.1509	0.1793	0.2122	0.2343
26	0.3297	0.3882	0.4534	0.4958	130	0.1449	0.1723	0.2039	0.2252

გამოყენებითი სტატისტიკის საფუძვლები

27	0.3233	0.3809	0.4451	0.4869	140	0.1396	0.1660	0.1965	0.2170
28	0.3172	0.3739	0.4372	0.4785	150	0.1348	0.1603	0.1898	0.2097
29	0.3115	0.3673	0.4297	0.4705	160	0.1305	0.1552	0.1838	0.2031
30	0.3061	0.3610	0.4226	0.4629	170	0.1266	0.1506	0.1783	0.1971
31	0.3009	0.3550	0.4158	0.4556	180	0.1230	0.1463	0.1733	0.1915
32	0.2960	0.3494	0.4093	0.4487	190	0.1197	0.1424	0.1687	0.1865
33	0.2913	0.3440	0.4032	0.4421	200	0.1166	0.1388	0.1644	0.1818
34	0.2869	0.3388	0.3972	0.4357	210	0.1138	0.1354	0.1604	0.1774
35	0.2826	0.3338	0.3916	0.4296	220	0.1112	0.1323	0.1568	0.1733
36	0.2785	0.3291	0.3862	0.4238	230	0.1087	0.1294	0.1533	0.1695
37	0.2746	0.3246	0.3810	0.4182	240	0.1064	0.1267	0.1501	0.1660
38	0.2709	0.3202	0.3760	0.4128	250	0.1043	0.1241	0.1471	0.1626
39	0.2673	0.3160	0.3712	0.4076	500	0.0736	0.0877	0.1040	0.1151

სპირმენის კორელაციის კოეფიციენტის (r_s) კრიტიკული მნიშვნელობები

"	$\alpha = .100$	$\alpha = .050$	$\alpha = .025$	$\alpha = .010$	$\alpha = .005$	$\alpha = .001$
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.8167	.8667
11	.4182	.5273	.6091	.7000	.7818	.8364
12	.3986	.4965	.5804	.6713	.7455	.8182
13	.3791	.4760	.5549	.6429	.7273	.7912
14	.3626	.4593	.5341	.6220	.6978	.7670
15	.3500	.4429	.5179	.6000	.6747	.7464
16	.3382	.4265	.5000	.5824	.6536	.7265
17	.3260	.4118	.4853	.5637	.6324	.7083
18	.3148	.3994	.4716	.5480	.6152	.6904
19	.3070	.3895	.4579	.5333	.5975	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

კენდალის კრიტიკული მნიშვნელობები

 $\alpha = .05$

m	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
3			64.4	103.9	157.3
4		49.5	88.4	143.3	217.0
5		62.6	112.3	182.4	276.2
6		75.7	136.1	221.4	335.2
8	48.1	101.7	183.7	299.0	453.1
10	60.0	127.8	231.2	376.7	571.0
15	89.8	192.9	349.8	570.5	864.9
20	119.7	258.0	468.5	764.4	1158.7

 $\alpha = .01$

m	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
3			75.6	122.8	185.6
4		61.4	109.3	176.2	265.0
5		80.5	142.8	229.4	343.8
6		99.5	176.1	282.4	422.6
8	66.8	137.4	242.7	388.3	579.9
10	85.1	175.3	309.1	494.0	737.0
15	131.0	269.8	475.2	758.2	1129.5
20	177.0	364.2	641.2	764.4	1158.7

დამატებითი მნიშვნელობები, როცა $n = 3$

m	$\alpha = .05$	$\alpha = .01$
9	54.0	75.9
12	71.9	103.5
14	83.8	121.9
16	95.8	140.2
18	107.7	158.6

მან-უიტნი/უილკოქსონის კრიტერიუმის კრიტიკული მნიშვნელობები
(α : 5%-იანი ცალმხრივი, 10%-იანი ორმხრივი)

n_1

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	—	—	—	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6
3	—	—	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11	11	12	13	13	14
4	—	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18	19	20	21	22	23
5	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25	26	28	29	30	32
6	0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32	34	36	37	39	41
7	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39	41	44	46	48	50
8	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47	49	52	54	57	60
9	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69
10	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62	65	68	72	75	79
11	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69	73	77	81	85	89
12	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77	81	85	90	94	98
13	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84	89	94	98	103	108
14	2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92	97	102	107	113	118
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100	105	111	116	122	128
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107	113	119	125	131	137
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115	121	128	134	141	147
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123	130	136	143	150	157
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130	138	145	152	160	167
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138	146	154	161	169	177
21	5	11	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146	154	162	170	179	187
22	5	12	20	28	36	44	52	60	68	77	85	94	102	111	119	128	136	145	154	162	171	179	188	197
23	5	13	21	29	37	46	54	63	72	81	90	98	107	116	125	134	143	152	161	170	179	189	198	207
24	6	13	22	30	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169	179	188	198	207	217
25	6	14	23	32	41	50	60	69	79	89	98	108	118	128	137	147	157	167	177	187	197	207	217	227

მან-უიტნი/უილკოქსონის კრიტერიუმის კრიტიკული მნიშვნელობები
(α : 2.5%-იანი ცალმხრივი, 5%-იანი ორმხრივი)

n_2

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2	3	3	3	3	3
3	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10
4	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14	15	16	17	17	18
5	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20	22	23	24	25	27
6	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	29	30	32	33	35
7	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41	43	45	48	50	53
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48	50	53	56	59	62
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55	58	61	64	67	71
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62	65	69	73	76	80
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69	73	77	81	85	89
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76	80	85	89	94	98
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83	88	93	98	102	107
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90	96	101	106	111	117
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98	103	109	115	120	126
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105	111	117	123	129	135
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112	119	125	132	138	145
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119	126	133	140	147	154
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127	134	141	149	156	163
21	3	8	15	22	29	36	43	50	58	65	73	80	88	96	103	111	119	126	134	142	150	157	165	173
22	3	9	16	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141	150	158	166	174	182
23	3	9	17	24	32	40	48	56	64	73	81	89	98	106	115	123	132	140	149	157	166	175	183	192
24	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156	165	174	183	192	201
25	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	163	173	182	192	201	211

უილკოქსონის კრიტერიუმის კრიტიკული მნიშვნელობები

α (ცალმხრივი)					α (ცალმხრივი)					α (ცალმხრივი)				
5%	2.5%	1%	0.5%		5%	2.5%	1%	0.5%		5%	2.5%	1%	0.5%	
α (ორმხრივი)					α (ორმხრივი)					α (ორმხრივი)				
n	10%	5%	2%	1%	n	10%	5%	2%	1%	n	10%	5%	2%	1%
1	—	—	—	—	31	163	147	130	118	61	715	672	623	589
2	—	—	—	—	32	175	159	140	128	62	741	697	646	611
3	—	—	—	—	33	187	170	151	138	63	767	721	669	634
4	—	—	—	—	34	200	182	162	148	64	793	747	693	657
5	0	—	—	—	35	213	195	173	159	65	820	772	718	681
6	2	0	—	—	36	227	208	185	171	66	847	798	742	705
7	3	2	0	—	37	241	221	198	182	67	875	825	768	729
8	5	3	1	0	38	256	235	211	194	68	903	852	793	754
9	8	5	3	1	39	271	249	224	207	69	931	879	819	779
10	10	8	5	3	40	286	264	238	220	70	960	907	846	805
11	13	10	7	5	41	302	279	252	233	71	990	936	873	831
12	17	13	9	7	42	319	294	266	247	72	1020	964	901	858
13	21	17	12	9	43	336	310	281	261	73	1050	994	928	884
14	25	21	15	12	44	353	327	296	276	74	1081	1023	957	912
15	30	25	19	15	45	371	343	312	291	75	1112	1053	986	940
16	35	29	23	19	46	389	361	328	307	76	1144	1084	1015	968
17	41	34	27	23	47	407	378	345	322	77	1176	1115	1044	997
18	47	40	32	27	48	426	396	362	339	78	1209	1147	1075	1026
19	53	46	37	32	49	446	415	379	355	79	1242	1179	1105	1056
20	60	52	43	37	50	466	434	397	373	80	1276	1211	1136	1086
21	67	58	49	42	51	486	453	416	390	81	1310	1244	1168	1116
22	75	65	55	48	52	507	473	434	408	82	1345	1277	1200	1147
23	83	73	62	54	53	529	494	454	427	83	1380	1311	1232	1178
24	91	81	69	61	54	550	514	473	445	84	1415	1345	1265	1210
25	100	89	76	68	55	573	536	493	465	85	1451	1380	1298	1242
26	110	98	84	75	56	595	557	514	484	86	1487	1415	1332	1275
27	119	107	92	83	57	618	579	535	504	87	1524	1451	1366	1308

კოლმოგოროვ-სმირნოვის კრიტერიუმის კრიტიკული მნიშვნელობები

n	$\alpha = .20$	$\alpha = .10$	$\alpha = .05$	$\alpha = .02$	$\alpha = .01$	n	$\alpha = .20$	$\alpha = .10$	$\alpha = .05$	$\alpha = .02$	$\alpha = .01$
1	.900	.950	.975	.990	.995	21	.226	.259	.287	.321	.344
2	.684	.776	.842	.900	.929	22	.221	.253	.281	.314	.337
3	.565	.636	.708	.785	.829	23	.216	.247	.275	.307	.330
4	.493	.565	.624	.689	.734	24	.212	.242	.269	.301	.323
5	.447	.509	.563	.627	.669	25	.208	.238	.264	.295	.317
6	.410	.468	.519	.577	.617	26	.204	.233	.259	.290	.311
7	.381	.436	.483	.538	.576	27	.200	.229	.254	.284	.305
8	.358	.410	.454	.507	.542	28	.197	.225	.250	.279	.300
9	.339	.387	.430	.480	.513	29	.193	.221	.246	.275	.295
10	.323	.369	.409	.457	.489	30	.190	.218	.242	.270	.290
11	.308	.352	.391	.437	.468	31	.187	.214	.238	.266	.285
12	.296	.338	.375	.419	.449	32	.184	.211	.234	.262	.281
13	.285	.325	.361	.404	.432	33	.182	.208	.231	.258	.277
14	.275	.314	.349	.390	.418	34	.179	.205	.227	.254	.273
15	.266	.304	.338	.377	.404	35	.177	.202	.224	.251	.269
16	.258	.295	.327	.366	.392	36	.174	.199	.221	.247	.265
17	.250	.286	.318	.355	.381	37	.172	.196	.218	.244	.262
18	.244	.279	.309	.346	.371	38	.170	.194	.215	.241	.258
19	.237	.271	.301	.337	.361	39	.168	.191	.213	.238	.255
20	.232	.265	.294	.329	.352	40	.165	.189	.210	.235	.252
						> 40	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

კრუსკალ-უოლისის სტატისტიკის p -მნიშვნელობები

n_1	n_2	n_3	H	p -მნიშვნელობა	n_1	n_2	n_3	H	p -მნიშვნელობა
2	1	1	2.7000	.500	4	3	2	6.4444	.009
								6.3000	.011
2	2	1	3.6000	.267				5.444	.046
								5.400	.051
2	2	2	4.5714	.067				4.5111	.098
			3.7143	.200				4.4444	.102
3	1	1	3.2000	.300	4	3	3	6.7455	.010
								6.7091	.013
3	2	1	4.2857	.100				5.7909	.046
			3.8571	.133				5.7273	.050
								4.7091	.092
3	2	2	5.3572	.029				4.7000	.101
			4.7143	.048					
			4.5000	.067	4	4	1	6.6667	.010
			4.4643	.105				6.1667	.022
								4.9667	.048
3	3	1	5.1429	.043				4.8667	.054
			4.5714	.100				4.1667	.082
			4.0000	.129				4.0667	.102
3	3	2	6.2500	.011	4	4	2	7.0364	.006
			5.3611	.032				8.8727	.011
			5.1389	.061				5.4545	.046
			4.5556	.100				5.2364	.052
			4.2500	.121				4.5545	.098
								4.4455	.103
3	3	3	7.2000	.004					
			6.4889	.001	4	4	3	7.1439	.010
			5.6889	.029				7.1364	.011
			5.6000	.050				5.5985	.049
			5.0667	.086				5.5758	.051
			4.6222	.100				4.5455	.099
								4.4773	.102
4	1	1	3.5714	.200	4	4	4	7.6538	.008
4	2	1	4.8214	.057				7.5385	.011
			4.5000	.076				5.6923	.049
			4.0179	.114				5.6538	.054
								4.6539	.097
4	3	1	5.8333	.021				4.5001	.104
			5.2083	.050					
			5.0000	.057	5	1	1	3.8571	.143
			4.0556	.093					
			3.8889	.129	5	2	1	5.2500	.036
								5.0000	.048
								4.4500	.071
								4.2000	.095
								4.0500	.119

n_1	n_1	n_1	H	p -მნიშვნელობა	n_1	n_1	n_1	H	p -მნიშვნელობა
5	2	2	6.5333	.008	5	4	4	7.7604	.009
			6.1333	.013				7.7440	.011
			5.1600	.034				5.6571	.049
			5.0400	.056				5.6176	.050
			4.3733	.090				4.6187	.100
			4.2933	.112				4.5527	.102
5	3	1	6.4000	.012	5	5	1	7.3091	.009
			4.9600	.048				6.8364	.011
			4.8711	.052				5.1273	.046
			4.0178	.095				4.9091	.053
			3.8400	.123				4.1091	.086
								4.0364	.105
5	3	2	6.9091	.009	5	5	2	7.3385	.010
			6.8281	.010				7.2692	.010
			5.2509	.049				5.3385	.047
			5.1055	.052				5.2462	.051
			4.6509	.091				4.6231	.097
			4.4945	.101				4.5077	.100
5	3	3	7.0788	.009	5	5	3	7.5780	.010
			6.9818	.011				7.5429	.010
			5.6485	.049				5.7055	.046
			5.5152	.051				5.6264	.051
			4.5333	.097				4.5451	.100
			4.4121	.109				4.5363	.102
5	4	1	6.9545	.008	5	5	4	7.8229	.010
			6.8400	.011				7.7914	.010
			4.9855	.044				5.6657	.049
			4.8600	.056				5.6429	.050
			3.9873	.098				4.5229	.100
			3.9600	.102				4.5200	.101
5	4	2	7.2045	.009	5	5	5	8.0000	.009
			7.1182	.010				7.9800	.010
			5.2727	.049				5.7800	.049
			5.2682	.050				5.6600	.051
			4.5609	.098				4.5600	.100
			4.5182	.101				4.5000	.102
5	4	3	7.4449	.010					
			7.3949	.011					
			5.6564	.049					
			5.6308	.050					
			4.5487	.099					

პასუხები

თავი 1

3. სახელდების სკალა. 5. რიგის სკალა.

თავი 2

5. ა) დაბალი ქულა – 6, მაღალი ქულა – 30; ბ) ექვსმა; გ) შვიდმა; დ) 37. 6. ა) ხუთის ტოლი ინტერვალი; ბ) 50; გ) "5" ნიშნავს 0–4 ინტერვალის სიხშირეს.

თავი 3

3. 0.32. 5. 24.3. 7. ბ) 3.4. 9. 134. 11. 12.89; 12.39. 13. 5.4; 5.7. 19. ა) მოდა 0; ბ) მოდა არ აქვს; დ) თაფლისფერი.

თავი 4

1. ა) 6, $\sigma_1 = 8$; ბ) 15, $\sigma_1 = 11$. 3. $\sigma_1 = 57$, $\sigma_1 = 64$. 5. ა) $d = 6$, $x_{35} = 1$; ბ) $x_{95} = 4$.

9. $\bar{x}_{25} = 1268 / 25 = 50.728$; დისპერსია 10.074. 11. 149.2; 5.32.

თავი 5

5. ბ) 0.904, რეკლამაზე დახარჯულ თანხასა და გასაღების დონეს შორის არსებობს ძლიერი დადებითი კავშირი; გ) 0.82; დ) $Y = 7.43X - 1.206$. 8. $r_s = 0.937$ – გოგონებისათვის პოპულარობას და მომხიბვლელობას შორის ძლიერი დადებითი კავშირია; $r_s = 0.134$ – ბიჭებში პოპულარობას და მომხიბვლელობას შორის სუსტი კავშირია.

9. ა) ხუთი შვილი; ბ) ოთხი, ორი; გ) შენიშვნა: პირსონის კორელაციის კოეფიციენტი

გამოთვალეთ ფორმულით: $r = \frac{s_{xy}}{s_x s_y} = -0.356$.

თავი 6

3. შემოვიღოთ აღნიშვნა: დაბალი – L , მაღალი – H , $P(L; L) = 0.425$, $P(L; H) = 0.075$, $P(2; H) = 0.0144$, $P(2; L) = 0.345$, $P(3; H) = 0.0112$, $P(3; L) = 0.1288$, $P(L) = 0.8994$, $P(H) = 0.425 / 0.8994 = 0.473$. 7. შემოვიღოთ აღნიშვნა: კარგი – G , ცუდი – B , $P(G|B) = 0.05$, $P(B|G) = 0.01$, $P(G) = 0.95$, $P(B) = 0.05$. 9. $P(E_2) = 1100 / 2500 = 11 / 25$, $P(E_3) = 1650 / 2500 = 33 / 50$, $P(F_3) = 300 / 2500 = 3 / 25$, $P(E_2 \cap E_3) = P(F_3) = 300 / 2500 = 3 / 25$,

$P(E_2 | E_4) = \frac{P(E_2 \cap E_4)}{P(E_4)} = \frac{300}{2500} : \frac{850}{2500} = 30 / 85$, $P(E_4 | E_2) = 300 / 1100 = 3 / 11$. 11. ა) $4 / 19$;

ბ) $9 / 19$; გ) $14 / 19$; დ) $15 / 19$. 12. ა) $1 / 56$; ბ) $15 / 56$; გ) $15 / 28$; დ) $5 / 28$. 13. ა) $1 / 21$; ბ) $5 / 42$; გ) $5 / 6$; დ) $10 / 21$. 14. ა) $1 / 2$; ბ) $3 / 8$; გ) $1 / 8$; დ) $1 / 2$. 15. ა) $1 / 2$; ბ) $7 / 16$; გ) $11 / 16$; დ) $15 / 16$. 16. ა) $5 / 16$; ბ) $11 / 16$; გ) $15 / 16$; დ) $1 / 16$. 17. ა) $15 / 32$; ბ) $1 / 2$; გ) $15 / 32$; დ) $15 / 32$. 18. ა) $1 / 2$; ბ) $3 / 16$; გ) $15 / 32$; დ) $5 / 16$; ე) $5 / 8$. 20. $1 / 6$. 21. ყველა შემთხვევაში $11 / 36$. 22. მეტია პირველ შემთხვევაში $1 / 36$ -ით. 23. ორივე ტოლია $1 / 2$ -ის. 24. ა) $1 / 2$; ბ) $5 / 9$; გ) $1 / 3$; დ) $1 / 4$; ე) $7 / 36$; ვ) $1 / 6$. 25. ორივე ტოლია $1 / 3$ -ის. 26. ა) $1 / 36$; ბ) $5 / 9$; გ) $25 / 216$; დ) $1 / 8$; ე) $43 / 216$. 27. ყველა ალბათობა ტოლია $91 / 216$. 28. ა) ორჯერ გაგორებისას $1 / 12$ -ით; ბ) ორჯერ გაგორებისას $5 / 54$ -ით; გ) სამჯერ გაგორებისას $5 / 72$ -ით. 29. ა) $0.5 - p$;

- ბ) $1-2p$; გ) p . 30. 99.4%. 31. ა) 0.35; ბ) 0.875; გ) 0.55. 32. ა) $(0.4 \cdot 0.16)^2 = 0.0041$; ბ) $2 \cdot 0.378 \cdot (1-0.378) = 0.47$; გ) $1-(1-0.45 \cdot 0.84)^2 = 0.61$; დ) $0.96 \cdot 0.84 + 0.96 \cdot 0.84 \cdot 0.16 = 0.98$; ე) $0.45^2 + 0.4^2 + 0.11^2 + 0.04^2 = 0.38$; ვ) $0.38 \cdot 2 \cdot 0.84 \cdot 0.16 = 0.1$. 33. ა) 19/45; ბ) 7/19. 34. $0.99 \cdot 0.001 / (0.99 \cdot 0.001 + 0.01 \cdot 0.999) = 0.09$. 35. ა) 27/80; ბ) 5/9.

თავი 7

5. 0.2641. 7. 7.150; 127.5; 11.29.

9. ა)

X	0	1	2	3	4	5
P	0.00001	0.0004	0.0081	0.0729	0.3281	0.59049

ბ) $np = 5 \cdot 0.9 = 4.5$; $\sqrt{n(1-p)p} = 0.67$.

11. 0.984. 12. 1/4. 13. 8/27. 14. 0.115. 15. ა) 0.06; ბ) 0.11; გ) 0.82; დ) 0.87. 16. ა) 0.11; ბ) 0.08; გ) 0.92; დ) 0.92; ე) 0.5; ვ) 0.08; ზ) 0.025. 17. ა) 0.125; ბ) 0.656; გ) 0.986; დ) 0.735. 18. ა) 1.4-ჯერ; ბ) 3.4-ჯერ; გ) 7.5-ჯერ. 19. 0.92.

თავი 8

1. ა) 0.4332; ბ) 0.4515; გ) $0.4452+0.4772$; დ) $0.4938-0.3413$; ე) 0.025; ვ) 0.0495; ზ) $0.5+0.3413$. 5. ა) $\mu = 24$, $\sigma^2 = 16.8$, $P(X > 35) = P(Z > 2.68)$; ბ) $0.5-0.4963$.

9. ა) $0.5 - 0.3944$; ბ) 2.06; გ) 0.1. 11. -0.68; 0; 0.68. 12. ა) ინდექსების განაწილება გადახრილია მარჯვნივ, ნულიდან საშუალომდე თავსდება მხოლოდ 1.425 (და არა 3) სტანდარტული გადახრა; ბ) $Z = 3.73$. ასეთი Z და ამაზე მეტი სიდიდის მიღების ალბათობა, თუ ჩავთვლით, რომ ინდექსებს აქვთ ნორმალური განაწილება არის 0.0001.

15.

ერთი მონეტა		ორი მონეტა		სამი მონეტა		ოთხი მონეტა	
ფარდ. სიხშ.	ალბათობა	ფარდ. სიხშ.	ალბათობა	ფარდ. სიხშ.	ალბათობა	ფარდ. სიხშ.	ალბათობა
0	0.5	0	0.25	0	0.125	0	0.0625
1	0.5	1/2	0.5	1/3	0.375	1/4	0.2500
		1	0.25	2/3	0.375	2/4	0.3750
				1	0.125	3/4	0.2500
						1	0.0625

ე) განაწილების ფორმა მონეტების რაოდენობის ზრდასთან ერთად უფრო ზარისებური ხდება. 23. ა) 12.9. 25. 23.756.

თავი 9

1. ა) 58, $18/\sqrt{9} = 6$; ბ) 6; გ) საშუალო არ შეიცვლება, სტანდარტული გადახრა ორჯერ შემცირდება; დ) იმისათვის, რომ საშუალოს სტანდარტული შეცდომა ორჯერ შემცირდეს, შერჩევის მოცულობა ოთხჯერ უნდა გაიზარდოს. 3. ა) 17.5; ბ) $2.2/\sqrt{35}$; გ) $n = 35$, $N = 100$, $n/N = 0.35 > 0.05$ ამიტომ საჭიროა სტანდარტული გადახრა გავამრავლოთ $\sqrt{\frac{100-35}{100-1}}$ -ზე. 5. $z = \frac{395-456}{13.51} = 4.52$, $P(Z < -4.54) = 0$, სტუდენტების შერჩევა

არ არის შემთხვევითი. 7. $8/\sqrt{40}$. 9. $54.4 \pm 1.65 \frac{14}{\sqrt{200}} = (38.12, 70.88)$. 11. $z = 11$ -ის მიღების

ალბათობა ძალიან პატარაა, უნდა დავასკვნათ, რომ $\mu \neq 35$. 13. $z = 2.99$, ნულოვანი იპოთეზა უნდა უარყოფთ. 15. ა) $(24.86 - 1.96 \cdot 3.92/8, 24.86 + 1.96 \cdot 3.92/8) = (23.9, 25.82)$; ბ) 236. 16. ა) $\bar{x} = 4.8125$, $s = 0.81009$, $n = 8 < 30$, ამიტომ 95%-იან ნდობის ინტერვალს აქვს სახე; ბ) პირველი კვარტილი - 4.125, მედიანა - 4.5, მესამე კვარტილი - 5.625; მანძილი მესამე კვარტილიდან მედიანამდე სამჯერ მეტია ვიდრე მანძილი მედიანიდან პირველ კვარტილამდე, ანუ სახეზეა მკვეთრად მარჯვნივ გადახრილი განაწილება; ნდობის ინტერვალი არ წარმოადგენს კარგ შეფასებას პოპულაციის საშუალოსათვის.

19. $0.28 \pm 1.96 \sqrt{\frac{0.28 \times 0.72}{50}}$. 23. ა) $4.032 \pm 1.96 \cdot 0.05$, $4.032 \pm 2.58 \cdot 0.05$, 99%-იანი ნდობის ინტერვალი უფრო განიერია, ვინაიდან ცდომილება მეტია; ბ) არ არსებობს; გ) დიახ, მოიცავენ. 25. $z = 1.43$, H_0 ჰიპოთეზას, $p = 0.5$ რომ ვერ უარყოფთ. 27. ა) $2.8 \pm 2.33 \cdot 0.05 = (2.6835, 2.9165)$; ბ) პოპულაციის სტანდარტული გადახრა - $0.05 \cdot \sqrt{1964} = 2.22$, საშუალო ნულისგან (მინიმალური მნიშვნელობა) დაშორებულია $2.8/2.22 = 1.26$ სტანდარტული გადახრით (და არა სამი სტანდარტული გადახრით), ამიტომ განანილება მარჯვნივ ასიმეტრიულია.

თაპი 10

2. ა) $(7.92 - 5.80) / \sqrt{\frac{3.45^2}{48} + \frac{2.87^2}{53}} = 3.34$, $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$, $z_{კრიტ.} = 1.65$, ნულოვანი ჰიპოთეზა უნდა უარყოფთ; ბ) $P = P(Z > 3.34) = 0.0004$; გ) $(1.075, 3.165)$;

დ) $1.645 \sqrt{\frac{3.45^2}{n} + \frac{2.87^2}{n}} = 0.7$, აქედან $n = 112$. 6. $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 > \mu_2$,

$s_p^2 = (11 \times 2.065^2 + 11 \times 1.782^2) / (11 + 11) = 3.72$, $t = (28.58 - 27.42) / \sqrt{\frac{3.72}{12} + \frac{3.72}{12}} = 1.473$,

$df = 22$, $\alpha = 0.1$, $t_{კრიტ.} = 1.321$, H_0 უნდა უარყოფთ. 7.) $H_0: p_1 = p_2$, $H_1: p_1 < p_2$,

$\hat{p}_1 = \frac{135}{218} = 0.62$, $\hat{p}_2 = \frac{220}{300} = 0.7$, $\hat{p} = \frac{135 + 220}{218 + 300} = 0.69$, $\sqrt{0.69 \times 0.31 \left(\frac{1}{218} + \frac{1}{300} \right)} = 0.041$,

$z = \frac{0.62 - 0.73}{0.041} = -2.68$, ნულოვანი ჰიპოთეზა უნდა უარყოფთ; ბ) 99%-იანი ნდობის

ინტერვალი მწვევლთა და არამწვევლთა პროპორციების სხვაობისათვის - $(-0.216, -0.004)$. ქირურგიული ოპერაციის შემდეგ მწვევლებს უფრო დიდი დრო

სჭირდებათ ანესთეზიიდან გამოსასვლელად ვიდრე არამწვევლებს. 10. ა) $\hat{p} = \frac{144}{700}$;

ბ) $\hat{p}_1 = 0.2$, $\hat{p}_2 = 0.22$, $H_0: p_1 = p_2$, $H_1: p_1 < p_2$, სტანდარტული შეცდომა -

$\sqrt{0.206 \cdot 0.794 \cdot \left(\frac{1}{500} + \frac{1}{200} \right)} = 0.034$, $z = \frac{0.2 - 0.22 - 0}{0.034} = -0.59$, $z_{კრიტ.} = 1.645$, ნულოვანი

ჰიპოთეზას ვერ უარყოფთ. 13. ა) $(4.1 - 2.8) \pm 1.96 \cdot \sqrt{\frac{3.3^2}{350} + \frac{2.0^2}{1965}} = 1.3 \pm 0.4 = (0.9, 1.7)$,

საშუალო მნიშვნელოვნად მეტი იყო 1982 წელს; ბ) $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$, $z = \frac{4.1 - 2.8}{0.18} = 7.1$,

$p < 0.0001$; განაწილება მარჯვნივ არის გადახრილი, მინიმალური შესაძლო მნიშვნელობა – ნული საშუალოსაგან 1982 წელს მხოლოდ მხოლოდ $4.1/3.3 = 1.2$ სტანდარტული გადახრით განსხვავდება; 1994 წელს კი მხოლოდ 1.4 სტანდარტული გადახრით.

თავი 11

1. ა) $n = 6$, $k = 3$, $N = 18$; ბ) $\bar{x}_3 = 15$, $\bar{x} = \frac{30 + 23 + 15}{3} = 22.78$. 2. $SSR = 173.75$, $MSR = 57.917$,

$SSE = 2647$, $MSE = 165.4375$, $F = 0.35$, $F_{კრიტ.} = 3.24$, $\alpha = 0.05$. ნულოვან ჰიპოთეზას $\mu_1 = \mu_2 = \mu_3 = \mu_4$ ვერ უარვყოფთ. ამ შერჩევების მიხედვით საცხოვრებელი ადგილი არ მოქმედებს მომზადების დონეზე. 10. $MSR = 475.66/4 = 118.92$, $MSE = 161.30/20 = 8.065$, $F = 118.92/8.065 = 14.745$, $F_{კრიტ.} = F(4, 20, 0.05) = 2.866$, ნულოვანი ჰიპოთეზა, მდინარის ნაპირზე მომუშავე სანარმოები არ მოქმედებენ მდინარის ტემპერატურაზე, უნდა უარვყოთ. 13. ა) $H_0: \mu_1 = \mu_2 = \mu_3$, H_1 : ყველა საშუალო ერთმანეთის ტოლი არ არის; ბ) $F = 25.6$, $F_{კრიტ.} = F(2, 940, 0.05) = 2.9957$, ნულოვანი ჰიპოთეზა უნდა უარვყოთ;

გ) საჭიროა ტუკის კრიტერიუმის გამოყენება; დ) როცა $\alpha = 0.05$, $CD = 4.12 \cdot \sqrt{\frac{1.67}{940}} = 0.174$,

$\mu_1 \neq \mu_3$, $\mu_2 \neq \mu_3$, $\mu_1 = \mu_2$. 14. გ) $F = \frac{1035.67}{51.8} = 20$, $df_1 = 3$, $df_2 = 12$, $F_{კრიტ.}(0.05) = 3.49$,

H_0 უნდა უარვყოთ, $CD = 4.2 \cdot \sqrt{\frac{51.8}{12}} = 8.73$. 17. ა) H_0 : სამივე ქარხნის საბურავების

პროფილი ერთნაირია, ალტერნატივა: სამივე ქარხნის საბურავების პროფილი ერთნაირი არ არის;

ბ)

	SS	df	MS	F
ქარხნებს შორის	50	2	25	11.848
საბურავებს შორის	70	19	3.68	1.744
შეცდომა	80	38	2.11	
სულ	200	59		

როცა $\alpha = 0.05$, $F_{კრიტ.} = F(2, 38) = 2.84$, ნულოვანი ჰიპოთეზა უნდა უარვყოთ; გ) განსხვავება არის; დ) საქმე გვაქვს ცვალებადობასთან, რომელიც გამოწვეულია სხვადასხვა ქარხნით და ცვალებადობასთან, რომელიც გამოწვეულია საბურავების ინდივიდუალური განსხვავებით. 19. ა) A ფაქტორის 4 დონე; ბ) B ფაქტორის 3 დონე; გ) ცდაში მონაწილეობდა 72 ცდისპირი; დ) როცა $\alpha = 0.05$, $F_{კრიტ.} = F(3, 60) = 2.79$, $F_{კრიტ.}(2, 60) = 3.18$,

$F_{კრიტ.} = F(6, 60) = 2.29$. B ფაქტორის მოქმედება მნიშვნელოვანია, $F = 4 > 3.18$.

20. ა) 11 კონკურსანტი; ბ) 9 მსაჯი; გ) კონკურსანტებს შორის არის განსხვავება, $F = 205.9 > F_{კრიტ.} = F(10, 297, 0.05) = 1.87$; დ) მსაჯებს შორის განსხვავება არ არის,

$F = 1.45 < F_{კრიტ.} = F(8, 297, 0.05) = 1.98$; ე) ურთიერთქმედება უმნიშვნელოა,

$F = 1.01 < F_{კრიტ.} = F(80, 297, 0.05) = 1.42$. ურთიერთქმედება მნიშვნელოვანი რომ

ყოფილიყო, ეს აღნიშნავდა „პატრიოტულ“ დამოკიდებულებას თანამემამულე კონკურსანტების მიმართ.

თავი 12

1. მოსალოდნელი სიხშირე – 1632, $\chi^2 = 40.345$, $\chi^2_{კრიტ.} = 9.236$, მყიდველის რაოდენობა დამოკიდებულია კვირის დღეზე. 2. ა) მოსალოდნელი სიხშირე $= (3+7+8+18)/4 = 9 > 5$; ბ) ნულოვანი ჰიპოთეზა: უბედური შემთხვევის რაოდენობა არ არის დამოკიდებული მუშის პროფესიონალიზმის დონეზე, ალტერნატივა: უბედური შემთხვევის რაოდენობა დამოკიდებულია მუშის პროფესიონალიზმის დონეზე, $\chi^2 = 2.334$, $\chi^2_{კრიტ.} = 5.991$, ნულოვან ჰიპოთეზას უარყოფთ. 6. $\chi^2 = 23.17$, $\chi^2_{კრიტ.} = 12.592$, ნულოვან ჰიპოთეზას უარყოფთ, გარიგების რაოდენობა დამოკიდებულია კლიენტის ოჯახურ მდგომარეობაზე. 9. როცა შერჩევის მოცულობა 10-ზე მეტია, მან-უიტნის სტატისტიკას გააჩნია

მოახლოებით ნორმალური განაწილება საშუალოთი $\frac{n_1 n_2}{2}$ და სტანდარტული გადახრით $\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$. $\mu = 324$, $\sigma = 37.47$, $U_1 = 27 \times 24 + \frac{27(27+1)}{2} - \sum R_i = 678$,

$U_2 = 624$, $U = 624$, $Z = \frac{624 - 324}{37.47} = 8$, ნულოვანი ჰიპოთეზა, პროფესიონალები და მომხმარებლები ერთნაირად აფასებენ პროგრამას, უნდა უარყოფთ. 10. უნდა გამოვიყენოთ უილკოქსონის კრიტერიუმი: $T = 15$; როცა $\alpha = 0.05$, $T_{კრიტ.} = 25$; ნულოვანი ჰიპოთეზა უნდა უარყოფთ და დავასკვნათ, რომ კურსები ეფექტურია. 11. ა) ყველა შერჩევა მიღებული უნდა იყოს ნორმალური პოპულაციიდან; ბ) უნდა გამოვიყენოთ კრუსკალ-უოლისის კრიტერიუმი: $H = 5.416$, $H_{კრიტ.} = 5.991$, ნულოვანი ჰიპოთეზა, სამი წამყვანი ფირმის ფეხსაცმლის ხარისხი ერთნაირია, უნდა დავადასტუროთ. 14. $E_1 = 58.45$, $E_2 = 63.78$, ..., $E_{24} = 177.59$, $\chi^2 = 51.28$, $\chi^2_{კრიტ.} = 7.815$. ნულოვანი ჰიპოთეზა, ავტო-საგზაო შემთხვევის რაოდენობა არ არის დამოკიდებული ასაკზე, უნდა უარყოფთ. ახალგაზრდების სადაზღვევო გადასახადი უნდა იყოს მეტი. 17. ა) დაავადებულია 500, ჯანმრთელია 99500; ბ) $475/500 = 0.95$, $25/500 = 0.05$; გ) $0.95/0.05 = 19$; დ) $4975/99500 = 0.05$, $94525/99500 = 0.95$; ე) $0.05/0.95 = 0.053$; ვ) 361-ჯერ; ზ) დამოკიდებულია; თ) $475/5450 = 8.7\%$, $4975/5450 = 91.3\%$.

თავი 13

1. ბ) $r = -0.567$, $t = -2.065$, $t_{კრიტ.} = -1.833$. 3. $H_0: \rho = 0$, $H_1: \rho \neq 0$, $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.104\sqrt{333}}{\sqrt{1-0.01}} = 1.92$, $t_{კრიტ.} = 1.96$, H_0 -ს ვადასტურებთ. 6. ა) $t_{კრიტ.} = 2.145$, $(Y, X_1)t = 6.01$, $(Y, X_2)t = 1.70$, $(Y, X_3)t = 2.80$, $(Y, X_4)t = -2.03$, $(X_1, X_2)t = 1.12$, $(X_1, X_3)t = 2.65$, $(X_1, X_4)t = -1.31$, $(X_2, X_3)t = 1.65$, $(X_2, X_4)t = -1.87$, $(X_3, X_4)t = -2.46$; გ) 14.122 + 63.2 X_1 + 10.1 X_2 + 31.5 X_3 - 55.5 X_4 ; დ) 0.775; ე) X_1 ; ვ) $X_1: (30.27, 96.03)$, $X_2: (-21.40, 41.60)$, $X_3: (-358.30, 421.30)$, $X_4: (-192.12, 81.20)$. 7. ა) სამი; ბ) თუ დანყებითა $X_2 = 1$, სხვა შემთხვევაში $X_2 = 0$, თუ არასრული საშუალო $X_3 = 1$, სხვა შემთხვევაში $X_3 = 0$, თუ საშუალო ზოგადი $X_4 = 1$, სხვა შემთხვევაში $X_4 = 0$, საშუალო პროფესიონალიზმის თუ დანყებითა $X_2 = 1$, სხვა შემთხვევაში $X_2 = 0$, $X_3 = 0$, $X_4 = 0$. 8. ბ) $-774 + 66.2X_1$; გ) $-586.256 + 22.861X_1 + 2302.267X_2 + 1869.813X_3$.

ლიტერატურა

1. ნ. ლაზრიევა, მ. მანია, გ. მარი, ა. მოსიძე, ა. ტორონჯაძე, თ. ტორონჯაძე, თ. შერვაშიძე. ალბათობის თეორია და მათემატიკური სტატისტიკა ეკონომისტებისათვის. ფონდი «ევრაზია», თბილისი, 2000. 661 გვ.
2. პარალელ კისი. სტატისტიკა სოციალურ მეცნიერებებში. თსუ გამომცემლობა, თბილისი, 2008. 532 გვ.
3. ო. ფურთუხია. აღწერითი სტატისტიკა, ალბათობა, სტატისტიკური დასკვნების თეორია. თბილისი, 2008. 418 გვ.
4. ო. ფურთუხია. ალბათობა და სტატისტიკა მაგალითებსა და ამოცანებში. თბილისი, 2009. 268 გვ.
5. ო. ფურთუხია, ზ. ციგროშვილი, ქ. მანჯგალაძე. უმაღლესი მათემატიკა, ნაწილი III, ალბათობა და მათემატიკური სტატისტიკა. თსუ გამომცემლობა, თბილისი, 2009. 251 გვ.
6. ო. ფურთუხია. ალბათურ-სტატისტიკური ამოცანები. თსუ გამომცემლობა, თბილისი, მეორე გამოცემა, 2013. 562 გვ.
7. კეკელია ნორა, სოხაძე გრიგოლი, ფაცაცია მზევინარი. მათემატიკის ელემენტები (რიცხვითი სისტემები, ფუნქციები, ალბათობა). თბილისი, 2014, 334 გვ.
8. Denis Anthony. Statistics for Health, Life and Social Sciences. Free Book. Bookboon.com. 2011. 292 p.
9. Dato N. M. de Gruijter, Leo J. Th. Van der Kamp. Statistical Test Theory for the Behavioral Sciences. Chapman&Hall/CRC. 2008. 282 p.
10. Mark S. Handcock, Martina Morris. Relative Distribution Methods in the Social Sciences. Springer. 1999. 280 p.
11. David A. Kenny. Statistics for the Social and Behavioral Sciences. Library of Congress Cataloging-in-Publication Data. 1987. 434 p.
12. James Stevens. Applied Multivariate Statistics for the Social Sciences. Lawtence Erlbaum Associates, Publishers. 2002. 699 p.
13. Артемиева Е. Ю. Сборник задач по теории вероятностей и математической статистике для психологов. Изд-во МГУ. 1969. 92 стр.
14. Ахтямов А. М. Математика для социологов и экономистов. ФИЗМАТЛИТ. Москва. 2004. 464 стр.
15. Крамер Д. Математическая обработка данных в социальных науках. Современные методы. Издательский центр «Академия», Москва. 2007. 288 стр.
16. Мациевский С. В. Высшая математика для гуманитариев. Изд-во Российского государственного университета им. И. Канта. Москва. 2010. 299 с.
17. Наследов А. Д. Математические методы психологического исследования. Анализ и интерпретация данных. Изд-во «Речь», Санкт-Петербург. 2004. 388 стр.
18. Рамси Д. Статистика для «Чайников». Изд-во «Диалектика», Москва-Санкт-Петербург-Киев. 2008. 316 стр.
19. Ростовцев П. С., Ковалева Г. Д., Нуртдинов А. Н. Курс лекций по теории вероятностей и математической статистике для социологов. Изд-во Новосибирского государственного университета, Новосибирск. 2004. 95 стр.
20. Megan J. Cleark, John A. Randal, A First Course in Applied Statistics. 2005, 309 p.
21. Alan Agresti, Barbara Finlay Statistical Methods for the Social Sciences. 1997, 667 p.
22. David F., G. Rochni, Paattrick W. Shannon, Business Statistics. 1981, 931 p.

ა

ალბათობის თვისებები 107
ალბათური და არაალბათური შერჩევა
და ა.შ. 17
აღწერა და დასკვნა 7, 8
აღწერითი სტატისტიკები 8, 40

ბ

ბაიესის ფორმულა 111
ბერნულის ფორმულა 120, 123
ბინომური ალბათობების გამოთვლა 119
ბინომური განაწილება 118, 119, 120, 123,
124, 126, 155, 156
ბინომური განაწილების ნორმალური
აპროქსიმაცია 134
ბინომური განაწილების თვისებები 118
ბინომური შემთხვევითი სიდიდის
საშუალო და დისპერსია 122

გ

გაბნევის დიაგრამა 84, 86, 87, 88, 89, 90,
94, 98, 101, 102, 103, 105, 248, 249, 258,
259, 261
გაბნევის დიაპაზონი 26, 40, 59, 60, 61, 78,
79, 80
გადაწყვეტილების მიღება 71, 145, 148,
149, 150
განაწილების კანონი 110, 276

დ

დამოკიდებული სიდიდე 83
დამოუკიდებელი სიდიდე 83
დამოუკიდებლობის ხი-კვადრატ
კრიტერიუმი 213, 230
დამრგვალების ზოგადი წესი 40
დასკვნები კორელაციის კოეფიციენტის
შესახებ 240, 244
დასკვნითი სტატისტიკური მეთოდები
8, 9
დენდროგრამა (ხისებრი დიაგრამა) 108

დეტერმინაციის კოეფიციენტი 100, 102,
104, 246, 247, 253, 257, 272, 284
დისკრეტული და უწყვეტი ცვლადები 15
დისპერსია და სტანდარტული გადახრა
66, 70, 73, 78, 124

ე

ემპირიული წესი 75
ერთფაქტორიანი დისპერსიული
ანალიზი ცდის პირთა შორის
სქემისათვის (პიპოთეზები) 188
EXCEL-ის სტატისტიკური ფუნქციების
აღწერა 273

ვ

ვარიაციის კოეფიციენტი 75, 78, 81, 253

თ

თვისებრივი და რაოდენობრივი
მონაცემები 13
თვისებრივი მონაცემების სიხშირეთა
განაწილება 28

ი

ინტერვალების სკალა 14, 15, 20, 21, 94,
213, 271

კ

კითხვები 6, 11, 20, 36, 54, 65, 78, 101, 112,
124, 139, 166, 182, 201, 230, 258
კომპიუტერის როლი სტატისტიკაში 10
კორელაციის კოეფიციენტი 87, 88, 89, 90,
91, 92, 93, 94, 95, 96, 99, 100, 101, 102,
103, 104, 105, 239, 240, 243, 244, 247,
251, 252, 256, 258, 259, 261, 271, 283,
284, 325, 326
კორელაციური მატრიცა 251, 283
კოვარიაციის კოეფიციენტი 88
კრიტერიუმის მნიშვნელოვნების დონე
147

კრიტიკურობის სტატისტიკის
მნიშვნელობის გამოთვლა 145, 146
კრუსკალ-უოლისის კრიტიკურობი 226,
228, 230, 236
კვარტილთშორისი გაბნევის დიაპაზონი
63, 78, 264

ბ

ბან-უიტნის კრიტიკურობი 222, 224
მარტივი შემთხვევითი შერჩევა 16, 18, 20
მედიანის შეფასება 50, 57
მედიანის თვისებები 47
მისადაგების წირი 86, 97
მოცემული ალბათობით Z სიდიდის
პოვნა, 129
მოდალური ინტერვალი, 5, 52
მოდელის ვარგისიანობის შემოწმება 246
მრავლობითი რეგრესია და კორელაცია
249

ვ

ნამრავლის ალბათობის ფორმულა 111
ნდობის ინტერვალი დისპერსიისათვის
151
ნდობის ინტერვალი პოპულაციის
საშუალოსთვის 143
ნორმალური განაწილების თვისებები 128
ნორმალური მიახლოება 123

ო

ორამოკრეფიანი ამოცანები ორი
ალბათობის შედარებისას 179
ორი დამოუკიდებელი პოპულაციის
საშუალოთა შორის განსხვავების
ჰიპოთეზათა შემოწმება (ორივე, ან
ერთ-ერთი შერჩევის მოცულობა
ნაკლებია 30-ზე) 174

პ

პარამეტრები 119, 136, 155, 165, 180, 228
პარამეტრები და სტატისტიკა 9
პირობითი ალბათობა 111
პირველი და მეორე გვარის შეცდომები
სტატისტიკური კრიტიკურობების
გამოყენებისას 149

პოპულაცია და შერჩევა 8
პოპულაციის და შერჩევის განაწილება
34
პოპულაციის განსაზღვრა 9
პუასონის მიახლოება 122

რ

რა არის ალბათობა? 106
რამდენად დიდი უნდა იყოს შერჩევის
მოცულობა n , რომ სრულდებოდეს
ცხო? 133
რანგები და პროცენტული რანგები 63
რაოდენობრივი მონაცემების სიხშირეთა
განაწილება და მისი გრაფიკული
წარმოდგენა 22
რატომ არის საჭირო სტატისტიკა 10
რეგრესიის წრფე 98, 99, 100, 101, 102, 105,
247, 283
რეგრესიული ანალიზი 6, 86, 97, 248, 249,
283
რეგრესიული ანალიზი მრუდწირული
(არანრფივი) კავშირისთვის 248
რიგის სკალა 14, 15, 20, 21, 91, 94, 101,
104, 169, 214
რიგის სკალის მონაცემთა
რაოდენობრივი ბუნება 14

ს

საიმედოობა 19, 158, 168, 169
საპროგნოზო ინტერვალები 245
სასრული პოპულაციები 156
საშუალოს დამრგვალების წესი, 41
საშუალოს შეფასება 43, 191
საშუალოს თვისებები 44
საშუალოთა შორის განსხვავების
ჰიპოთეზის შემოწმება დიდი
მოცულობის მქონე შერჩევებისას 172
სიხშირეთა განაწილება 22, 23, 24, 28, 34,
36, 37, 38
სიხშირეთა განაწილების გრაფიკულად
გამოსახვა 25
სიხშირეთა პოლიგონი 27, 28
სპირმენის რანგობრივი კორელაციის
კოეფიციენტი 94
სრული ალბათობის ფორმულით 111

გამოყენებითი სტატისტიკის საფუძვლები

სტანდარტული გადახრა 40, 70, 71, 74, 75, 78, 79, 80, 81, 82, 92, 99, 104, 122, 123, 125, 127, 133, 138, 140, 141, 142, 144, 151, 152, 153, 157, 159, 160, 162, 163, 166, 167, 168, 169, 170, 171, 175, 183, 184, 185, 189, 242, 246, 251, 265, 269, 276

სტანდარტული ნორმალური განაწილების ცხრილის გამოყენება 128

სტანდარტული ნორმალური განაწილების წირის ქვეშ მდებარე არის ფართობის მოსაძებნი პროცენტული ცხრილი 131

სტატისტიკური ცხრილები, 6, 285

სტატისტიკური მეთოდები და გაზომვის ტიპი, 15

სტატისტიკური მეთოდების დანიშნულება 7

სტატისტიკური მეთოდოლოგია 7

სვეტოვანი და წრიული დიაგრამები 28

ტ

ტესტი პროპორციათა სხვაობისთვის 181

უ

უალბათესი რიცხვი 120, 123

უარყოფის არის დადგენა 147

უილკოქსონის ნიშნის რანგების კრიტერიუმი 225, 230

უმცირეს კვადრატთა მეთოდი 86, 98, 100, 240, 245, 247, 250

ფ

ფარდობითი სიხშირეები 28, 30, 38, 123, 126

ფიქტიური ცვლადები რეგრესიულ ანალიზში 254

ფოთლებიანი ღეროების მსგავსი დიაგრამები 33

შ

შანსების შეფარდება 221, 222

შანსების შეფარდების თვისებები 2×2 შეუღლების ცხრილებისათვის 122

შეჯამება, 5, 6, 10, 19, 22, 36, 77, 100, 112, 123, 138, 162, 200, 230, 255

შემთხვევითი სიდიდეები და მათი განაწილებები 110

შერჩევის ფარდობითი სიხშირის საშუალო და დისპერსია 123

შერჩევის სახეობები 15

შერჩევის შეცდომა 16, 20

შერჩევითი განაწილებები 133

შესწორებულ შერჩევით დისპერსია 67

შესწორებული სტანდარტული გადახრა 70, 152, 162

ჩ

ჩებიშევის თეორემა 76

ც

ცდისპირთა შიდა ორფაქტორიანი სქემა დადისპერსიული ანალიზი 193

ცდისპირთა შორის ორფაქტორიანი სქემა და დისპერსიული ანალიზი 196

ცდომილება 19, 157, 158, 169, 170, 241

ცდომილება, სიზუსტე და შერჩევის მოცულობა 157

ცენტრალური ზღვარიით თეორემა (ცხთ) 132

ცენტრალური ზღვარიით თეორემა და სიხშირეები 136

ცვლადები და მათი გაზომვა, 5, 13

ცვლადი 13, 15, 20, 37, 98, 102, 169, 214, 218, 219, 237, 238, 240, 241, 242, 249, 251, 252, 254, 255, 262, 263, 271

ნ

ნინასწარმეტყველება (პროგნოზირება) 100

ჟ

ჯამის ალბათობის ფორმულა 107

ხ

ხი-კვადრატ (χ^2) თანხმობის კრიტერიუმი 211

პ

პიპოთეზათა შემონმება 2×2
 შეუღლების ცხრილის შემთხვევაში
 216
 პიპოთეზათა შემონმება $\mu_1 - \mu_2$ -ის
 შესახებ 173
 პიპოთეზათა შემონმება (პოპულაცია
 ნორმალურია, $n < 30$) 153
 პიპოთეზათა შემონმება დანყვილებულ
 მონაცემთა საშუალოების
 სხვაობებისათვის 178
 პიპოთეზათა შემონმება
 დისპერსიისათვის 160
 პიპოთეზათა შემონმება ნორმალური
 პოპულაციის საშუალოსათვის ($n < 30$)
 150

პიპოთეზათა შემონმება ორი
 დამოუკიდებელი ნორმალური
 პოპულაციის დისპერსიების შესახებ
 177
 პიპოთეზათა შემონმება პოპულაციის
 პროპორციისათვის ($n \geq 30$) 154
 პიპოთეზათა შემონმება პოპულაციის
 პროპორციისათვის, როცა შერჩევის
 მოცულობა პატარაა 155
 პიპოთეზათა შემონმება საშუალოსთვის
 145
 პიპოთეზების ფორმულირება 145
 პიპოთეზის შემონმება დახრილობის
 შესახებ 241
 პიპოთეზის შემონმება თანაკვეთის
 შესახებ 242
 პისტოგრამები 25

გამომცემლობის რედაქტორი
გარეკანის დიზაინი
კომპ. უზრუნველყოფა

მარინე ვარაშაშვილი
ნინო ებრალიძე
ლალი კურდღელაშვილი

0179 თბილისი, ი. ჭავჭავაძის გამზირი 14
14 Ilia Chavchavadze Avenue, Tbilisi 0179
Tel 995(32) 225 14 32
www.press.lsu.ge