



საქართველოს ტექნიკური უნივერსიტეტი

სტუ

ყოფნა ძეგლი...

საბაზისო ინჟინერინგი მონაცემთა უმჯობესება და ანალიზი

სტეფან ბ. ვარდემანი

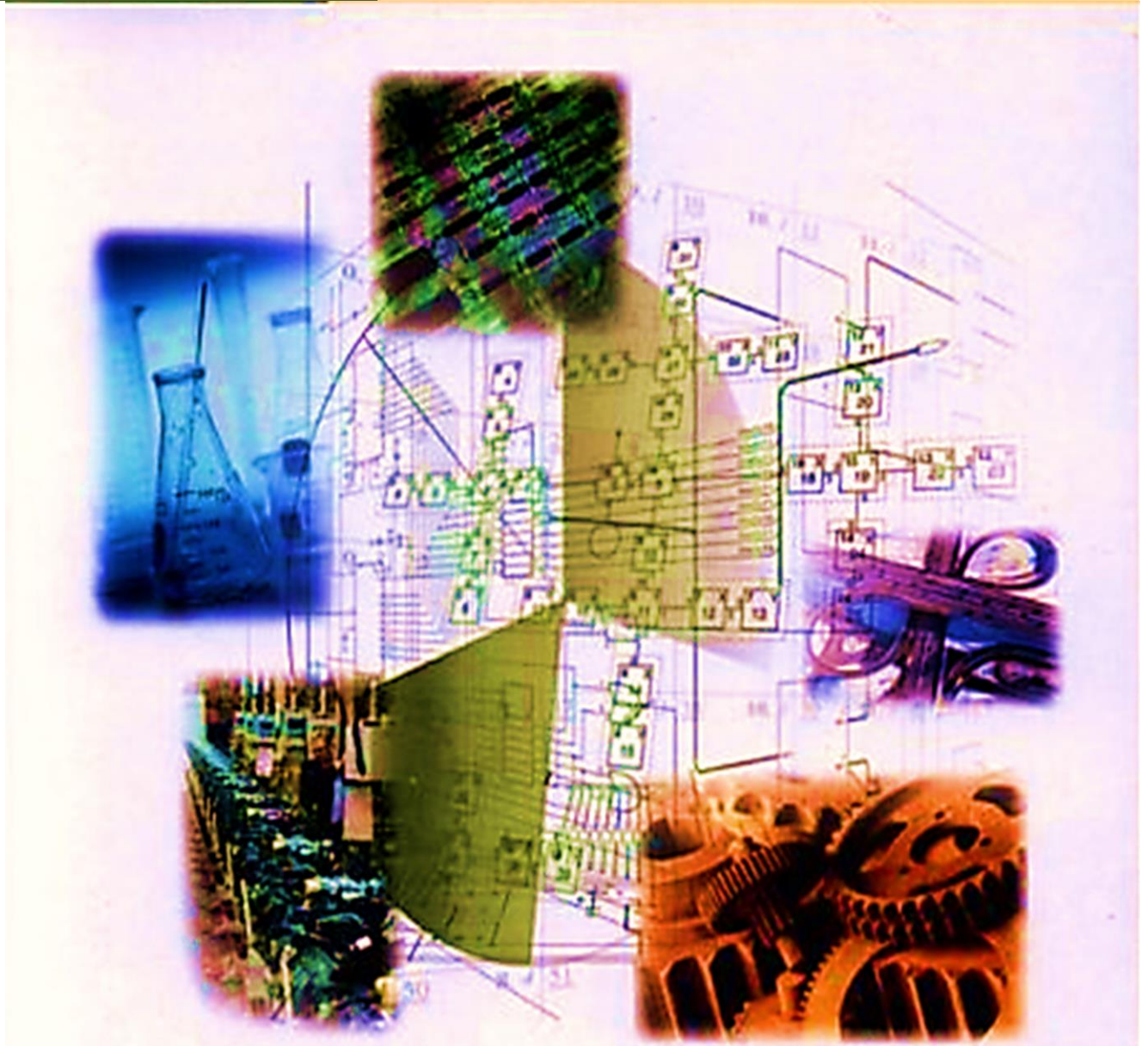
აიოვას სახელმწიფო უნივერსიტეტი

ჯ. მარკუს ჯოუბი

მაიამის უნივერსიტეტი

საბაზისო
ინჟინერინგი

მონაცემთა შეგროვება და ანალიზი



სტეფან ბ. ვარდემანი ჯ. მარკუს ჯოუბი



საქართველოს ტექნიკური უნივერსიტეტი

სტუ

ყოფნა ძეგლი...

საბაზისო ინჟინერინგი

მონაცემთა შეგროვება და ანალიზი

ლექციების კურსი

Stephen B. Vardeman, J. Marcus Jobe

“Basic Engineering. Data Collection and Analysis”

მიხედვით

(სტუ ბიბლიოთეკა)

საინჟინრო სპეციალობის სტუდენტებისთვის

დ. ნატროშვილი თ. ალიაშვილი

საბაზისო ინჟინერინგი მონაცემთა შეგროვება და ანალიზი

ლექციების კურსი მომზადებულია

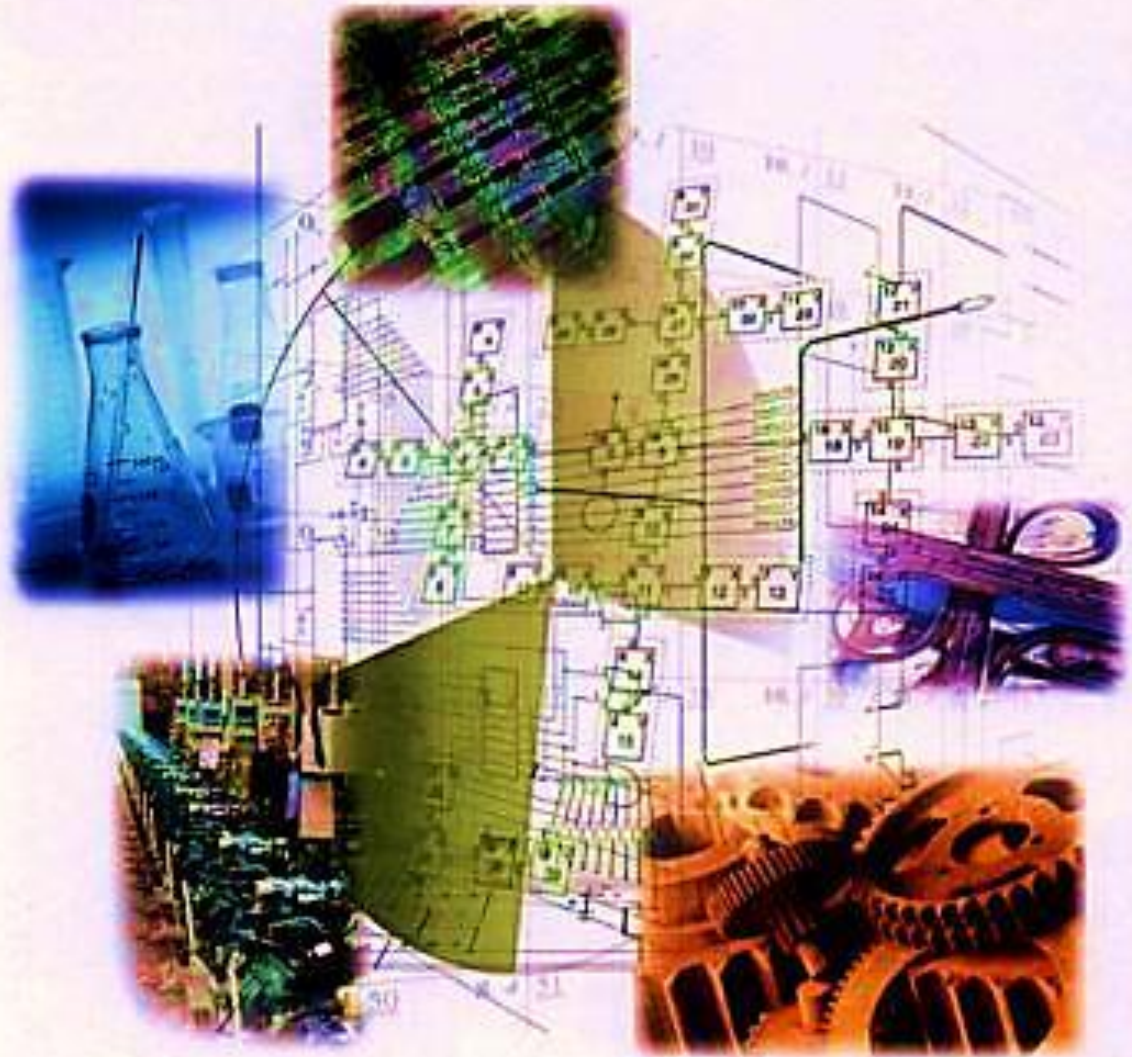
Stephen B. Vardeman, J. Marcus Jobe

“Basic Engineering. Data Collection and Analysis”

მიხედვით

**Basic
Engineering**

Data Collection and Analysis



Stephen B. Vardeman

J. Marcus Jobe

სარჩევი

თავი 1 შესავალი

- 1.1 საინჟინრო სტატისტიკა: რა და რატომ
- 1.2 საბაზისო ტერმინოლოგია

სავარჯიშოები

თავი 2 ელემენტარული აღწერითი სტატისტიკა

- 2.1. რაოდენობრივი მონაცემების ელემენტარული გრაფიკული და ცხრილური დამუშავება
- 2.2. სიხშირეთა ცხრილები და ჰისტოგრამები
- 2.3. კვანტილები და მათთან დაკავშირებული გრაფიკები
- 2.4. სტანდარტული რიცხვითი მახასიათებლები
- 2.5 დამოკიდებულებები ცვლადებს შორის

სავარჯიშოები

თავი 3 დისკრეტული შემთხვევითი სიდიდე

- 3.1. შემთხვევითი სიდიდეები და მათი განაწილებები
- 3.2. დისკრეტული ალბათობის ფუნქცია და დაგროვებითი ალბათობის ფუნქცია
- 3.3. დისკრეტული ალბათობის განაწილების შეჯამება
- 3.4. ბინომური და გეომეტრიული განაწილებები
- 3.5. პუასონის განაწილება.

სავარჯიშოები

თავი 4 უწყვეტი შემთხვევითი სიდიდე

- 4.1. ალბათობის სიმკვრივის ფუნქციები და დაგროვებითი ალბათობის ფუნქციები
- 4.2. უწყვეტი განაწილების საშუალო და დისპერსია
- 4.3. ალბათობის ნორმალური განაწილება
- 4.4. მაჩვენებლიანი განაწილება
- 4.5. ვეიბულის განაწილება

სავარჯიშოები

თავი 5 ერთობლივი განაწილებები და დამოუკიდებლობა

- 5.1. დისკრეტულ შემთხვევით სიდიდეთა ერთობლობა
 - 5.2. პირობითი ალბათობა და დამოუკიდებლობა დისკრეტული შემთხვევითი სიდიდეებისთვის
 - 5.3. მრავალი ცვლადის ფუნქციები
 - 5.4. საშუალო და დისპერსია შემთხვევით სიდიდეთა წრფივი კომბინაციებისთვის
 - 5.5 ცენტრალური ზღვარითი ეფექტი
- სავარჯიშოები

თავი 6 ფორმალური სტატისტიკური დასკვნები

- 6.1. დიდი შერჩევის ნდობის ინტერვალი საშუალოსთვის
 - 6.2. ზოგადად გამოყენებადი ნდობის ინტერვალი μ -სთვის
 - 6.3. რამდენიმე დამატებითი კომენტარი ნდობის ინტერვალთან დაკავშირებით
 - 6.4. დიდი შერჩევის მნიშვნელობის ტესტი საშუალოსთვის
- სავარჯიშოები

დანართი

- A. 1. ელემენტარული ალბათობის ძირითადი ცნებები და აქსიომები

- A. 2. უმარტივესი თეორემები ალბათობათა თეორიიდან.
- A. 3. პირობითი ალბათობა და ხდომილებათა დამოუკიდებლობა
- A. 4. კომბინატორიკის ელემენტები

სავარჯიშოები

- B. ცხრილები

1. შესავალი

1.1. საინჟინრო სტატისტიკა: რა და რატომ

ზოგადად დავსვით კითხვა: რა არის ინჟინრის სამუშაო? გააკეთოს პროექტი, აშენოს, იმოქმედოს და/თუ გააუმჯობესოს ფიზიკური სისტემები და აწარმოოს პროდუქცია. ყოველივე ამ საქმიანობას მართავენ იმ მათემატიკური და ფიზიკური თეორიების ცოდნის საფუძველზე, რომელთა შესწავლაც ხდება საინჟინრო საბაკალავრიატო სასწავლო გეგმის შესაბამისად. როცა იზრდება საინჟინრო გამოცდილება და მიიღება რაიმე ინჟინრული გადაწყვეტილება, მაშინვე მასთან ერთად, იგივე ხმაში იწყებს მუშაობას რაოდენობრივი და სამეცნიერო პრინციპები. ამავე დროს ტექნოლოგიური მიღწევები სულ უფრო ვითარდება და გვხვდება ახალი სისტემები და პროდუქციები. ყოველივე ეს კი ინჟინერს და მის მუშაობას აყენებს პირისპირ ახალ გამოწვევებთან, სადაც უკვე თეორიები და გამოცდილებები გვთავაზობენ მცირეოდენ დახმარებას. როდის მოხდება ეს და რა არის გასაკეთებელი?

რასაკვირველია, ამ შემთხვევაში შესაძლოა დავუძახოთ კონსულტანტებს, მაგრამ მაინც უმეტეს შემთხვევაში ინჟინერს დამოუკიდებლად უწევს გამოსავლის პოვნა, ანუ გადაწყვეტი ნაბიჯის გადადგმა. ამისათვის აუცილებელია **მონაცემთა შეგროვება და ანალიზი**. იგი დაგვეხმარება უკეთ გავიგოთ, როგორ მუშაობენ ახალი სისტემები. მონაცემთა შეგროვებასა და ანალიზზე სპეციალური ტრენინგებისა გარეშე ინჟინრის მცდელობა შეიძლება იყოს შემთხვევითი და უხარისხოდ ჩაფიქრებული. ამას ეწირება ძვირფასი დრო და რესურსები. შედეგად კი მიიღება ზოგჯერ მცდარი და ზოგჯერ სულ მცირე არააუცილებელი ან ორაზროვანი დასკვნები. ამის თავიდან ასაცილებლად მომუშავე ინჟინრისთვის სასიცოცხლოდ აუცილებელია, რომ ხელთ ჰქონდეს ინსტრუმენტი (იარაღი), რომელიც მოიცავს საუკეთესო შესაძლო პრინციპებსა და მეთოდებს მონაცემთა შეგროვებისა და ინტერპრეტაციისთვის.

საინჟინრო სტატისტიკის მიზანია, საჭირო კონცეფციებითა და მეთოდებით უზრუნველყოს ის ინჟინრები, რომლებიც პირისპირ აღმოჩნდნენ მოვლენებთან და რომელთა ცოდნის დონე ვერ უწევს სრულყოფილ ადექვატურ დახმარებას. იგი

წარმოგვიდგენს პრინციპებს, თუ როგორ უნდა შეძლონ მათ წარმატებით შეისწავლონ ემპირიული სახის ინფორმაცია და მართონ საინჟინრო ტიპის სისტემები.

საინჟინრო სტატისტიკა არის კვლევა (შესწავლა) იმისა, თუ როგორ არის უკეთესი

1. შევადგოთ საინჟინრო მონაცემები,
2. შევაჯამოთ და აღვწეროთ საინჟინრო მონაცემები, და
3. გავაკეთოთ ფორმალური დასკვნა და პრაქტიკული გათვლები საინჟინრო მონაცემთა ბაზის შესახებ,

სანამ მივიღებთ გადაწყვეტილებას რაიმეს ცვლილების შესახებ.

1.2. საბაზისო ტერმინოლოგია

საინჟინრო სტატისტიკა მოითხოვს შევისწავლოთ ახალი ცნებები და ნაცნობი სიტყვების ტექნიკური მნიშვნელობები. ჩვენ შემოვიტანთ სტატისტიკისთვის დამახასიათებელ ტერმინებსაც, რომლებიც შეიძლება გამოჩნდნენ მონაცემთა შესწავლისას ან მათ სტრუქტურებში.

როდესაც ინჟინერი აგროვებს მონაცემებს, მან უნდა გადაწყვიტოს როგორი სახის აქტიურობით უნდა მოხდეს ეს. გადაატრიალოს ჩამრთვლები და მართოს პროცესში მონაწილე ცვლადები, თუ მარტივად, დაელოდოს მოვლენის მოხდენას და შეეცადოს ჩაიწეროს მისი დამახასიათებელი თვისებები?

დაკვირვებითი კვლევა (შესწავლა) არის ისეთი ფორმა, როდესაც მკვლევარის როლი ძირითადად პასიურია. პროცესს ან მოვლენას უყურებს და იწერს მონაცემებს, მაგრამ პიროვნების მხრიდან არ არის არანაირი ჩარევა და მონაწილეობის მიღება კვლევის პროცესში.

ექსპერიმენტული კვლევა (ან უფრო მარტივად, **ექსპერიმენტი**) არის ისეთი ფორმა, როდესაც მკვლევარი აქტიურადაა ჩართული პროცესის მიმდინარეობაში. საპროცესო ცვლადები არის მართვადი და კვლევის გარემო რეგულირებადია.

რეალურად, სტატისტიკური კვლევების უმრავლესობა არის ერთროულად ორივე ტიპის, და ორივე განმარტება უნდა იყოს გააზრებული, როგორც ინტერვალის საპირისპირო მეორე ბოლო. ამ კონტინუუმში ექსპერიმენტული მეთოდი საზოგადოდ წარმოგვიდგენს ყველაზე უფრო ეფექტურ და საიმედო გზას საინჟინრო

მონაცემების შესაგროვებლად. როგორც წესი, ასე გაცილებით სწრაფად ხდება საპროცესო ცვლადების მართვა და იმის დანახვა, თუ როგორ პასუხობს სისტემა ცვლილებებს, ვიდრე პასიური დაკვირვება იმ იმედით, რომ იქნებ დავინახოთ რამე საინტერესო გამოვლინება.

გარდა ამისა, გაცილებით უსაფრთხოა დავადგინოთ მიზეზ-შედეგობრივი კავშირები ექსპერიმენტის დროს, ვიდრე ეს შეიძლება გავაკეთოთ დაკვირვებების დროს. ცხოვრებისეული სისტემები მრავალმხრივი, ანუ კომპლექსური ხასიათის არიან. შესაძლებელია დავაკვირდეთ რამდენიმე კარგად ჩატარებულ პროცესს და დავასკვნათ, რომ ყველა ეს კონკრეტული შემთხვევა აღებული იყო X -ის მიდამოდან, ამიტომ ჩავთვალოთ, რომ X-ის მიდამოში პროცესი კარგად წარიმართება, რაც ცხადია არ არის უსაფრთხო. აქ შესაძლებელია მოხდეს, რომ მნიშვნელოვანი ცვლადები იყოს მხოლოდ ფონი. ისინი ცვალებადობენ და წარმოადგენენ დასაკვირვებელი სისტემის ყოფაქცევის ჭეშმარიტ მიზეზს. ესენი არიან ე. წ. **დაფარული ცვლადები**, რომლებიც მართავდნენ ან არეგულირებდნენ ორივე პროცესს. შესაძლებელია მოხდეს ისიც, რომ გვექონდეს უფრო მარტივი მრავალი ცვლადის შემთხვევა, სადაც ისინი ცვალებადობენ შემთხვევითად, ისე რომ მათი ზემოქმედება სისტემაზე უმნიშვნელოა, მითუმეტეს რომ ასეთ შემთხვევაში მათზე დაკვირვების დრო შეზღუდულია. ამ დროში შესაძლებელი იყო სხვა უკეთესი შედეგის მიღებაც.

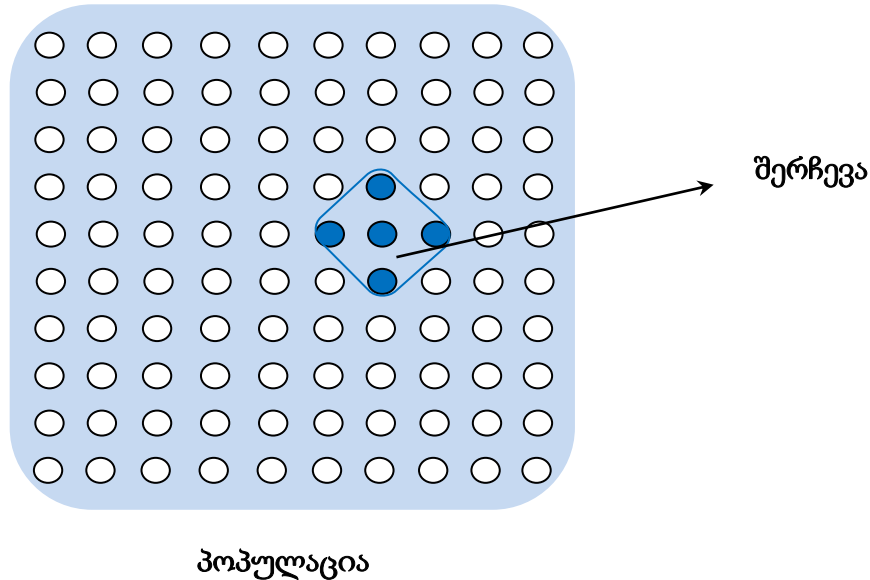
მეორეს მხრივ, ექსპერიმენტის დროს, როდესაც დასაკვირვებელი გარემო ფართოა, მაგრამ რეგულირებადი, გარდა მცირე რაოდენობა ცვლადებისა ინჟინერი ცვლის მათ მიზანმიმართულად და ამის შედეგად დასკვნები მიზეზების შესახებ გაცილებით უფრო ძლიერია. თუ გარემოებები შექმნილია მკვლვარის მიერ, მაშინ მას მუდმივად თან ახლავს ხელსაყრელი შედეგები. ამ დროს უეჭველად უნდა ვიყოთ დარწმუნებული არჩეული მეთოდის გონივრულობაში, რომლის გამოც მივიღეთ ხელსაყრელი შედეგი.

***განმარტება 1.1. პოპულაცია** არის საგნების მთლიანი ჯგუფი, რომლის შესახებაც გვინდა ინფორმაციის შეგროვება სტატისტიკური კვლევისთვის. (ქართულ ტერმინოლოგიაში მას გენერალურ ერთობლიობასაც უწოდებენ).*

***განმარტება 1.2. შერჩევა** (ან ამორჩევა) ეწოდება საგნების იმ ჯგუფს, რომელზედაც რეალურად ხდება მონაცემთა შეგროვება. იმ შემთხვევაში, თუ მიმდინარეობს აღწერითი კვლევა (ანუ როდესაც მონაცემები გროვდება ზოგიერთ საგანზე ან მთლიან პოპულაციაზე და დასკვნები განკუთვნილია მხოლოდ ამ საგნებისთვის),*

შერჩევა არის პოპულაციის საგანი. (ზოგიერთ შემთხვევაში შესაძლებელია ჩავრთოთ მთელი პოპულაცია).

შემდეგი ნახაზი გვიჩვენებს პოპულაციასა და შერჩევას შორის დამოკიდებულებას.



ნახაზი 1.1. პოპულაცია და შერჩევა

დავუშვათ, რომ გადმოსატვირთ ტერმინალში ყუთით მიღებულია მანქანის 100 ნაწილი. ხდება 5 მათგანის შემოწმება იმაზე, მისაღებია თუ არა ეს ტვირთი. ასი ნაწილი შეადგენს ინტერესის პოპულაციას და ხუთი ნაწილი (ერთხელ აღებული) არის შერჩევა, რომლის ზომა არის 5. (შევნიშნოთ, აქ რომ გამოყენებული სიტყვები ნიშნავს შემდეგს: გვაქვს ერთი შერჩევა და არა ხუთი.)

არსებობს შემდეგი სიტყვების, „პოპულაცია“ და „შერჩევა“ მნიშვნელობების გაფართოვების რამდენიმე გზა, რომლებიც ხშირად გამოიყენება სხვადასხვა დროს. ერთ-ერთი ასეთი გაფართოვება არის ორივესთვის საერთო, ისინი გამოიყენებიან არა მხოლოდ შესასწავლი საგნების შედარებისთვის, არამედ იმ მონაცემთა მნიშვნელობების შესაფასებლადაც, რომლებიც ასოცირდებიან მათთან. ამის საილუსტრაციოდ შეიძლება განვიხილოთ ერთი ასეთი მაგალითი. დავუშვათ, გვინდა შევამოწმოთ ყუთში მოთავსებული მანქანის ნაწილები სიმტკიცე. მათთვის გვექნება სიმტკიცის 100 სხვადასხვა მნიშვნელობა. ჩავთვალოთ, რომ რიცხვების ეს ერთობლიობა არის პოპულაცია. ხუთი სიმტკიცის მნიშვნელობა კი, რომლებიც

შეესაბამებთან შერჩევის ნაწილებს, შეიძლება განვსაზღვროთ როგორც შერჩევა რიცხვების ამ პოპულაციიდან.

მონაცემთა ტიპები: ინჟინრებს მუშაობის კვლევების პროცესში შეხვედრა უწევთ მრავალი ტიპის მონაცემებთან. ერთიერთი მნიშვნელოვანი მომენტი, რაც აინტერესებთ ინჟინრებს, არის მონაცემების განსხვავება რაიმე ნორმისგან. ეს განსხვავება უფრო ხშირად ეხება ხოლმე ხარისხს, რომლის მნიშვნელობებიც შეიძლება იყოს კატეგორიული ან რაოდენობრივი მახასიათებლები. საინჟინრო საქმეში ძალიან ხშირად კატეგორიული ცვლადების მაგიერაც ძირითად რიცხვითი მონაცემები გამოიყენება, რაც გამოიხატება იმაში, რომ სტანდარტებიდან გადახრებს რიცხვები უფრო კარგად გამოსახვენ, ვიდრე კატეგორიული ცვლადები.

განმარტება 1.3. თვისებრივი ანუ კატეგორიული მონაცემები არის მნიშვნელობები, ძირითადად არარიცხვითი ხასიათის, და ასოცირდება შერჩევიდან აღებული საგნის რაიმე მახასიათებელთან. აქ შეიძლება შეგვხვდეს დალაგება ასოცირებული თვისებრივ მონაცემებთან, მაგრამ მათი შეერთება და დათვლა მაინც მოითხოვდეს რაიმე აზრით ამ მონაცემების რიცხვითი მნიშვნელობებით წარმოდგენას.

ისევ განვიხილოთ მანქანის 5 ნაწილისგან შემდგარი შერჩევა ყუთში არსებული 100 ნაწილიდან თუ შესაძლებელია თითოეული ნაწილის კლასიფიცირება რაიმე (დალაგებული) კატეგორიების მიხედვით; მაგალითად,

- (1) - დადასტურებული (რომ კარგია),
- (2) - აღდგენილი,
- (3) - ჯართი.

დავუშვათ, რომ ცნობილია ხუთივე ნაწილის კლასიფიკაცია, ანუ გვაქვს 5 თვისებრივი მონაცემი. თუ მთელი ერთობლიობიდან (იგულისხმება 5 დეტალი შერჩევიდან) 3 - დადასტურებულია, 1 - აღდგენილი და 1 - ჯართი. ამგვარად რიცხვითი შედეგი მიიღება საწყისი კატეგორიული მონაცემებიდან.

კატეგორიული მონაცემების საპირისპიროდ არსებობს რიცხვითი მონაცემებიც.

განმარტება 1.4. რაოდენობრივი ანუ რიცხვითი მონაცემები არის რიცხვითი მახასიათებლების მნიშვნელობა, რომელიც ასოცირდება შერჩევიდან აღებულ საგნებთან. ზოგადად, ესენი არიან იმ მოვლენის მოხდენათა რაოდენობა, რომელიც

გვანტერესებს, ან საგანთა ფიზიკური თვისების *გაზომვის* შედეგად მიღებული რიცხვითი მნიშვნელობა.

კვლავ დავუბრუნდეთ ჩვენს ყუთს მანქანის ნაწილებით. როკველის სიმტკიცის მნიშვნელობა 5 შერჩეული ნაწილისთვის შეიძლება წარმოვადგინოთ როგორც *გაზომვების* რაოდენობრივი მნიშვნელობების მონაცემების სიმრავლე. ხოლო ხილული დაზიანებების რაოდენობა მანქანის ზედაპირზე 5 შერჩეული ნაწილისთვის – როგორც სიმრავლე *რაოდენობრივი* მონაცემების.

ზოგჯერ მოსახერხებელია ვიმოქმედოთ ისე, თითქოს შესაძლოა გაზომვათა უსასრულო სიზუსტე. ამ თვალსაზრისით გაზომვის შედეგად მიღებული ცვლადები არიან უწყვეტნი იმ აზრით, რომ მათი შესაძლო მნიშვნელობების სიმრავლე არის მთლიანი (*უწყვეტი*) რიცხვითი ინტერვალი. მაგალითად, იდეალიზაცია მოსახერხებელია როკველის სიმტკიცის შემთხვევაში, რადგან სიმტკიცის მნიშვნელობები მდებარეობენ $(0, +\infty)$ შუალედში. ცხადია, ეს მართლაც მხოლოდ იდეალიზებაა. ყველა რეალური გაზომვის შედეგი არის უახლოესი ერთეული (არ აქვს მნიშვნელობა, რომელი იქნება). ამას მივყავართ იქამდე, რომ საჭიროა მაღალი დონის აჭურვილობის საზომი ხელსაწყო ციფრული ეკრანით. აქედან გამომდინარე, როდესაც ვუყურებთ გაზომვის რიცხვით მონაცემებს, ისინი არიან *დისკრეტული* სიდიდეები, თითქოს ისინი არიან იზოლირებული შესაძლო მნიშვნელობები, მაშინ როცა კონტინუუმ რაოდენობის შედეგია მოსალოდნელი. მთელი $(0, +\infty)$ შუალედის აღება მათემატიკურად მოსახერხებელია და სრულიად ადექვატურია პრაქტიკული მიზნებისთვის.

საყოველთაოდ ცნობილია, რომ გაზომვის მონაცემები უკეთესია, ვიდრე კატეგორიული ან რაოდენობრივი (დათვლილი) მონაცემები. სტატისტიკური მეთოდები გაზომვითი მონაცემებისთვის უფრო მარტივი და მეტი ინფორმაციის შემცველია, ვიდრე რაოდენობრივი მონაცემებისთვის. ამიტომ უმეტესად შეისწავლიან გაზომვით მონაცემებს, ვიდრე ფიზიკური ობიექტების რაოდენობრივ მონაცემებს. თუმცა, ზოგჯერ ამას აბალანსებს საპირისპირო ფაქტი, რომ გაზომვები არის ძალიან შრომატევადი საქმიანობა (რაც ნიშნავს სიძვირეს), ვიდრე რაოდენობრივი მონაცემების შეგროვება.

განმარტება 1.5. ვიტყვი, რომ გვაქვს *ერთცვლადიანი მონაცემები*, თუ შერჩევის ყოველ ობიექტს ვაკვირდებით მხოლოდ ერთი მახასიათებლის მიხედვით.

განმარტება 1.6. ვიტყვი, რომ გვაქვს მრავალცვლადიანი მონაცემები, თუ შერჩევის ყოველ ობიექტს ვაკვირდებით ერთზე მეტი მახასიათებლის მიხედვით. კერძო შემთხვევაში, როდესაც დაკვირვება ხდება მხოლოდ ორი მახასიათებლის მიხედვით, მაშინ ვამბობთ რომ მოცემული გაქვს ბივარიანტული მონაცემები.

განმარტება 1.7. როდესაც მრავალცვლადიანი მონაცემები შედგება რამდენიმე განსაზღვრულობისგან ძირითადად ერთი და იმავე მახასიათებლების მიხედვით (მაგ., დაზადებულია სხვადასხვა ხელსაწყოთი, ან დამზადებულია სხვადასხვა დროს), მაშინ ასეთ მონაცემებს ვუწოდებთ განმეორებითი გაზომვის მონაცემებს. იმ კერძო შემთხვევაში, როცა გვაქვს ბივარიანტული პასუხები, მაშინ ვიყენებთ ტერმინს შეწყვილებული მონაცემები.

მნიშვნელოვანია ამოვიცნოთ მონაცემთა მრავალცვლადიანი მახასიათებელი, როდესაც ის სახეზეა. როკველის სიმტკიცის მნიშვნელობები შერჩევის 5 ელემენტისთვის 100 ნაწილიანი ყუთიდან და ნახშირბადის პროცენტული შემცველობა სხვა 5 ნაწილში არ არის ეკვივალენტური იმისა, რომ ერთდროულად მოგვეთხოვა ორივე (სიმტკიცე და ნახშირბადის შემცველობა) ერთი და იმავე 5 დეტალიანი შერჩევისთვის. ამ შემთხვევაში გვაქვს ორი შერჩევა: 5 ერთცვლადიანი მონაცემებისთვის პირველ შემთხვევაში და ერთი შერჩევა 5 ბივარიანტული მონაცემებისთვის მეორე შემთხვევაში. მეორე სიტუაცია უკეთესია ვიდრე პირველი, რადგან ის იძლევა ანალიზის და ექსპლუატაციის საშუალებას მათი ურთიერთდამოკიდებულების შესახებ, რომელიც შეიძლება არსებობდეს სიმტკიცესა და ნახშირბადის პროცენტულ შემცველობას შორის.

მონაცემთასტრუქტურების ტიპები: სტატისტიკური საინჟინრო კვლევები ზოგჯერ ტარდება პროცესის წარმადობის, არსებულ სტანდარტებთან შედარების მიზნით. ასეთი გამოკვლევები შეიცავენ მხოლოდ ერთ შერჩევას. მაგრამ არსებობს გაცილებით მეტი საერთო სიმრავლე პირობებისა, რომლის მიხედვითაც შესაძლებელია მათი შედარება. არსებობს სხვადასხვა სტანდარტული ცნებები სტრუქტურების და მეთოდების მრავალშერჩევანი კვლევებისთვის. მაგალითად, საილუსტრაციოდ განვიხილოთ ორი მათგანი.

განმარტება 1.8. (სრული) ფაქტორიალური კვლევა არის ისეთი კვლევა, როდესაც მონაცემები შეგროვილია ყველა საპროცესო ცვლადისა და პარამეტრის ყველა შესაძლო კომბინაციის მიხედვით. საპროცესო ცვლადებს უწოდებენ ფაქტორებს, ხოლო თითოეული ცვლადის პარამეტრებს ფაქტორის დონეებს.

მაგალითად, დავუშვათ, რომ გვაქვს ჩვენთვის საინტერესო ოთხი ფაქტორი. მოხერხებულობისთვის აღვნიშნო ისინი A, B, C და D ასოებით. თუ A-ს აქვს 3 დონე, B-ს აქვს 2 დონე, C-ს აქვს 2 დონე და D-ს აქვს 4 დონე. კვლევა, რომელიც მოიცავს ყველა შესაძლო შერჩევას, არის $3 \times 2 \times 2 \times 4 = 48$ პირობის სხვადასხვა შესაძლო სიმრავლე. ვუწოდოთ მას $3 \times 2 \times 2 \times 4$ ფაქტორიალური კვლევა.

მაგალითი 1.1. ვთქვათ, ტარდება ექსპერიმენტი მოწყობილობის მიერ გრანულების წარმოებაზე $2 \times 2 \times 2$ ფაქტორიალური სტრუქტურით. კვლევის ფაქტორები და შესაბამისი დონეებია:

მოცულობის მიხედვით: მცირე მოცულობა vs. მეტი მოცულობა

მასალის ნაკადი: არსებული მეთოდით vs. ხელით შევსება

ნარევის ტიპი: აუკინძავი vs. აკინძული

(vs. არის შემოკლებული აღნიშვნა სიტყვის, “versus” - წინააღმდეგ, პირისპირ, ოპოზიციაში).

თუ ამათ ყველას გავაერთიანებთ რვა სიმრავლედ პირობების მიხედვით, გვექნება ქვემოთ მოცემული სურათი.

მდგომარეობა	მოცულობა	ნაკადი	ნარევი
1	მცირე	არსებული	აუკინძავი
2	მეტი	არსებული	აუკინძავი
3	მცირე	ხელით	აუკინძავი
4	მეტი	ხელით	აუკინძავი
5	მცირე	არსებული	აკინძული
6	მეტი	არსებული	აკინძული
7	მცირე	ხელით	აკინძული
8	მეტი	ხელით	აკინძული

ცხრილი 1. 2. ²³ ფაქტორიალური კვლევის კომბინაციები

როდესაც გვაქვს ბევრი ფაქტორი და მოიცავს და/ან დონეებს, მაშინ შერჩევათა რაოდენობა ფაქტორიალურ კვლევაში სწრაფად იზრდება არაპრაქტიკულ

ზომებამდე. ინჟინრებს ხშირად სურთ მოაგროვონ საჭირო მონაცემები მხოლოდ იმ კომბინაციებისთვის, რომლითაც შემდგომ შესაძლებელი იქნება შედგეს სრული ფაქტორიალური კვლევა.

განმარტება 1.9. *ნაწილობრივ ფაქტორიალური კვლევა არის ისეთი კვლევა, როდესაც მონაცემები შეგროვილია მხოლოდ ზოგიერთი კომბინაციისთვის, რომელთაგანაც შესაძლებელი იქნება შედგეს სრული ფაქტორიალური კვლევა.*

არ ჩავთვალოთ, რომ ნაწილობრივ ფაქტორიალური კვლევა მოგვცემს იმავე შედეგს, რასაც შესაბამისი სრული ფაქტორიალური კვლევა. ამ დროს შესაძლებელია, რომ რომელიმე ინფორმაცია დაგვეკარგოს, რადგან შევისწავლით ყველა შესაძლო კომბინაციების მხოლოდ ნაწილს. თუმცა, ზოგჯერ შესაძლებელია ისიც, რომ ზოგიერთი ნაწილობრივ ფაქტორიალური კვლევა პოტენციურად იყოს მეტი ინფორმაციის მომცველი, ვიდრე სხვა რომელიმე. თუ შერჩევის ზომა ფიქსირებულია, გარკვევა იმისა, რომელი შერჩევა ავიღოთ, წარმოადგენს საკითხს, რომელიც მოითხოვს ფრთხილ განხილვას. არსებობს მეთოდები, რომელთა გათვალისწინება საშუალებას იძლევა, გავარკვიოთ, სრული ფაქტორიალური კვლევის რომელი ნაწილი უნდა ავიღოთ, რომ იგი იყოს პოტენციური ინფორმაციის მომცველი.

მაგალითი 1.1. *(გაგრძელება).* რეალურად განხორციელებული ექსპერიმენტის პროცესის კომბინაციები წარმოდგენილია ზემოთ მოყვანილ ცხრილში. შემდეგ ცხრილში კი ჩამოთვლილია ოთხი სხვადასხვა ექსპერიმენტული კომბინაცია, რომელიც წარმოადგენს კარგად ამორჩეულ ნახევარს 8 შესაძლო კომბინაციიდან. (ესენი არიან კომბინაციები წინა ცხრილიდან, რომელთა ნომრებია 2, 3, 5 და 8).

მოცულობა	ნაკადი	ნარევი
მეტი	არსებული	აუკინძავი
მცირე	ხელით	აუკინძავი
მცირე	არსებული	აკინძული
მეტი	ხელით	აკინძული

ცხრილი 1. 3. *2³ ფაქტორიალური კვლევის ნახევარი.*

სავარჯიშოები:

1. აღწერეთ სიტუაცია, როდესაც მნიშვნელოვანია დაკვირვებითი კვლევა. აღწერეთ სხვა სიტუაცია, როდესაც გამოიყენება ექსპერიმენტული კვლევა.
2. აღწერეთ ორი განსხვავებული გარემოება, სადაც საჭირო იქნება შესაბამისად თვისებრივი და რაოდენობრივი მონაცემები.
3. რა სახის ინფორმაცია შეიძლება მივიღოთ ერთი შერჩევის n ბივარიანტული მონაცემებიდან (x, y) წერტილისთვის, რომლის მიღებაც არ შეიძლება ორი განცალკევებული შერჩევიდან შესაბამისად n მონაცემებიდან x წერტილისთვის და n მონაცემებიდან y წერტილისთვის.
4. აღწერეთ სიტუაცია, როდესაც საჭირო ხდება შეწყვილებული მონაცემების გამოყენება.
5. განვიხილოთ კვლევა, ქაღალდის თვითმფრინავების გაკეთება, სადაც გვაქვს ორი დიზაინი (ვთქვათ, დელტა vs. ფრთა), ორი განსხვავებული ქაღალდი (ვთქვათ, კონსტრუქციული vs. საბეჭდი) და ორი განსხვავებული დატვირთულობის პირობა (ვთქვათ, ქაღალდის სამაგრი კლიპებით vs. სამაგრი კლიპების გარეშე). გვანტერესებს მათი მოქმედების ეფექტი ფრენის დისტანციაზე. აღწერეთ სრული ფაქტორიალური და ნაწილობრივ ფაქტორიალური მონცემთა სტრუქტურები, რომლებიც შეიძლება გამოჩნდნენ ასეთი კვლევის დროს.
6. ახსენით, რატომ არის უფრო საიმედო გავაკეთოთ მიზეზობრივი დასკვნა ექსპერიმენტული კვლევიდან ვიდრე დაკვირვებითი კვლევიდან.

2. ელემენტარული აღწერითი სტატისტიკა

2.1. რაოდენობრივი მონაცემების ელემენტარული გრაფიკული და ცხრილური დამუშავება

საინჟინრო მონაცემები ყოველთვის ცვლადი სიდიდეებია. თითქმის ყოველთვის მონაცემთა ანალიზი იწყება შესაბამისი გრაფიკულ და/ან ცხრილურ გამოსახულებებზე დაკვირვებით. მართლაც, როდესაც რამდენიმე შერჩევა ჩართულია კარგ ცხრილში ან ფორმაში, მაშინ მათი საშუალებით შესაძლებელია ბევრი რამ გავიგოთ მონაცემთა ისტორიაზე. ჩვენ განვიხილავთ წერტილოვან, ღეროვან-ფოთლოვან, სიხშირის დიაგრამებს, ჰისტოგრამებს, გაბნევის დიაგრამებს და ვისაუბრებთ მათ სასარგებლო მხარეებზე.

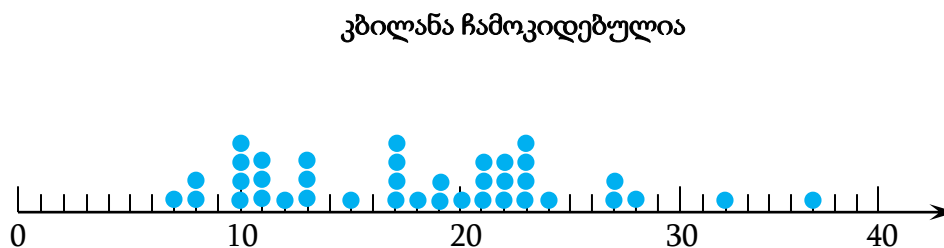
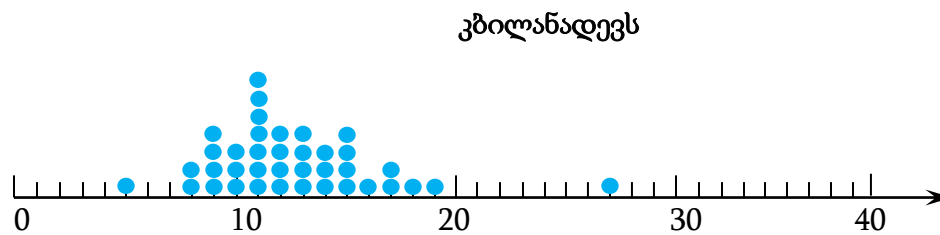
როდესაც საინჟინრო კვლევა წარმოებს მცირე ან ზომიერი რაოდენობის ერთცვლადიან რაოდენობრივ მონაცემებზე, ძალიან იოლია ფანქრისა და ფურცლის საშუალებით დაიხაზოს მარტივი წერტილოვანი დიაგრამა, რომელიც საკმაოდ კარგად გამოსახავს შექმნილ სიტუაციას. წერტილოვანი დიაგრამაზე ყოველი დაკვირვების შედეგი გამოსახულია წერტილით, რომელიც მდებარეობს რიცხვითი ღერძის იმ ადგილზე, რომელიც შეესაბამება მონაცემის მნიშვნელობას.

მაგალითი 2.1. მოვიყვანოთ მონაცემები ექსპერიმენტიდან, რომელიც შეეხება საკითხს: როგორ დავტვირთოთ კბილანა უწვეტად მომუშავე ნახშირის ღუმელში ისე, რომ მინიმუმამდე იქნეს დაყვანილი დამახინჯებები თერმული დამუშავების დროს. ექსპერიმენტის შედეგად მიღებულია ორი რიცხვითი სიმრავლე კბილანის ორი მდგომარეობისთვის (კბილანა დევს და კბილანა ჩამოკიდებულია), რომლებიც მოცემულია .0001 დუიმებში.

კბილანა დევს	კბილანა ჩამოკიდებულია
5, 8, 8, 9, 9,	7, 8, 8, 10, 10,
9, 9, 10, 10, 10,	10, 10, 11, 11, 11,
11, 11, 11, 11, 11,	12, 13, 13, 13, 15,
11, 11, 12, 12, 12,	17, 17, 17, 17, 18,
12, 13, 13, 13, 13,	19, 19, 20, 21, 21,
14, 14, 14, 15, 15,	21, 22, 22, 22, 23,
15, 15, 16, 17, 17,	23, 23, 23, 24, 27,
18, 19, 27,	27, 28, 31, 36,

ცხრილი 2.1. *ორი რიცხვითი სიმრავლე კბილანის ორი მდგომარეობისთვის*

ეს მონაცემები უფრო ნათლად შეიძლება გამოვსახოთ ქვემოთ მოყვანილი წერტილოვანი დიაგრამის სახით, რომელიც მაგალითად გვიჩვენებს, რომ პირველი მონაცემები მეორესთან შედარებით თითქმის ერთადაა თავმოყრილი.



ნახაზი 2.2. *წერტილოვანი დიაგრამები (.0001 in.)*

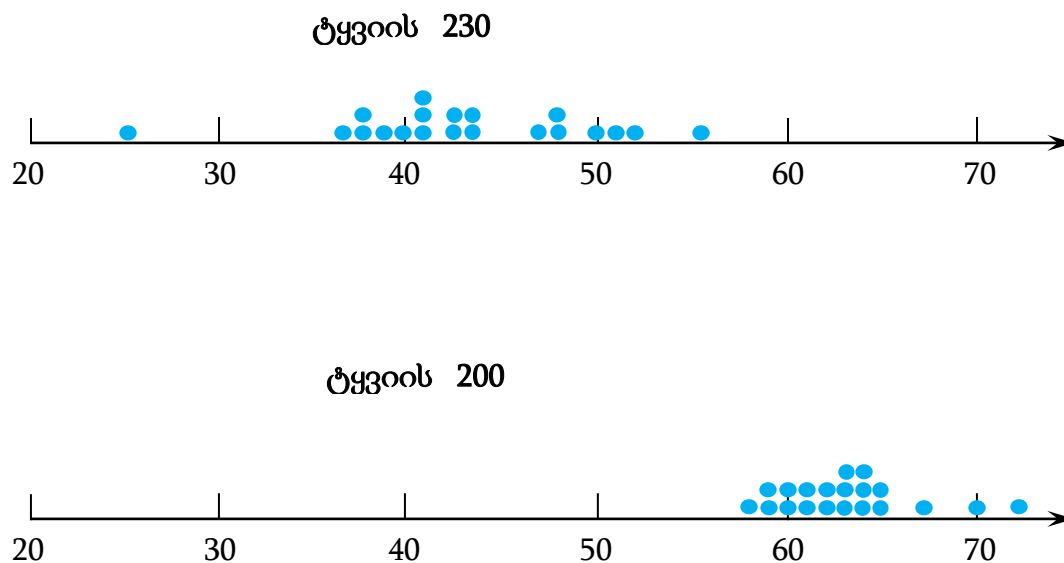
მაგალითი 2.2. *ტყვიის შეღწევადობის სიღრმე*

სტუდენტები ადარებენ 0.45 კალიბრიანი სხვადასხვა ტყვიის შეღწევადობის სიღრმეს მუხაში 15 ფუტი მანძილიდან. შემდეგი ცხრილი გვიჩვენებს შეღწევადობის სიღრმეს (მილიმეტრებში) ტყვიის ორი ტიპისთვის.

230 გრანულიანი ტყვია	200 გრანულიანი ტყვია
40.50, 38.35, 56.00, 42.55,	63.80, 64.65, 59.50, 60.70,
38.35, 27.75, 49.85, 43.60,	61.30, 61.50, 59.80, 59.10,
38.75, 51.25, 47.90, 48.15,	62.95, 63.55, 58.65, 71.70,
42.90, 43.85, 37.35, 47.30,	63.30, 62.65, 67.75, 62.30,
41.15, 51.60, 39.75, 41.00,	70.40, 64.05, 65.00, 58.00,

ცხრილი 2.3. ტყვიის შეღწევადობის სიღრმე (მმ)

შემდეგი გრაფიკები კი წარმოდგენენ შესაბამის წერტილოვან დიაგრამებს.



ნახაზი 2.4. ტყვიის შეღწევადობის სიღრმის წერტილოვანი დიაგრამები (მმ)

ეს დიაგრამები გვიჩვენებს, რომ შეღწევადობა 200 მარცვლოვანი ტყვიის უფრო დიდია და თანმიმდევრული, ვიდრე 230 მარცვლოვანის. (სტუდენტებმა გააკეთეს პროგნოზი, ტყვიების შეღწევადობა მეტია, რაც მსუბუქია ტყვია, მეტია საწყისი სიჩქარე და მცირეა ის ფართობი, რომელზედაც შესაძლებელია იყოს ხახუნის.)

წერტილოვან დიაგრამაზე ზოგადად უკეთესად შეიგრძნობა მონაცემთა სიმრავლე, მაგრამ ყოველთვის არ ხდება მონაცემების მნიშვნელობების აღდგენა ისეთნაირად,

რომ შესაძლებელი იყოს მათზე მოქმედებების ჩატარება დიდი სიზუსტით. **ლეროვან - ფოთლოვანი** დიაგრამა ზოგადად იძლევა მეტ ვიზუალურ ინფორმაციას, თანაც საწყისი ზუსტი მნიშვნელობების შენარჩუნებით. ლეროვან - ფოთლოვანი დიაგრამა შედგენილია ისე, რომ ითვალისწინებს თითოეული მონაცემის მცირე ციფრებსაც და თან გვიჩვენებს სად ვარდებიან ისინი.

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

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

ნახაზი 2.5. ლეროვან-ფოთლოვანი დიაგრამები დადებული კბილანის შემთხვევაში (მაგალითი 1 -სთვის)

ამ დიაგრამაზე გამოსახულია ორი შესაძლო გრაფიკი. ორივეზე დიაგრამაზე პირველი ციფრი ვერტიკალური წრფის მარცხნივ წარმოადგენს „ლეროს“. ციფრებს ამ წრფის მარჯვნივ უწოდებენ ფოთლებს და გვამლევენ დაკვირვების მეორე ციფრებს. მეორე დიაგრამა გვიჩვენებს იგივეს, მაგრამ უფრო დეტალურად ვიდრე პირველი. მაგ. ფოთლებს წარმოგვიდგენს ორ ნაწილად, „0 — 4“ და „5 — 9“ პოზიციებად. პირველში არის ყველა ციფრი ერთად „0 — 9“ -მდე.

მაგალითი 2.2. (გაგრძელება). ახლა შევადგინოთ ლეროვან-ფოთლოვანი დიაგრამა 200 მარცვლოვანი ტყვიის მონაცემებისთვის. აქაც შესაძლებელი იქნება ორი ვარიანტი. პირველში მოსახერხებელია, რომ წრფის მარჯვენა მხარეს დაიწეროს ორი ციფრი, რომელიც წარმოადგენს ლეროს და დანარჩენი ციფრები წარმოადგენენ ფოთლებს.

58	.64, .00
59	.50, .80, .10

58	.00, .65
59	.10, .50, .80

60	.70	60	.70
61	.30, .50	61	.30, .50
62	.95, .65, .30	62	.30, .65, .95
63	.80, .55, .30	63	.30, .55, .80
64	.65, .05	64	.05, .65
65	.00	65	.00
66		66	
67	.75	67	.75
68		68	
69		69	
70	.40	70	.40
71	.70	71	.70

ნახაზი 2.6. 200 მარცვლოვანი ტყვიის შეღწევადობის სიღრმის ღეროვან - ფოთლოვანი დიაგრამა

პირველი გამოსახულებაში ციფრები ჩაწერილია იმ მიმდევრობით, როგორც ეს მოცემულია ცხრილში (მონაცემებში), მარცხნიდან მარჯვნივ და ზემოდან ქვემოთ. მეორე დიაგრამა უკეთესია იმ მხრივ, რომ მონაცემები დალაგებულია ზრდის მიხედვით. შევნიშნოთ, რომ ორივე დიაგრამა არსებითად ერთსა და იმავე მიზანს ემსახურება, ვიზუალურად წარმოადგინოს მონაცემები.

რჩევა: როდესაც ვადარებთ მონაცემთა სიმრავლეებს, სასარგებლოა ღეროვან - ფოთლოვანი დიაგრამები წარმოვაგინოთ ერთად, **ზურგი - ზურგთან**.

კბილანა დევს	კბილანა ჩამოკიდებულია
9 9 9 9 8 8 5	0 7 8 8
4 4 4 3 3 3 3 2 2 2 2 1 1 1 1 1 1 0 0 0	1 0 0 0 0 1 1 1 2 3 3 3
9 8 7 7 6 5 5 5 5	1 5 7 7 7 7 8 9 9
	2 0 1 1 1 2 2 2 3 3 3 4
	7 2 7 7 8
	3
	3

ნახაზი 2.7. ღეროვან-ფოთლოვანი დიაგრამები ზურგი-ზურგთან (მაგალითი 2.1 -სთვის)

ეს ნახაზი თვალნათლივ გვიჩვენებს განსხვავებას მონაცემთა ორ სიმრავლეს შორის. გვიჩვენებს აგრეთვე, განსხვავების ადგილმდებარეობას, თითოეული სიმრავლის გაფანტულობებსა და განსხვავებებს მათ შორის.

2.2. სიხშირეთა ცხრილები და ჰისტოგრამები

წერტილოვანი და ღეროვანი - ფოთლოვანი დიაგრამები იოლად და კარგად გამოსაყენებელი სქემებია მონაცემთა სიმრავლეებზე მსჯელობის დროს. თუმცა ისინი არ არიან ხშირად ხმარებადი ფორმები პრეზენტაციებისა და მოხსენებების დროს. ამ მეტად ოფიციალურ გარემოებებში უფრო გამოყენებადია სიხშირეთა ცხრილი და ჰისტოგრამები.

სიხშირეთა ცხრილი შედგენილია შემდეგნაირად: პირველ რიგში ვიღებთ ინტერვალს, რომელიც მთლიანად შეიცავს მონაცემთა სიმრავლის შესაბამის რიცხვებს. შემდეგ ამ მონაკვეთს ვყოფთ მცირე და ტოლ მონაკვეთებად, ჩავინიშნავთ მონაცემთა რაოდენობას რომლებიც მოხვდებიან შესაბამის ინტერვალებში. ბოლოს სიხშირე, ფარდობითი სიხშირე და დაგროვებითი სიხშირე შეიძლება შევკრიბოთ.

მაგალითი 2.1. (გაგრძელება). შემდეგი ცხრილი წარმოადგენს ერთერთ შესაძლო სიხშირეთა ცხრილს. ფარდობითი სიხშირის მნიშვნელობა მიიღება სიხშირის შეფარდებით 38 -თან, ანუ მონაცემთა რაოდენობასთან.

შუალედები (.0001 დუიმი)	იარლიყი	სიხშირე	ფარდობითი სიხშირე	დაგროვებითი ფარდობითი სიხშირე
5-8		3	.079	.079
9-12		18	.474	.553
13-16		12	.316	.869
17-20		4	.105	.974
21-24		0	0	.974
25-28		1	.026	1.000
		38	1.000	

ცხრილი 2.8. დადებული კბილანის მონაცემების სიხშირეთა ცხრილი

ჩანაწერები დაგროვებითი ფარდობითი სიხშირის სვეტში არიან წილადები, რომლებიც მიიღებიან წინა ფარდობითი სიხშირეების შეკრებით. „იარლიყების“ სვეტი გვაძლევს იგივე ტიპის ინფორმაციას განაწილების ფორმაზე, რასაც წერტილოვანი ან ღეროვანი - ფოთლოვანი დიაგრამა.

(დამოუკიდებლად გააკეთეთ ყველაფერი ანლოგიურად ჩამოკიდებული კბილანისთვისაც!)

სიხშირეთა ცხრილის შედგენისთვის საჭირო ინტერვალის შედგენა არის განსჯის საგანი. არ ხდება, რომ ორი ადამიანი აუცილებელად აირჩევს ერთსა იმავე ინტერვალს. ვინაიდან არსებობს მთელი რიგი მიზეზები, რის გამოც ადამიანი გონებაში იმახსოვრებს რაღაც რიცხვებს, სანამ დაიწყებდეს მათ ამორჩევას. ესთეტიკური თვალსაზრისით, მაგალითად, დამრგვალებულ რიცხვებს უპირეტესობა აქვთ მონაკვეთის კიდურა წერტილების შერჩევის დროს. ცხადია, ამის გამო ადგილი აქვს ინფორმაციის დაკარგვას. პრობლემურია ინტერვალების რაოდენობის შერჩევა, რადგან თითოეულ გადაწყვეტილებას აქვს თავისი ურთიერთსაპირისპირო მხარეები და სხვა.

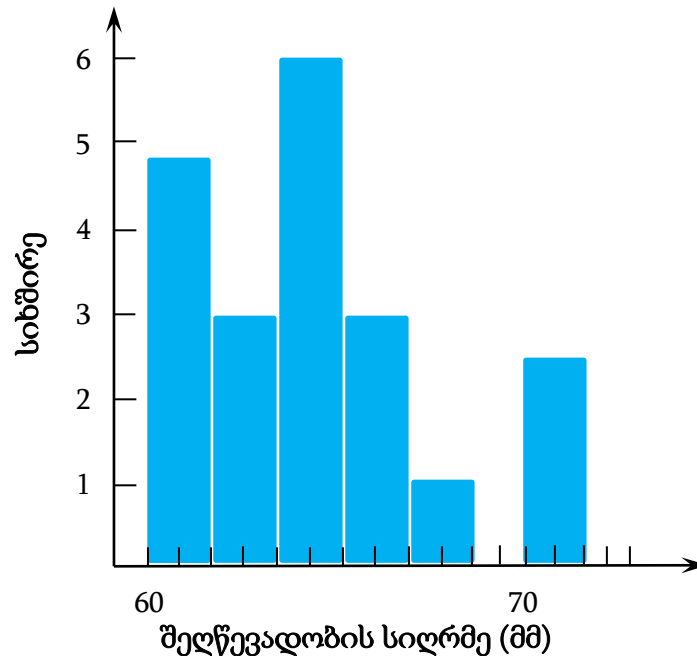
სიხშირეთა ცხრილის შედგენის შემდეგ უკეთესია მისი წარმოდგენა *ჰისტოგრამის* სახითაც. სიხშირეთა ან ფარდობით სიხშირეთა *ჰისტოგრამა* არის სვეტოვანი გრაფიკები, რომლებიც გამოიყენება მონაცემთა განაწილების ფორმის წარმოსაჩენად.

მაგალითი 2.2. (გაგრძელება).

შელწევადობის სიღრმე (მმ)	იარლიყი	სიხშირე	ფარდობითი სიხშირე	დაგროვებითი ფარდობითი სიხშირე
58.00 – 59.00		5	.25	.25
60.00 – 61.00		3	.15	.40
62.00 – 63.00		6	.30	.70
64.00 – 65.99		3	.15	.85
66.00 – 67.00		1	.05	.90
68.00 – 69.00		0	0	.90
70.00 – 71.99		2	.10	1.00
		20	1.00	

ცხრილი 2.9. *200 მარცვლოვანი ტყვიის შელწევადობის სიღრმის სიხშირეთა ცხრილი*

თუ ამ ცხრილის გადავიყვანთ ჰისტოგრამაში მივიღებთ შემდეგ სურათს:



გრაფიკი 2.10. *200 მარცვლოვანი ტყვიის შელწევადობის სიღრმის ჰისტოგრამა*

ამ ნახაზის ვერტიკალურულ ღერძი არის სიხშირეთა სკალა, ამიტომ ამ გრაფიკს ეწოდება **სიხშირეთა ჰისტოგრამა**. თუ ვერტიკალულ სკალაზე დავიტანთ ფარდობით სიხშირეებს, მაშინ მივიღებთ **ფარდობით სიხშირეთა ჰისტოგრამა**.

ზოგადი მითითებები, როგორ ავაგოთ ჰისტოგრამები

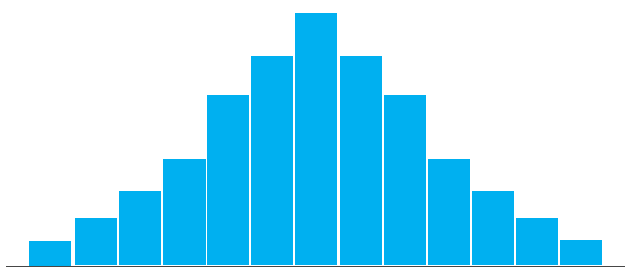
- გამოვიყენოთ ტოლი სიგრძის ინტერვალები
- ვერტიკალურ ღერძზე მთელი რიცხვები დაწყებული 0-დან
- ავერიდოთ ღერძების დაწყებას
- დავიტანოთ ღერძებზე საინფორმაციო შუალედები
- სვეტის ცენტრი იყოს ინტერვალის შუა წერილში.

ასეთნაირად აგებული ჰისტოგრამის ტოლი ფართობები შეესაბამება ტოლ რიცხვებს მონაცემთა სიმრავლიდან. სვეტების განლაგების წესი თვალნათლივ გამოხატავს წრტილების განლაგებას მონაცემთა სიმრავლეში. თუ არ გავყვებით ამ მითითებებს, შედეგად შეიძლება მივიღოთ ჰისტოგრამა, რომელმაც შეიძლება შესაბამისად ვერ წარმოადგინოს მონაცემთა სიმრავლე.

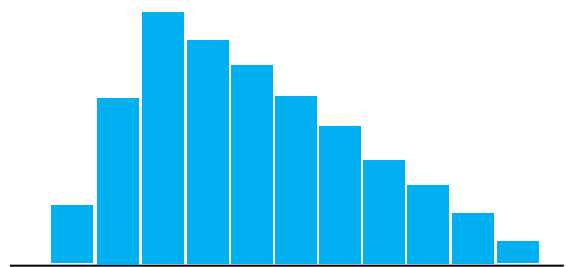
შემდეგი ფიგურები გვიჩვენებენ ტერმინოლოგიების შესაბამის განაწილების ფორმებს, რომლებიც გვხვდებიან მაშინ, როდესაც ვადგენთ წერტილოვან გრაფიკებს, ღეროვან - ფოთლოვან დიაგრამებს და ჰისტოგრამებს.

გრაფიკული და ცხრილური გამოსახულებები ერთი შეხედვით ძალზე მარტივი შეიძლება მოგვეჩვენოს, მაგრამ ეს პირველი შეხედულება მატყუარაა. მისი რეგულარულად და ჰკვიანურად გამოყენება არის მძლავრი ინჟინრული იარაღი. ინფორმაცია მონაცემების ადგილმდებარეობის, გაზნევის, ფორმის შესახებ და უფრო ნათლად კი ჰისტოგრამა იძლევიან ზუსტ მინიშნებებს, იმაზე თუ როგორ ფუნქციონერებენ ფიზიკური პროცესები, რომლებიც წარმოშობენ ამ მონაცემებს.

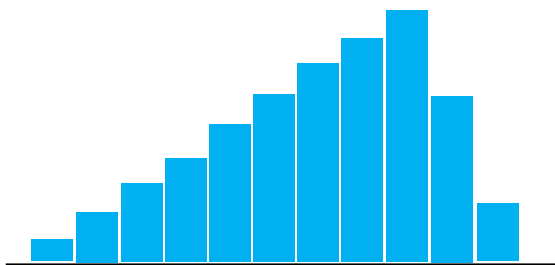
განაწილების ფორმის ინჟინრული ინტერპრეტაციის მაგალითები



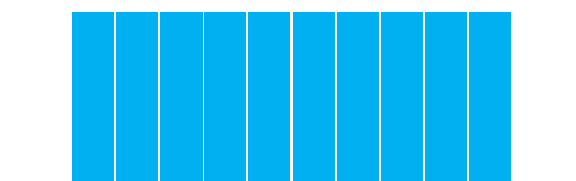
ზარის-ფორმის



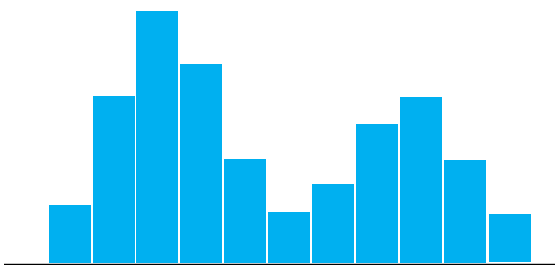
მარჯვნივ დახრილი



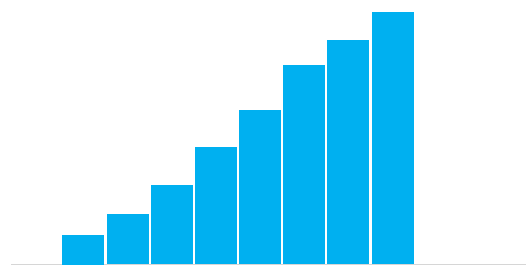
მარცხნივ დახრილი



თანაბარი



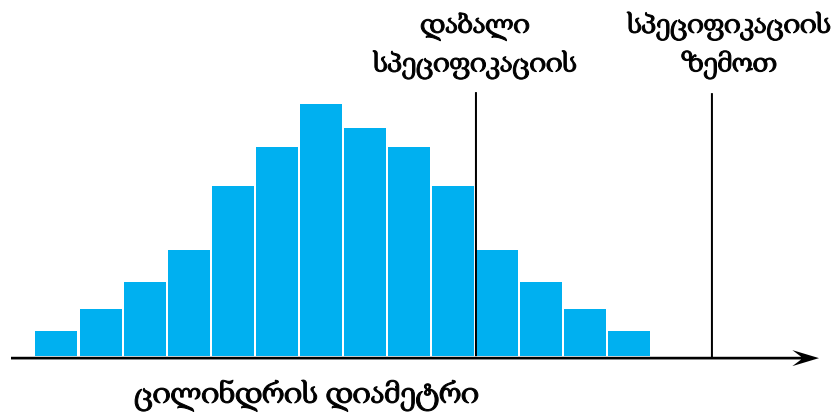
ბიმოდალური



ჩამოჭრილი

გრაფიკები 2.11. *განაწილების ფორმები*

მაგალითად, თუ მონაცემები მექანიკურად დამუშავებული ლითონის ცილინდრების დიამეტრების შესახებ წარმოდგენილია ჰისტოგრამის სახით, დიდი შანსია რომ ეს ჰისტოგრამა იყოს **ბიმოდალური** (ან უფრო **მულტიმოდალური**, რომელსაც აქვს რამდენიმე ამოზურცულობა). ამ დროს უნდა ვივარაუდოთ, რომ დეტალის მექანიკური დამუშავება წარმოებდა ერთზე მეტ დანადგარზე, ან ერთზე მეტი ოპერატორის მიერ, ან სხვადასხვა დროს. პრაქტიკულად ამ მრავალმექანიკური დამუშავების პროცესის გრძელ თანმიმდევრობაში დიამეტრების ზომების განაწილებას ექნება მეტი ცვალებადობა ვიდრე მაშინ, როცა არის ერთი მოწყობილობა, ერთი ოპერატორი და ერთი რეგულაცია. სხვა შესაძლებლობა ჰისტოგრამის, მაგ. **ჩამოჭრილი**, შეიძლება შედეგი იყოს ასეთი ფაქტის; შემოწმდა ცილინდრების დიდი რაოდენობა (შეიძლება 100%) და რაღაც მნიშვნელობის მერე ყველა დაწუნებულ იქნა გადაჭარბებული დიამეტრის გამო. ან, შესაძლებელია ჰისტოგრამაზე იყოს საინჟინრო სპეციფიკაციის (მოთხოვნების) მონიშვნა, რომლის იქითაც პროდუქტი წუნდებულია. ეს სიტუაცია წარმოდგენილია ქვედა სურათზე.



გრაფიკი 2.12. საინჟინრო სპეციფიკაციის მონიშნული ჰისტოგრამა

ამ ფორმების კვლევა სასარგებლოა არა მარტო იმიტომ, რომ იძლევა ინფორმაციის ფიზიკურ პროცესებზე, არამედ ეს ფორმები მეტად მნიშვნელოვანია იმიტაც, რომ იძლევა საშუალებას შემუშავდეს შესაბამისი მეთოდები ფორმალური სტატისტიკური დასკვნების გასაკეთებლად, რაც თავის მხრივ არის საფუძველი სწორი პროგნოზების გასაკეთებლად. აქვე შევნიშნოთ, რომ მეთოდოთოგია, რომელიც შეესაბამება ერთ რომელიმე განაწილების ფორმას, შეიძლება არ იყოს შესაბამისი სხვა ფორმისთვის.

ცოტა მოგვიანებით, ჭანჭიკის მოჭერის ბრუნვის მომენტის მაგალითზე ჩვენ განვიხილავთ მონაცემების ცვლადებს შორის დამოკიდებულებას, რომელიც წარმოდგენილი ე.წ. **გაბნევის დიაგრამის** სახით.

2.3. კვანტილები და მათთან დაკავშირებული გრაფიკები

მკითხველთა უდიდესმა ნაწილმა იცის რა არის *პროცენტული მაჩვენებელი*. მაგალითად, ეს კარგად ჩანს საგანმანათლებლო ტესტების შესრულების შემდეგ მიღწეული წარმატებების სტატისტიკური შეფასების დროს. დავუშვათ, ვიღაც კონკრეტული პიროვნება შედეგების მიხედვით მოხვდა გამოსაცდელთა მიერ მიღებული ქულების 80% -მდე მაჩვენებელში. უხეშად რომ ვთქვათ, ეს იმას ნიშნავს, რომ მასზე უკეთესი შედეგები აქვს გამოსაცდელთა 20% -ს. ასეთი კონცეფცია ძალიან ხშირად მოსახერხებელია საინჟინრო მონაცემების შესწავლისას. თუმცა ასევე ხშირად უფრო მოსახერხებელია ვიმუშაოთ წილადებთან 0 -დან 1-მდე, ვიდრე რიცხვებზე 0-დან 100 -მდე. ამისთვის გამოვიყენებთ ახალ ტერმინოლოგიას „კვანტილი“.

კვანტილები. მიახლოებით რომ ვთქვათ, ავიღოთ 0 -სა და 1-ს შორის რაიმე p რიცხვი. მაშინ მონაცემთა განაწილების p **კვანტილი** არის ისეთი რიცხვი, რომ წილადი p მდებარეობს ამ რიცხვიდან მარცხნივ და $1 - p$ წილადი კი მარჯვნივ. მონაცემთა სიმრავლის დისკრეტულობისა და სასრულობის გამო აუცილებელია ჩამოვაყალიბოთ ზუსტი განმარტება იმ ტერმინებში, რომელთაც შემდგომში მოვიხმართ.

განმარტება 2.1. ვთქვათ, მონაცემთა სიმრავლე შედგება n დალაგებული მნიშვნელობებისგან $x_1 \leq x_2 \leq \dots \leq x_n$,

1. თუ დადებითი მთელი $i \leq n$ რიცხვებისთვის $p = \frac{i-.5}{n}$, მაშინ მონაცემთა სიმრავლის p **კვანტილი** არის

$$Q(p) = Q\left(\frac{i-.5}{n}\right) = x_i$$

(მონაცემთა სიმრავლის i -ურ წერტილს ეწოდება $\frac{i-.5}{n}$ კვანტილი.)

2. i მთელი რიცხვისთვის დანებისმიერი p რიცხვისთვის, რომელიც არის $\frac{.5}{n}$ -სა და $\frac{i-.5}{n}$ რიცხვებს შორის და არ არის $\frac{i-.5}{n}$ ფორმის, მაშინ მონაცემთა სიმრავლის p **კვანტილი** მიიღება წრფივი ინტერპოლაციის საშუალებით $Q\left(\frac{i-.5}{n}\right)$ და $\frac{i-.5}{n}$ წერილებს შორის.
(ორივე ამ შემთხვევაში აღნიშვნა $Q(p)$ გამოყენება p კვანტილის აღსანიშნავად).

ეს განმარტება განსაზღვრავს $Q(p)$ -ს ყოველი p -სთვის $.5/n$ -დან $(n-1)/n$ -მდე. $Q(p)$ -ს საპოვნელად შეიძლება ამოვხსნათ განტოლება $p = (i - .5)/n$. საიდანაც,

$$i = np + .5$$

და არის $(np + .5)$ -ე წევრი დალაგებულ მონაცემთა სიმრავლეში.

მაგალითი 2.3. მშრალი ქაღალდის პირსახოცის მოხვეის ძალის კვანტილები.

სტუდენტებმა: ლიმ, სებათიმ და სტრაუზმა გამოიკვლიეს მშრალი ქაღალდის პირსახოცის მოხვეის ძალები რამდენიმე ბრენდისთვის. გაზომვის შედეგები (გრამებში) მოყვანილია შემდეგ ცხრილში.

ტესტი	მოხვეის ძალა(გრ)
1	8,577
2	9,471
3	9,011
4	7,583
5	8,572
6	10,688
7	9,614
8	9,614
9	8,527
10	9,165

ცხრილი 2.13. ქაღალდის ათი პირსახოცის მოხვეის ძალები

თუ დავალაგებთ ძალების ამ მონაცემებს და გამოვითვლით $\frac{i-.5}{10}$ მნიშვნელობებს მივიღებთ: .05, .15, .25, . . . , .85 და .95 ძალთა განაწილების კვანტილებს, რომლებიც შემდეგ ცხრილშია მოცემული:

i	$\frac{i - 5}{10}$	i -ური მონაცემის წერტილი $x_i = Q\left(\frac{i - 5}{10}\right)$
1	.05	7,583 = $Q(.05)$
2	.15	8,527 = $Q(.15)$
3	.25	8,572 = $Q(.25)$
4	.35	8,577 = $Q(.35)$
5	.45	9,011 = $Q(.45)$
6	.55	9,165 = $Q(.55)$

7	.65	9,471 = $Q(.65)$
8	.75	9,614 = $Q(.75)$
9	.85	9,614 = $Q(.85)$
10	.95	10,688 = $Q(.95)$

ცხრილი 2.14. ქალაქის პირსახოცის მოხვევის ძალის განაწილების კვანტილები

რადგანაც მოცმული გვაქვს $n = 10$ მონაცემის წერილი, ამიტომ თითოეული მოიცავს მონაცემის 10% -ს. თუ გამოვიყენებთ განმარტება 2.1-ის პირველ ნაწილს ვიპოვით (მაგალითად) .35 კვანტილს, იქნება 3 უმცირესი მონაცემის წერტილი და მეოთხის ნახევარი მდებარეობს განსახილავი წერტილის მარცხნივ. ექვსი უდიდესი წერტილი და მეშვიდეს ნახევარი კი მდებარეობს მარჯვენა მხარეს. ე.ი. მონაცემთა მეოთხე უმცირესი იქნება .35 კვანტილი, როგორც ეს ნაჩვენებია ცხრილში 2.13.

განმარტების მეორე პუნქტის საილუსტრაციოდ განვიხილოთ .5 და .93 კვანტილები. რადგანაც .5 არის $\frac{.5 - .45}{.55 - .45} = .5$, მაშინ .45 -დან .55 -მდე წრფივი ინტერპოლაცია გვაძლევს:

$$Q(.5) = (1 - .5)Q(.45) + .5Q(.55) = .5 \cdot 9,011 + .5 \cdot 9,165 = 9,088 \text{ გრ}$$

მაშინ, .93 არის $\frac{.93 - .85}{.95 - .85} = .8$, ხოლო .85 -დან .95 -მდე წრფივი ინტერპოლაცია გვაძლევს:

$$Q(.93) = (1 - .8)Q(.85) + .8Q(.95) = .5 \cdot 9,614 + .8 \cdot 9,165 = 10,688 = 10,473.2 \text{ გრ}$$

კერძოდ, p -ს დამრგვალებული მნიშვნელობებისთვის $Q(p)$ კვანტილებს აქვთ სპეციალური სახელი.

განმარტება 2.2. $Q(.5)$ -ს ეწოდება განაწილების მედიანა .

განმარტება 2.3. $Q(.25)$ და $Q(.75)$ ეწოდებათ განაწილების პირველი (ანუ დაბალი) კვანტილი და მესამე (ანუ მაღალი) კვანტილი, შესაბამისად.

მაგალითი 2.3. (გაგრძელება) თუ დავუბრუნდებით ცხრილს 2.14 და მნიშვნელობას $Q(.5)$, რომელიც წინასწარ გამოვითვალეთ, მოხვევის ძალის განაწილებისთვის გვექნება:

$$\text{მედიანა} = Q(.5) = 9,088 \text{ გრ}$$

პირველი კვარტილი = $Q(.25) = 8,572$ გრ

მესამე კვარტილი = $Q(.75) = 9,614$ გრ

შესაძლებელია კვანტილის გრაფიკული წარმოდგენა.

განმარტება 2.4. კვანტილის გრაფიკი არის გრაფიკი $Q(p)$ vs. p , დალაგებული n ზომის მონაცემთა $x_1 \leq x_2 \leq \dots \leq x_n$ სიმრავლისთვის ისეთი, რომ ასახვა წერტილებს შორის დამყარებულია შემდეგი წესით $\left(\frac{i-.5}{n}, x_i\right)$ და სიბრტყეზე მიღებული წერტილები შეერთებულია წრფის მონაკვეთებით.

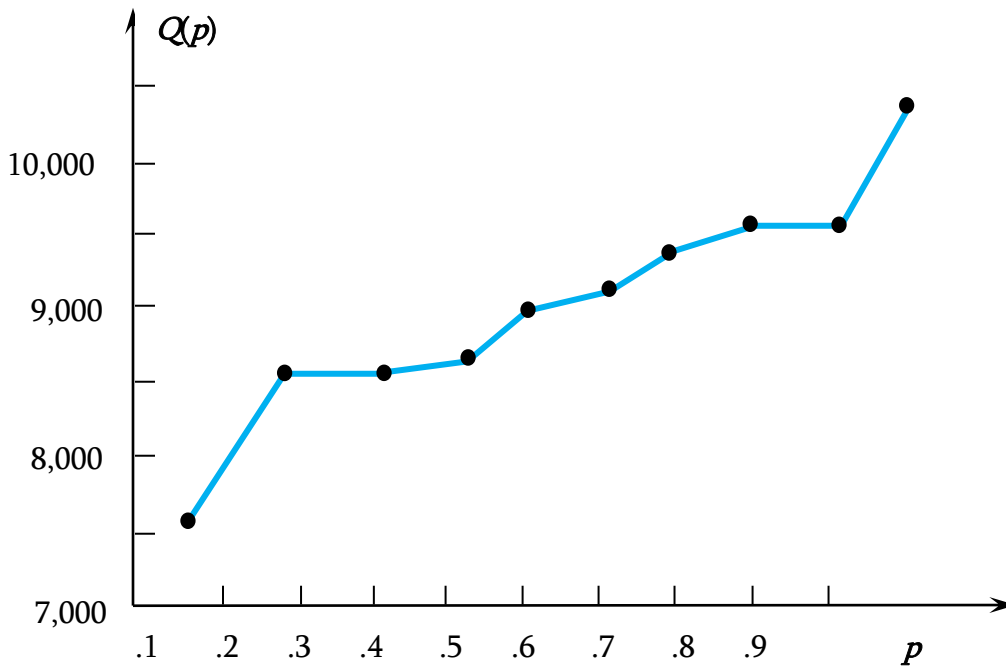
(vs. შემოკლებული აღნიშვნაა სიტყვისა “versus - წინააღმდეგ, პირისპირ, მიმართ, ოპოზიშიაში მყოფი“)

ამიტომაც, განმარტების მეორე პუნქტში ვიყენებთ ტერმინს წრფივი ინტერპოლაცია, რაც ნიშნავს წრფის მონაკვეთების დატანას სიბრტყეზე, და საბოლოოდ გვაძლევს კვანტილის გრაფიკს.

მაგალითი 2.3. (გაგრძელება). მივმართოთ კვლავ 2.14 ცხრილს $\frac{i-.5}{10}$ კვანტილებისთვის მოხვევის ძალის განაწილებისთვის შესახებ. ცხადია, რომ კვანტილების წერტილოვანი გრაფიკი უკვე შეიცავს მინიშნებას შემდეგი წყვილების მიმდევრობითი შეერთების შესახებ.

(.05; 7,583)	(.15; 8,527)	(.25; 8,572)	(.35; 8,577)	(.45; 9,011)
(.55; 9,165)	(.65; 9,471)	(.75; 9,614)	(.85; 9,614)	(.95; 10,688)

და ვლგებულობთ ასეთ გრაფიკს:



გრაფიკი 2.15. ქაღალდის პირსახოცების მოხვეის ალის კვანტილის გრაფიკი

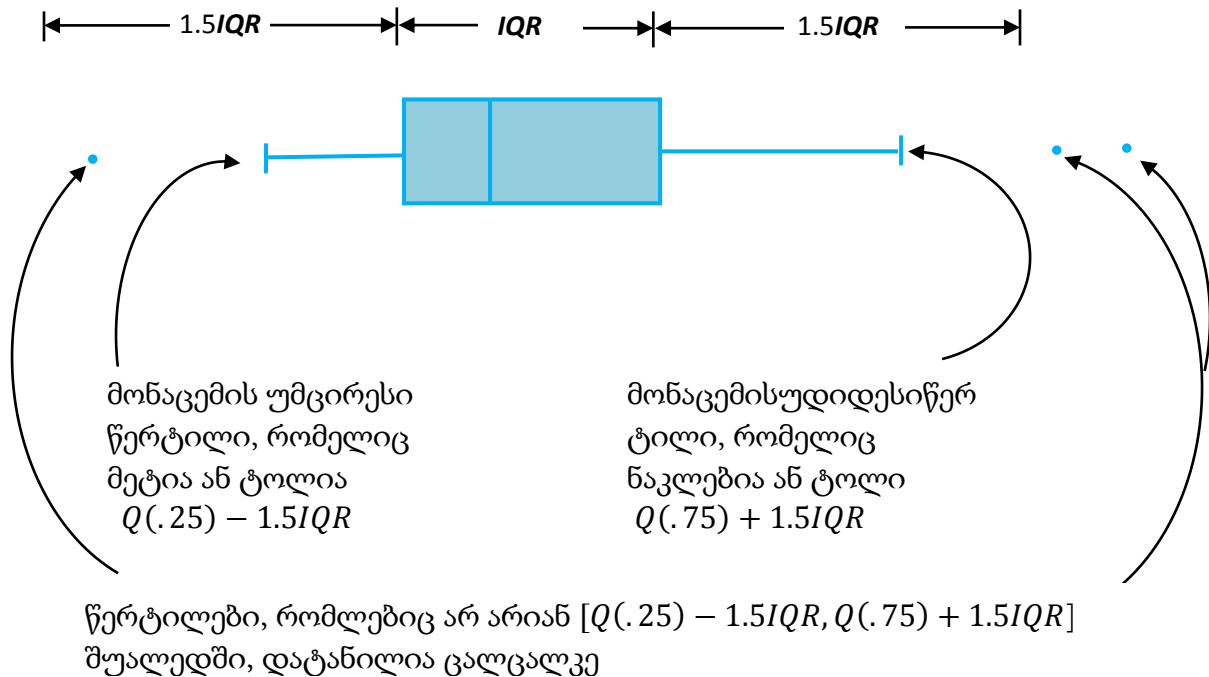
კვანტილის გრაფიკი ყოველთვის გამოიყენება ზოგიერთი ინფორმაციის მისაღებად. მაგალითად, ინფორმაცია გრაფიკის ვიზუალურ სიგლუვეზე, რომელიც შეიძლება იყოს კომპენსაცია რაიმე გადაწყვეტილებაზე. (ტაქტიკური დაშვება არის ის, რომ ძირითადად მონაცემებმა თავისთავად მოგვცენ უფრო და უფრო გლუვი კვანტილის გრაფიკები უფრო და უფრო დიდი შერჩევებისთვის.)

ბოქსპლოტი. კვანტილების იდეის გაცნობა წინა პირობაა ე.წ. ბოქსპლოტის (ბრტყელი გრაფიკების) ასაგებად, რომელიც არის წერტილოვანი დიაგრამების და ჰისტოგრამების ალტერნატივა. მართალია იგი ოდნავ ნაკლები ინფორმაციის შემცველია, მაგრამ მას აქვს ის უპირატესობა, რომ მართკუთხედები შეიძლება განთავსდნენ ერთმანეთის გვერდი-გვერდ შედარების მიზნით.

არსებობს რამდენიმე მოსაზრება ასეთი გრაფიკების ასაგებად. ერთი არის, რომ იგი გამოსახავს გრაფიკის ზოგად სახეს. შემდეგი, ბოქსპლოტი გამოსახულია პირველი კვარტილიდან მესამე კვარტილის ჩათვლით და გაყოფილია მედიანის ხაზით. ამიტომ კვარტილთშორისი რანგი (ანუ გაბნევის დიაპაზონი) არის

$$IQR = Q(.75) - Q(.25)$$

(IQR — interquartile range).



გრაფიკი 2.16. ზოგადი ბოქსპლოტი

და გამოითვლება როგორც, მონაცემის უმცირესი წერტილი $1.5IQR - Q(.25)$ -სა და მონაცემის უდიდესი წერტილი $1.5IQR + Q(.75)$ მნიშვნელობებს შორის. ზოგადად, მონაცემების უდიდესი რაოდენობა ვარდება შუალედში

$$[Q(.25) - 1.5 \cdot IQR; Q(.75) + 1.5 \cdot IQR].$$

ის წერტილები, რომლებიც ამ შუალედში არ მდებარეობენ, დატანილია ინდივიდუალურად. ამიტომ ის განისაზღვრება როგორც *შორეული* ან *უჩვეულო*.

მაგალითი 2.3 (გაგრძელება). განვიხილოთ ბოქსპლოტის აგება ქაღალდის პირსახოცის მოხვევის ძალის მონაცემთა სიმრავლისთვის. დავიწყეთ კვარტილებით,

$$Q(.25) = 8,572 \text{ გრ}, \quad Q(.5) = 9,088 \text{ გრ}, \quad Q(.75) = 9,614 \text{ გრ}.$$

ანუ,

$$IQR = Q(.75) - Q(.25) = 9,614 - 8,572 = 1,042 \text{ გრ}$$

და

$$1.5 \cdot IQR = 1,563 \text{ გრ.}$$

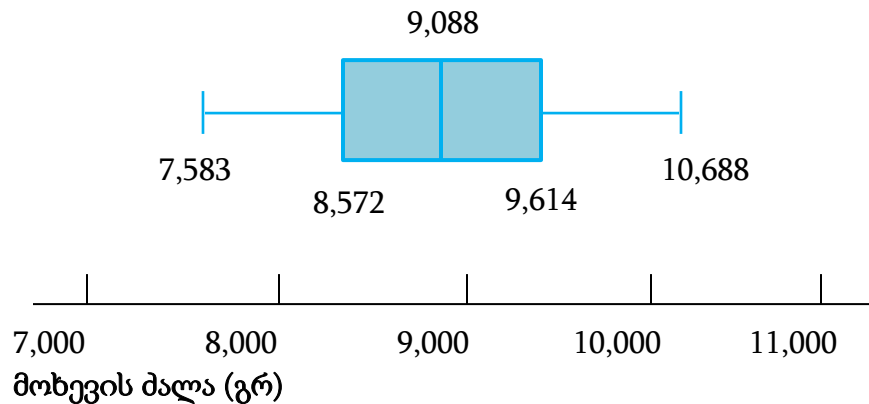
მაშინ,

$$Q(.75) + 1.5 \cdot IQR = 9,614 + 1,563 = 11,177 \text{ გრ}$$

და

$$Q(.25) - 1.5 \cdot IQR = 8,572 - 1,563 = 7,009 \text{ გრ.}$$

რადგანაც მონაცემის ყველა წერტილი მდებარეობს 7,009 გრ-სა და 11,177 გრ-ს შორის, ბოქსპლოტი ექნება ასეთი სახე:



გრაფიკი 2.17. პირსახოცის მოხვევის ძალის ბოქსპლოტი

ბოქსპლოტი გვიჩვენებს განაწილების განლაგებას და მართკუთხედების განთავსებას რიცხვითი ღერძის გასწვრივ. იგი გამოსახავს განაწილების გაფანტულობას და მის გავრცელებას მართკუთხედის საშუალებით, რომელიც შეიცავს საშუალოს, განაწილების 50% -ს. განაწილების ფორმის ზოგიერთი ელემენტი ნაჩვენებია სიმეტრიულობის გათვალისწინებით (ზოგჯერ კი მის გარეშე). წყვეტა მართკუთხედის ბოლოსა და იზოლირებულად მონიშნულ წერტილს შორის არის შეხსენება იმისა, რომ მონაცემის არც ერთი წერტილი არ ვარდება ამ ინტერვალში.

გამოვსახოთ კიდევ ორი ბოქსპლოტი ერთსა და იმავე სკალაზე „გვერდი - გვერდ“, რომელიც წარმოადგენს ეფექტურ გზას შერჩევათა შედარებისთვის.

მაგალითი 2.2. (გაგრძელება). მეტი რამ ტყვიის შეღწევადობაზე

შემდეგი ცხრილი მოიცავს ინფორმაციას, რომელიც საჭიროა $\frac{i-.5}{20}$ კვანტილების საპოვნელად ტყვიის შეღწევადობის სიღრმის ორი განაწილება 230 და 200 მარცვლოვანი ტყვიებისთვის.

i	$\frac{i-5}{20}$	i -ური უმცირესი (230) მონაცემის წერტილი $Q\left(\frac{i-5}{20}\right)$	i -ური უმცირესი (200) მონაცემის წერტილი $Q\left(\frac{i-5}{20}\right)$
1	.025	27.75	58.00
2	.075	37.35	58.65
3	.125	38.35	59.10
4	.175	38.35	59.50
5	.225	38.75	59.80
6	.275	39.75	60.70
7	.325	40.50	61.30
8	.375	41.00	61.50
9	.425	41.15	62.30
10	.475	42.55	62.65
11	.525	42.90	62.95
12	.575	43.60	63.30
13	.625	43.85	63.55
14	.675	47.30	63.80
15	.725	47.90	64.05
16	.775	48.15	64.65
17	.825	49.85	65.00
18	.875	51.25	67.75
19	.925	51.60	70.40
20	.975	56.00	71.70

ცხრილი 2.18. ტყვიის შეღწევადობის სიღრმის განაწილების კვანტილები

230 მარცვლოვანი ტყვიებისთვის შეღწევადობის სიღრმის შესახებ დამატებით შეიძლება გამოვითვალოთ

$$Q(.25) = .5 \cdot Q(.225) + .5 \cdot Q(.275) = .5 \cdot 38.75 + .5 \cdot 39.75 = 39.25 \text{ მმ},$$

$$Q(.5) = .5 \cdot Q(.475) + .5 \cdot Q(.525) = .5 \cdot 42.55 + .5 \cdot 42.90 = 42.725 \text{ მმ},$$

$$Q(.75) = .5 \cdot Q(.725) + .5 \cdot Q(.775) = .5 \cdot 47.90 + .5 \cdot 48.15 = 48.025 \text{ მმ}.$$

ანუ,

$$IQR = 48.025 - 39.25 = 8.775 \text{ მმ},$$

$$1.5 \cdot IQR = 13.163 \text{ მმ},$$

$$Q(.75) + 1.5 \cdot IQR = 61.188 \text{ მმ},$$

$$Q(.25) - 1.5 \cdot IQR = 26.087 \text{ მმ}.$$

(შეეცადეთ გააკეთოთ იგივე 200 -მარცვლოვანი ტყვიებისთვის)

ანუ იგივენაირად 200 -მარცვლოვანი ტყვიებისთვის გვექნება:

$$Q(.25) = 60.25 \text{ მმ}$$

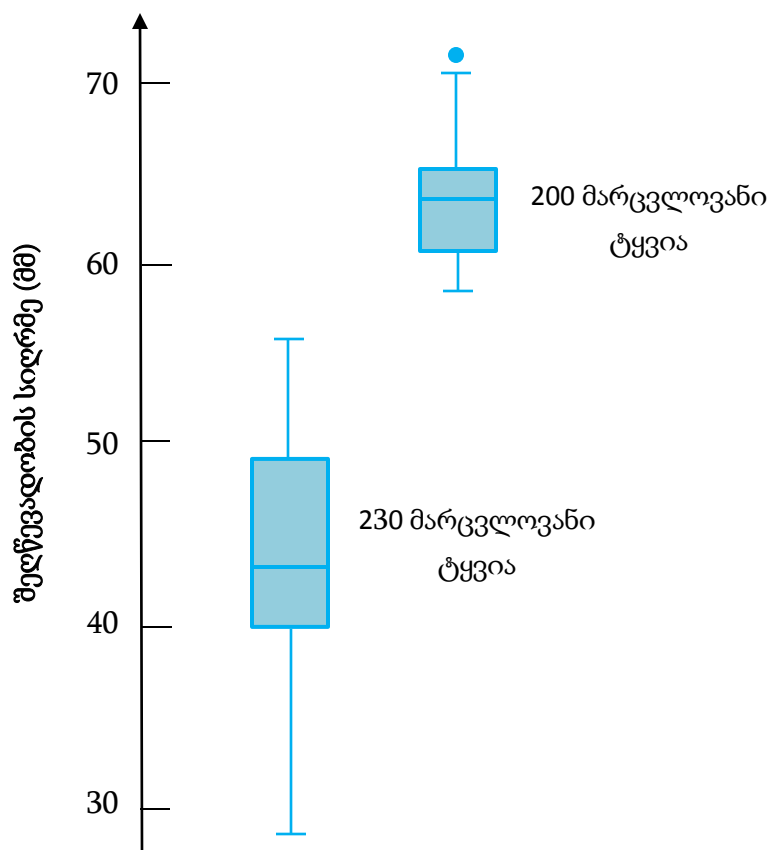
$$Q(.5) = 62.80 \text{ მმ}$$

$$Q(.75) = 64.35 \text{ მმ}$$

$$Q(.75) + 1.5 \cdot IQR = 70.50 \text{ მმ}$$

$$Q(.25) - 1.5 \cdot IQR = 54.10 \text{ მმ}.$$

შემდეგი ნახაზი კი გვიჩვენებს ბოქსპლოტის სურათს „გვერდი - გვერდ“.



ნახაზი 2.19. ბოქსპლოტი „გვერდი-გვერდ“ ტყვის შეღწევადობისთვის

მოცემული გრაფიკი გვიჩვენებს 200 მარცვლოვანი ტყვიების უფრო ღრმა და თანმიმდევრულ შეღწევადობას. იგი აგრეთვე გვიჩვენებს ერთი ექსტრემალური მონაცემის არსებობას და აქედან გამომდინარე, დახრილობას (გავიხსენოთ რას ნიშნავდა განაწილების ფორმის დახრილობა. გრაფიკები 2.11.) და ეს ყველაფერი საკმაოდ ლაკონურად და კომპაქტურადაა გაკეთებული. გარდა ამისა, შედარების მიზნით, ამ გრაფიკს კიდევ შეიძლება დაემატოს სხვა ბოქსპლოტებიც, (თუ გვექნება სხვა ტყვიების მონაცემებიც), ყოველგვარი ვიზუალური გადატვითვის გარეშე.

2.4. სტანდარტული რიცხვითი მახასიათებლები

ადგილმდებარეობის საზომები. ადამიანების უდიდესი ნაწილისთვის ცნობილია „საშუალო“, რომელიც წარმოადგენს მონაცემთა ცენტრს. მაგალითად, ბრძმედში ტემპერატურა შეიძლება ცვალებადობდეს ძალიან დიდ შუალედში, მაგრამ მათი

საშუალო მაინცგვეუბნება რაღაცას საშუალო ტემპერატურასა და იმ დროს წარმოდგენილ ტემპერატურებზე. გამოცდაზე მიღებული ქულებიც შეიძლება ძალიან ცვალებადობდეს, მაგრამ როგორც ვიცით, თითქმის ყველა სტუდენტის სურვილია, რომ მისი ქულები აჭარბებდეს სულ მცირე საშუალოს მაინც.

სიტყვა *საშუალო* შემოტანილია სასაუბრო ენიდან და ზოგჯერ სხვადასხვაგვარად გაიგება ხოლმე. ერთერთი არის მედიანა, რომელიც მონაცემებს ყოფს შუაზე. უხეშად რომ ვთქვათ, ეს არის იმ ფართობის ნახევარი, რომელიც წარმოდგენილია კარგად გაკეთებულ ჰისტოგრამაზე და მდებარეობს მედიანის რომელიმე მხარეს. როგორც ცენტრის საზომზე, კარგ წარმოდგენას იძლევა ამოვარდნები, ანუ ცოტა ექსტრემალური მნიშვნელობები ან ძალიან დაშორებული მონაცემები. მაგალითად განვიხილოთ, მონაცემთა პატარა სიმრავლე

2, 3, 6, 9, 10

მისი მედიანაა 6, და მისი მიშვნელობა დარჩება იგივე, 10 რომ გადავააგილოთ 10,000,000 ადგილზე და/ან 2 გადავიტანოთ -200,000 ადგილზე.

როგორც ვიცით, მედიანა არის მონაცემთა ცენტრალური მნიშვნელობა. თუმცა მხოლოდ ტექნიკური მხარე არ არის, სტატისტიკურ ანალიზში იგი ხშირად მიბმულია საშუალოზე. მართლაც, ძალიან ხშირად მის ნაცვლად განიხილება (არითმეტიკული) საშუალო.

განმარტება 2.5. რაოდენობრივი მონაცემებიდან ნებისმიერი შერჩევისთვის, (ვთქვათ, $x_1, x_2, \dots, x_n$) -სთვის (არითმეტიკული) საშუალო არის

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

საშუალოს ზოგჯერ უწოდებენ *პირველი რიგის მომენტს* ან განაწილების *მასების ცენტრს*. თუ წარმოვიდგენთ, რომ ერთეულოვანი მასები განლაგებულია რიცხვით ღერძზე მონაცემების ადილმდებარეობისა მნიშვნელობების მიხედვით, მაშინ მასების განაწილების წონასწორობის წერტილი იქნება $\bar{x}$.

გაფანტულობის საზომი. ცვლილების რაოდენობრივი მაჩვენებელი მონაცემთა სიმრავლის განლაგების ერთერთი მნიშვნელოვანი საზომია. განაწილების გაფანტულობის საზომი არის *რანგი*.

განმარტება 2.6. ვთქვათ, მოცემული გვაქვს ზრდის მიხედვით დალაგებული მონაცემთა სიმრავლე $x_1 \leq x_2 \leq \dots \leq x_n$. ამ მონაცემთა **რანგი** (ანუ გაბნევა, ანუ გაფანტულობა) არის

$$R = x_n - x_1$$

რაც რანგი დამოკიდებულია მონაცემების მხოლოდ უმცირეს და უდიდეს მნიშვნელობებზე, ამდენად მას გააჩნია მაღალი მგრძნობელობა ექსტრემალური მნიშვნელობების (ამოვარდნების) მიმართ. იმის გამო, რომ იგი ადვილად გამოთვლილია, ხშირად გამოიყენება დიდი პოპულაციისთვის, მაგალითად ინდუსტრიულ წარმოებაში, როგორც ხარისხის სტატისტიკური კონტროლის იარაღი.

თუმცა, სტატისტიკური დასკვნების გამოტანის მეთოდების უდიდესი ნაწილი ეყრდნობა განაწილების გაფანტულობის სხვა საზომებს. გამოსახულებების სახელები „გადახრის კვადრატის საშუალო“ ან „ფესვი გადახრის კვადრატის საშუალოდან“ გამოიყენებამ საზომების აღსანიშნავად, რომელთაც შესაბამისად ეწოდებათ დისპერსია და სტანდარტული გადახრა.

განმარტება 2.7. მონაცემთა სიმრავლის $x_1, x_2, \dots, x_n$ შერჩევის დისპერსია ეწოდება

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

შერჩევის სტანდარტული გადახრა, s, არის კვადრატული ფესვი შერჩევის დისპერსიიდან.

ფორმულაში $n-1$ გამყოფი შეიძლება შეიცვალოს n -ით, მაშინ s^2 იქნება ცენტრალური $\bar{x}$ მნიშვნელობიდან მონაცემთა წერტილებამდე მანძილების კვადრატების ჯამის საშუალო. ცხადია, s^2 არის არაუარყოფითი და 0 -ის ტოლია მხოლოდ მაშინ, როცა ყველა წერილი ერთი და იგივეა. ამასთან ადვილი მისახვედრია, რომ s^2 -ის

გამოსახვის ერთეულები კვადრატჯერ მეტია იმ ერთეულებზე, რომელშიც გამოსახულია საწყისი მონაცემთა სიმრავლე. ამიტომ კვადრატული ფესვის ამოღება s^2 -დან დაგვაბრუნებს ისევ იმ ერთეულებში, რა ერთეულებშიც იყო ორიგინალური მონაცემები.

სტატისტიკები და პარამეტრები. ახლა უკვე დროა შემოვიტანოთ ზოგიერთი მეტად მნიშვნელოვანი ტერმინი. შერჩევათა განაწილების ტერმინები ოდნავ განსხვავებულია პოპულაციის შესაბამისი ცნებებისგან.

შერჩევის მონაცემთა რიცხვით რეზიუმეებს უწოდებენ **სტატისტიკებს**, ხოლო პოპულაციის რიცხვით რეზიუმეებს კი **პარამეტრებს**. როგორც წესი სტატისტიკებს აღნიშნავენ ლათინური ასოებით, პარამეტრებს კი — ბერძნულით.

მაგალითად, გავიხილოთ საშუალო. *განმარტება 2.5* უშუალოდ ეხება შერჩევის საშუალოს გამოთვლას. მაგრამ თუ გვინდა, რომ საშუალო გამოვთვალოთ უშუალოდ მთელი პოპულაციისთვის, არსებობს საერთო შეთანხმება, რომ მას აღნიშნავენ ბერძნული პატარა ასოთი (μ) „მიუ“. ანუ **პოპულაციის საშუალო** გამოისახება ასე:

$$\bar{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

თუ ამ გამოსახულებას შევადარებთ განმარტება 2.5-ს, არა მარტო საშუალოს სიმბოლოა განსხვავებული. აქაც მიღებულია, რომ პოპულაციის ზომა აღინიშნებოდეს N -თ, ხოლო შერჩევის ზომა n -თ. შემდეგ თავებში მოცემულია თეორიული განაწილებების განმარტებები და იქაც μ სიმბოლო გამოიყენება იგივე კონტექსტში.

სხვა მაგალითი, იმის საილუსტრაციოდ, რაც შეეხება *განმარტება 2.7* -ს, შეიძლება განვიხილოთ დისპერსია და სტანდარტული გადახრა. *განმარტება 2.7* შეეხება კონკრეტულად *შერჩევის* დისპერსიას და სტანდარტულ გადახრას. თუ მონაცემთა სიმრავლე წარმოადგენს მთლიან პოპულაციას, აქაც შეთანხმებული აღნიშვნით, **პოპულაციის დისპერსია** აღინიშნება σ^2 („სიგმა კვადრატი“) და განმარტება ასე:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

კვადრატული ფესვი σ^2 -დან არის პოპულაციის სტანდარტული გადახრა. (ამ შემთხვევაში გამყოფი არის N და არა $N - 1$, როგორც ეს მოსალოდნელი იყო განმარტება 2.4 -დან. ამ ცვლილებას აქვს თავის მიზეზი, მაგრამამ ეტაპზე მისი ახსნა არ არის ხელმისაწვდომი). შემდეგ თავში შემოტანილია განმარტებები თეორიული განაწლების ვარიაციისა და სტანდარტული გადახრისთვისაც. აქაც კვლავ, σ^2 და σ გამოიყენება იგივე კონტექსტით.

მხოლოდ ერთ ადგილას გადავუხვევთ ლათინურ/ბერძნულ აღნიშვნებს, კვანტილის შემთხვევაში. $Q(p)$ აღნიშნავს p კვანტილს შერჩევაშიც, პოპულაციაშიც და თეორიულ მოდელშიც.

2.5 დამოკიდებულებები ცვლადებს შორის

ცხოვრებაში სტატისტიკური ამოცანების განხილვისას ხშირად გვხვდება ბივარიანტული მონაცემები. ამის მიზეზი არის ის, რომ ექსპერიმენტული რაოდენობრივი ცვლადი x ცვალებადობს რამდენიმე სხვადასხვა გარემოებების გამო, რომელიც თავის მხრივშესაბამისად იწვევს საპასუხო y ცვლადის სხვადასხვა შერჩევების წარმოშობას. ასეთ დროს x და y ცვლადების დამოკიდებულების აღწერისთვის ყველაზე მარტივი და სასურველია ისეთი დამკიდებულება, რომელიც კარგი მიახლოებით წრფივი განტოლებით აღიწერება. ეს უნდა გაკეთდეს მას შემდეგ რაც უკვე გვექნება გაფანტვის დიაგრამა.

უმცირეს კვადრატთა მეთოდი. ეს მეთოდი გამოიყენება წრფის მოსარგებად (x, y) მონაცემებზე. ეს შესაბამისობა დამყარებულია ე.წ. კორელაციის შეფასებაზე და განსაზღვრულობის კოეფიციენტზე.

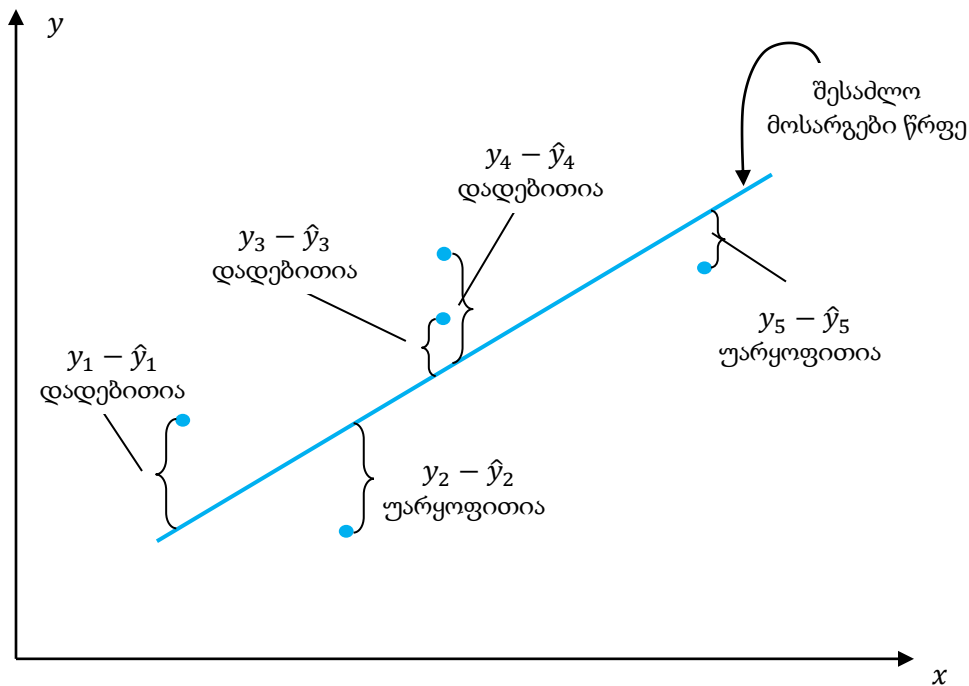
განმარტება 2.8. უმცირეს კვადრატთა მეთოდის გამოყენება განტოლების მორგებისთვის n წერტილიან მონაცემთა სიმრავლისთვის ნიშნავს, პარამეტრების

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

ისეთ შერჩევას, რომ მინიმუმადე იყოს დაყვანილი შემდეგი ჯამი:

სადაც, $y_1, y_2, \dots, y_n$ არის დაკვირვების შედეგად მიღებული საპასუხო ცვლადები, ხოლო $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ კი - შესაბამისი პროგნოზირებადი ან მოსარგები საპასუხო ცვლადები.

იმისათვის, რომ წრფე მოვარგოთ (x, y) მონაცემებს, მოყვანილი განმარტება წარმოგვიდგენს ამ რეკომენდაციებს. შევარჩიოთ წრფის დახრილობა და OY ღერძთან თანაკვეთა უმცირეს კვადრატთა მეთოდის გზით; ვერტიკალური მანძილების კვადრატთა ჯამი მონაცემთა წერტილებიდან წრფის შესაბამის წერტილებამდე, როგორც ეს ნაჩვენებია გრაფიკზე.



გრაფიკი 2.20. მონაცემთა ხუთი (x, y) წერტილი და შესაძლო მოსარგები წრფე

აქ მოცემულია ზოგადი სურათი მოგონილი ხუთი წერტილისთვის. (ეს არის კვადრატების ჯამი ხუთი მონიშნული წერტილის, შემდგომ სხვაობები უნდა შეიკრიბოს და შეირჩეს მინიმუმი).

დახრილობა და ღერძთან თანაკვეთა

გამოითვლება შემდეგი ფორმულებით:

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

და

$$b_0 = \bar{y} - b_1 \bar{x}$$

შერჩევის კორელაცია და განსაზღვრულობის კოეფიციენტი.

ვიზუალურად უმცირეს კვადრატთა მეთოდი კარგ სამსახურს გვიწევს წრფის მორგების საქმეში. თუმცა ამასთან ერთად კარგი იქნებოდა გვეჩვენა მეთოდი, რომელიც რიცხობრივად განსაზღვრავდა, თუ როგორაა მორგება განხორციელებული. ასეთ საზომს ეწოდება შერჩევის კორელაცია.

განმარტება 2.9. n წყვილიან (x_i, y_i) შერჩევაში x -ს და y -ს შორის შერჩევის (წრფივი) კორელაცია ეწოდება:

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

შერჩევის კორელაცია ყოველთვის მოთავსებულია -1 -სა და 1 -ს შორის. ამასთან -1 ან 1 -ის ტოლია იმ შემთხვევაში, თუ მონაცემის ყველა (x, y) წერტილი ვარდება უშუალოდ წრფეზე. ამ ფორმულების შედარება გვაძლევს, რომ

$$r = b_1(\sum(x_i - \bar{x})^2 / \sum(y_i - \bar{y})^2)^{1/2},$$

ანუ b_1 -ს და r -ს აქვს ერთი დ იგივე ნიშანი. ე.ი. შერჩევის კორელაცია ტოლია -1 -ის ნიშნავს, რომ x -ის ზრდას შეესაბამება y -ის პროპორციული (წრფივი) კლება. ხოლო შერჩევის კორელაცია ტოლია $+1$ -ის ნიშნავს, რომ x -ის ზრდას შეესაბამება y -ის პროპორციული (წრფივი) ზრდა.

რეალურად მონაცემები ხშირად არ არიან სრულად ($+1$ ან -1) კორელირებულნი. საზოგადოდ, r მოთავსებულია -1 -სა და 1 -ს შორის. მაგრამ თუ უნდათ, რომ ყურადღება მიაპყრონ ფაქტს იმას, თუ როგორია ამ მოვლენის ყოფაქცევა, ხალხი r -ს დებულობს როგორც *ამკარა წრფივი დამოკიდებულების ძალის ზომას*. r ახლოს არის

+1 -თან ან -1-თან გამოსახავს, თუ რამდენად მყარია წრფივი დამოკიდებულება. r ახლოს არის 0-თან გამოსახავს წრფივი დამოკიდებულების ნაკლებობას. r -ის ნიშანი კი გვიჩვენებს y -ის ზრდის ან კლების ტენდენციას, როდესაც იზრდება x .

განსაზღვრულობის კოეფიციენტი არის განტოლების მორგების სხვა რაოდენობრივი საზომი. ის შეიძლება გამოყენებულ იქნას არა მარტო წრფის მარტივი სიტუაციის შემთხვევაში, არამედ უფრო ფართო შემთხვევებშიც.

განმარტება 2.10. მონაცემთა n -წერტილიანი სიმრავლისთვის მორგებული განტოლების, რომელიც შედგენილია უმცირეს კვადრატთა მეთოდით და წარმოადგენს y -მორგებულ $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ მნიშვნელობებს, **განსაზღვრულობის კოეფიციენტი** ეწოდება:

$$R^2 = \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

R^2 შეიძლება გამოისახოს, როგორც წილადი y ნედლი ცვალებადობის შეფარდებისა მორგებულ განტოლებაზე, სადაც გათვალისწინებულია ეს ცვალებადობა. ასე, რომ მორგებული განტოლება შეიცავს მუდმივ წევრს, $\sum(y_i - \bar{y})^2 \geq \sum(y_i - \hat{y}_i)^2$. აქედან გამოდინარე, $\sum(y_i - \bar{y})^2$ არის საზომი y -ის ნედლი ცვალებადობისა, ხოლო $\sum(y_i - \hat{y}_i)^2$ არის მორგების შემდგომ y -ის დარჩენილი ნაწილის ცვალებადობა (ვარიაცია). მივიღეთ, რომ არაუარყოფითი სხვაობა $\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2$ არის მორგების - განტოლების პროცესის დროს y -ის ცვალებადობის საზომი. მაშინ R^2 გამოსახავს ამ განსხვავებას წილადის სახით (მთლიანი ნედლი ვარიაციის).

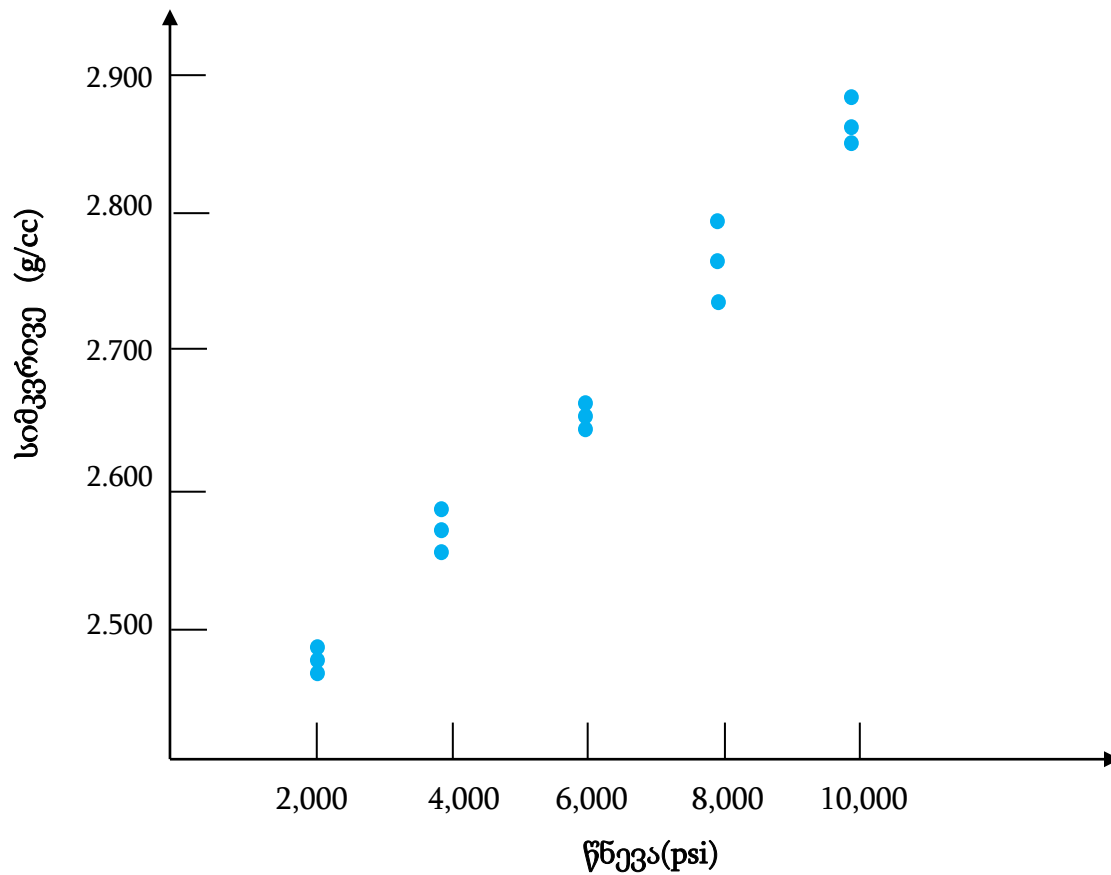
განსაზღვრულობის კოეფიციენტს გააჩნია მეორე სასარგებლო ინტერპრეტაცია. განტოლებებისთვის რომლებიც არიან წრფივები პარამეტრების მიმართ, აღმოჩნდა რომ R^2 არის კვადრატული კორელაცია. ანუ იგი არის კვადრატული კორელაცია დაკვირვებად y_i სიდიდეებსა და $\hat{y}_i$ მორგებულ სიდიდეებს შორის. (იმის გამო, რომ თავის მხრივ $\hat{y}_i$ სრულად (სრულყოფილად) კორელირებულია x_i სიდიდეებთან, აგრეთვე აღმოჩნდება, რომ R^2 არის კვადრატული კორელაცია y_i და x_i სიდიდეებს შორის.)

ვნახოთ, როგორ გამოითვლებიან ეს სიდიდეები კონკრეტული მონაცემთა სიმრავლისთვის. დავუშვათ, რომ ექსპერიმენტის საფუძველზე მიღებული გვაქვს სიმკვრივის წნევაზე დამოკიდებულების ასეთი მონაცემები. (საზომი ერთეულებია: psi - ფუნტი კვადრატულ დუიმზე, g/cc - გრამი/კუბურ სანტიმეტრზე).

x, წნევა	y, სიმკვრივე
2,000	2,486
2,000	2,479
2,000	2,472
4,000	2,558
4,000	2,570
4,000	2,580
6,000	2,646
6,000	2,657
6,000	2,653
8,000	2,724
8,000	2,774
8,000	2,808
10,000	2,861
10,000	2,879
10,000	2,858

ცხრილი 2.21. დაკვირვების მონაცემები

მის შესაბამის გაფანტულობის გრაფიკს აქვს ასეთი სახე:



გრაფიკი 2.22. *გაფანტულობის გრაფიკი, სიმკვრივე vs. წნევა*

ამ მონაცემებისთვის გამოვითვალთ სტატისტიკები:

$$\sum x_i = 2,000 + 2,000 + \dots + 10,000 = 90,000 ,$$

ამიტომ $\bar{x}_i = \frac{90,000}{15} = 6,000$.

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= (2,000 - 6,000)^2 + (2,000 - 6,000)^2 + \dots + \\ &+ (10,000 - 6,000)^2 = \mathbf{120,000,000} . \end{aligned}$$

$$\sum y_i = 2.486 + 2.479 + \dots + 2.858 = 40.005 ,$$

ამიტომ $\bar{y}_i = \frac{40.005}{15} = \mathbf{2.667}$.

$$\begin{aligned} \sum (y_i - \bar{y})^2 &= (2.486 - 2.667)^2 + (2.479 - 2.667)^2 + \dots + \\ &+ (2.858 - 2.667)^2 = \mathbf{.289366} . \end{aligned}$$

$$\begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= (2,000 - 6,000)(2.486 - 2.667) + \dots + \\ &+ (10,000 - 6,000)(2.858 - 2.667) = \mathbf{5,840} . \end{aligned}$$

მაშინ უმცირეს კვადრატთა მეთოდით დახრილობა და თანაკვეთა გამოითვლება ასე:

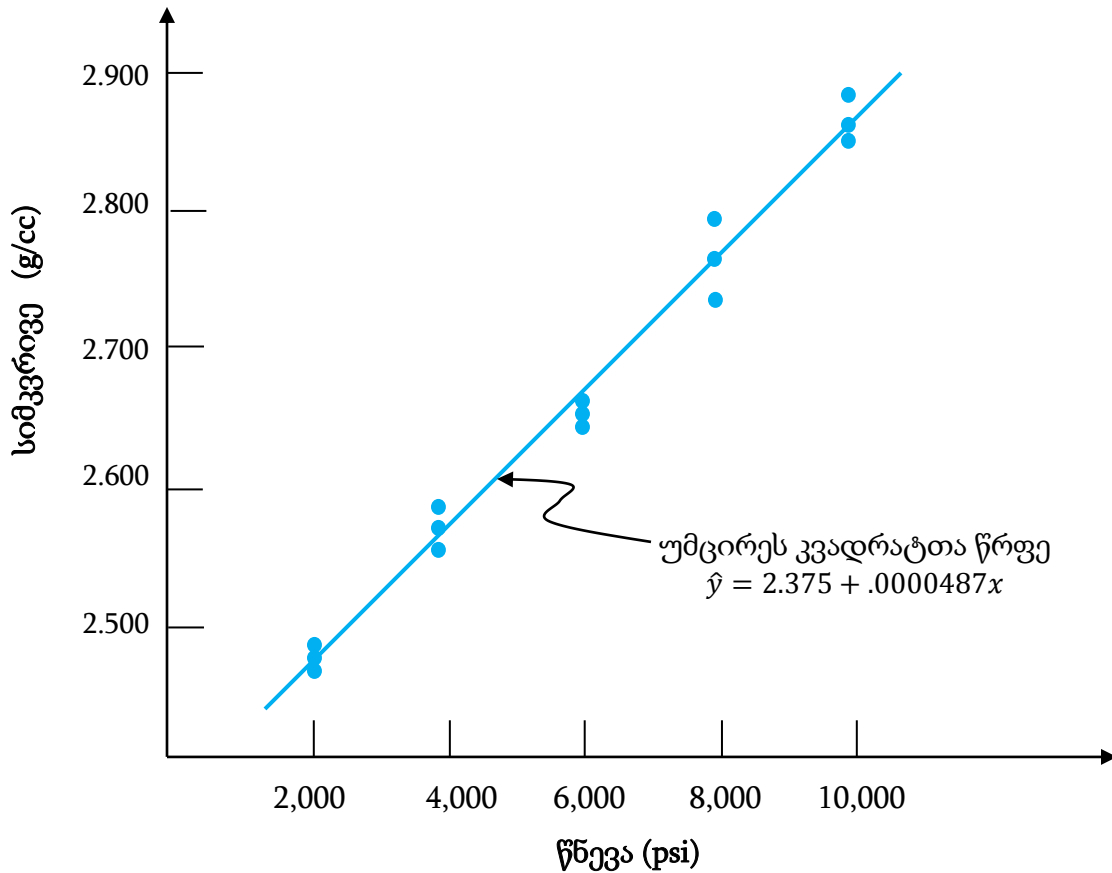
$$b_1 = \frac{5,840}{120,000,000} = .000048\bar{6}(g/cc)/psi$$

$$b_0 = 2.667 - (.000048\bar{6}) \cdot 6,000 = 2.375 g/cc$$

უმცირეს კვადრატთა წრფეს აქვს ასეთ სახე:

$$\hat{y} = 3.375 - .0000487x ,$$

რომელიც გამოსახულია შემდეგ გრაფიკზე.



გრაფიკი 2.23. გაფანტულობის გრაფიკი წნევა/სიმკვრივე მონაცემებით და უმცირეს კვადრატთა წრფე.

შევნიშნოთ, რომ დახრილობა $b_1 \approx .0000487 \text{ g/cc}$ აქვს ფიზიკური შინაარსი, რომ (დაახლოებით) ყოფილიყო.

განმარტება 2.5 -ში ნაჩვენებია უმცირეს კვადრატთა მეთოდში მონაწილე y , რომელიც შეესაბამება x -ს, შეიძლება განისაზღვროს როგორც **მორგებული ან პროგნოზირებადი სიდიდე**. ამის გამო შესაძლებელი უნდა იყოს y -ის გამოსახვა იმისდა მიხედვით, თუ როგორია x -ის ყოფაცევა. ნათქვამი უფრო გასაგები რომ გახდეს, განვიხილოთ ასეთი ამოცანა: ზემოთ მოყვანილ მონაცემებში გვაქვს სიმკვრივე, რომელიც შეესაბამება მაგალითად $4,000(\text{psi})$ და ცხრილზე დაკვირვებით განვსაზღვროთ სიმკვრივე, რომელიც შეიძლება რომ შეესაბამებოდეს $5,000(\text{psi})$, რომელიც არ გვაქვს ცხრილში. პირველი, რაც თვალში საცემია, მნიშვნელობა $x = 4,000$ ცხრილში გვხვდება სამჯერ და ბუნებრივია, რომ y მნიშვნელობა ასე განვსაზღვროთ,

$$\bar{y} = \frac{1}{3}(2.558 + 2.570 + 2.580) = 2.5693 \text{ g/cc}$$

და გამოვიყენოთ ეს მნიშვნელობა. მაგრამ თუ დავუშვებთ, რომ y დაახლოებით წრფივადაა დამოკიდებული მაშინ მორგებული სიდიდე

$$\hat{y} = 2.375 + .0000486 \cdot 4,000 = 2.5697 g/cc$$

შეიძლება იყოს უკეთესი ვიდრე საშუალო სიმკვრივე $x = 4,000$ -სთვის. შევხედოთ ახლა სიტუაციას, როცა $x = 5,000$ -ს. აქ უკვე ცხრილში აღარ გვაქვს შესაბამისი მონაცემი. ერთადერთი რაზეც შეიძლება გაჩნდეს კითხვა, არის ის რომ ინტერპოლაციით მიღებული მნიშვნელობა იქნება თუ არა აზრიანი ფიზიკური თვალსაზრისით. ასე, რომ მორგებული მნიშვნელობა

$$\hat{y} = 2.375 + .0000486 \cdot 5,000 = 2.6183 g/cc$$

შეიძლება გამოვიყენოთ როგორც სიმკვრივე, რომელიც შეესაბამება 5,000 (psi) წნევის მნიშვნელობას.

ამ მონაცემებისთვის შეგვიძლია გამოვითვალოთ კორელაცია:

$$r = \frac{5,840}{\sqrt{120,000,000 \cdot (.289366)}} = .9911 .$$

ეს მნიშვნელობა ახლოსაა +1 -თან და განსაზღვრავს მყარ დადებით წრფივ დამოკიდებულებას, როგორც ეს ჩანს გრაფიკებზე 2.22 და 2.23.

თუ გამოვიყენებთ მორგების წრფეს, ჩვენ შეგვიძლია ვიპოვოთ $\hat{y}$ ყველა $n = 15$ მონაცემთა წერტილისთვის მონაცემთა ორიგინალური სიმრავლიდან. ეს მონაცემები მოცემულია ქვედა ნახაზზე.

x , წნევა	y , მორგებული სიმკვრივე
2,000	2,4723
4,000	2,5697
6,000	2,6670
8,000	2,7643
10,000	2,8617

ცხრილი 2.24. მორგებული სიმკვრივის მნიშვნელობები

კიდევ ერთხელ მხედველობაში მივიღებთ ცხრილი 2.21 -ს და გვექნება რომ:

$$\begin{aligned}\sum (y_i - \hat{y}_i)^2 &= (2.486 - 2.4723)^2 + (2.479 - 2.4723)^2 + (2.472 - 2.4723)^2 + \\ &+ (2.588 - 2.5697)^2 + \dots + (2.879 - 2.8617)^2 + \\ &+ (2.858 - 2.8617)^2 = .005153\end{aligned}$$

რადგანაც $\sum (y_i - \bar{y})^2 = .289366$,

ბოლოს მივიღებთ, რომ

$$R^2 = \frac{.289366 - .005153}{.289366} = .9822$$

და მორგებული წრფის შეფასება გამოდის, რომ 98% -ზე მეტი სიზუსტითაა ახლოს ნედლ ცვლილებებთან და ამცირებს „აუხსნელ“ ცვლილებებს .289366-დან .005153-მდე.

მონცემებისთვის *წნევა/სიმკვრივე* კორელაცია x -სა და y -ს შორის არის

$$r = .9911$$

რადგანაც $\hat{y}$ სავსებით კარგადაა კორელირებული x -თან, ეს ნიშნავს რომ ასევე კარგადაა კორელირებული $\hat{y}$ და y .

$$r^2 = .9911^2 = .9822 = R^2$$

და R^2 მართლაც არის შერჩევითი კორელაციის კვადრატი y -სა და $\hat{y}$ -ს შორის.

სავარჯიშოები:

1. ააგეთ ზურგი - ზურგთან ღეროვან-ფოთლოვანი დიაგრამა ორი შერჩევისთვის, რომელიც მოცემულია ცხრილი 2.3-თ (ტყვიის შეღწევადობის სიღრმე).
2. ქვემოთ მოყვანილია საიზოლაციო ლითონის ფურცლის ორი სხვადასხვა გზით გასაჭრელად საჭირო ძალის მონაცემები (ft lb).

სიგრძივი დაჭრები	ჯვარედინი დაჭრები
1.15	.89
.84	.69
.88	.46
.91	.85
.86	.75
.88	.67
.92	.78
.87	.77
.93	.80
.95	.79

ა) შეადგინეთ კვანტილების წერტილოვანი გრაფიკი ორი შერჩევისთვის. იპოვეთ მედიანები, კვარტილები, და .37 კვანტილები მონაცემთა ორივე სიმრავლისთვის.

ბ) დახაზეთ ბოქსპლოტები გვერდი-გვერდ, გაჭრის ორი მეთოდის შედარების მიზნით. იმსჯელეთ რას გვიჩვენებს ეს გრაფიკი ორი მეთოდის შესახებ.

- წინა სავარჯიშოს ორივე მონაცემებისთვის გამოთვალეთ და შეადარეთ საშუალოები, რანგები, კვარტილთშორისი რანგი და სტანდარტული გადახრა.
- წინა სავარჯიშოში თქვენს მიერ მიღებული რიცხვები სტატისტიკებია თუ პარამეტრები? ახსენით.
- სიგრძივი დაჭრის ყველა მონაცემს დაუმატეთ 1.3 და გამოთვალეთ თავიდანსაშუალო, მედიანა, რანგი, კვარტილთშორისი რანგი და სტანდარტული გადახრა. რას გვამღევენ მათი შედარებები წინა მიღებულ შედეგთან? გაიმეორეთ იგივე სავარჯიშო მაგრამ ახლა ყველა მონაცემი გაამრავლეთ 2-ზე, 1.3-ის დამატების მაგიერ. ისევ შეადარეთ!
- მოცემულია ხელოვნური მონაცემების მცირე სიმრავლე.

x	1	2	3	4	5
y	8	8	6	6	4

გამოთვლები ჩაატარეთ აუცილებლად ხელით და უპასუხეთ შემდეგ კითხვებს:

(ა) მონაცემების საშუალებით მიიღეთ უმცირეს კვადრატთა წრფე. დახაზეთ გაფანტვის დიაგრამა და დახაზეთ ეს წრფე ამავე დიაგრამაზე.

(ბ) მიიღეთ შერჩევის კორელაცია x -სა და y -ს შორის.

(გ) მიიღეთ შერჩევის კორელაცია r -სა და y -ს შორის და შეადარეთ (ბ) პუნქტში მიღებულ პასუხს.

(დ) გამოიყენეთ შესაბამისი ფორმულა და გამოთვალეთ ამ მონაცემებისთვის R^2 . შეადარეთ ის (ბ) და (გ) პუნქტებში მიღებული პასუხების კვადრატებს.



3. დისკრეტული შემთხვევითი სიდიდე.

ალბათობის თეორია არის შემთხვევითი ცვალებადობის მათემატიკური აღწერა. ამ ნაწილში შევეცდებით შემოვიტანოთ ალბათობის ცნებებისა და დებულებების ის მინიმალური რაოდენობა, რომლებიც საკმარისი იქნება სტატისტიკური დასკვნების გასაკეთებლად.

დასაწყისში ჯერ განვიხილავთ დისკრეტულ შემთხვევით სიდიდეებს და მათი განაწილების კანონებს, შემდეგ უწყვეტ შემთხვევით სიდიდეებს, ალბათობებს, შემთხვევითი ცვლადის ფუნქციებს, შემთხვევითი მოვლენების ურთიერთდამოკიდებულებებს და სხვა.

3.1. შემთხვევითი სიდიდეები და მათი განაწილებები.

განმარტება 3.1. შემთხვევითი სიდიდე (ან შემთხვევითი ცვლადი) არის სიდიდე, რომელიც მიიღება პირველადი დაკვირვების საფუძველზე და დამოკიდებულია მოვლენების შესაძლო შედეგებზე. მათ ზოგადად აღნიშნავენ ლათინური ანბანის ბოლო ასოებით.

ეს ცნება თავისთავად მოიცავს ორ განსხვავებულ ცნებას, რომლებიც მომდინარეობს დისკრეტული და უწყვეტი მონაცემებიდან, შესაბამისად, ეწოდებათ დისკრეტული შემთხვევითი სიდიდე და უწყვეტი შემთხვევითი სიდიდე.

განმარტება 3.2. დისკრეტული შემთხვევითი სიდიდე არის ისეთი შემთხვევითი სიდიდე, რომელსაც გააჩნია მხოლოდ იზოლირებული ანუ განცალგებული შესაძლო მნიშვნელობები (და არა შედეგების კონტინუუმ სიმრავლე).

განმარტება 3.3. უწყვეტი შემთხვევითი სიდიდე არის ისეთი შემთხვევითი სიდიდე, რომლის შესაძლო მნიშვნელობები მთლიანად ავსებენ რიცხვით ინტერვალს, ანუ მნიშვნელობების სიმრავლე წარმოადგენს ინტერვალს.

შემთხვევითი სიდიდე, რომლის შედეგები შესაძლებელია რომ გადაითვალოს, არის დისკრეტული სიდიდე. თუმცა ძალიან ხშირად მათემატიკურად უფრო მოსახერხებელია და პრაქტიკული მიზნებისთვისაც ადექვატურია უწყვეტი შემთხვევითი

სიდიდის გამოყენება. შემთხვევითი სიდიდე არის გარკვეულწილად არაპროგნოზირებადი. ამიტომაც, მისი აღწერისას ან მოდელირების დროს მეტად მნიშვნელოვანია დავაკონკრეტოთ მისი შესაძლო მნიშვნელობების სიმრავლე და ამ შესაძლო მნიშვნელობებთან დავაკავშიროთ მათი შესაბამისი ალბათობები.

განმარტება 3.4. იმისათვის, რომ განვსაზღვროთ ალბათობის განაწილება, უნდა გვეჩვენოს შესაძლო მნიშვნელობათა სიმრავლე და შემდეგ (ამა თუ იმ გზით) მიმდევრობით ყველა მათგანს შევუსაბამოთ რიცხვები 0 -დან 1 -მდე, რომელთაც ეწოდებათ ალბათობები და წარმოადგენს საზომს იმისას, თუ რამდენჯერ გამოჩნდნენ სხვადასხვა რიცხვითი მნიშვნელობები.

დისკრეტული და უწყვეტი შემთხვევითი სიდიდეებისთვის ალბათობების განაწილების აგების მეთოდები განსხვავებულია. ამიტომ, ჩვენ მათ შევისწავლით ორ ეტაპად. ჯერ დავიწყებთ დისკრეტული განაწილებით.

3.2. დისკრეტული ალბათობის ფუნქცია და დაგროვებითი ალბათობის ფუნქცია.

ყველაზე ხშირად გამოყენებადი ინსტრუმენტი დისკრეტული შემთხვევითი სიდიდის ალბათობის განაწილების აღწერისთვის არის ალბათობის ფუნქცია.

განმარტება 3.5. X დისკრეტული შემთხვევითი სიდიდის (ცვლადის) ალბათობის ფუნქცია არის ცვლადის ყოველი შესაძლო $x_1, x_2, \dots$ მნიშვნელობებისთვის განსაზღვრული არაუარყოფითი $f(x)$ ფუნქცია, სადაც თითოეული $f(x_i)$ მნიშვნელობა წარმოადგენს X დისკრეტული შემთხვევითი სიდიდის, x_i შემთხვევითი მნიშვნელობის ალბათობას.

შემდგომში P ასოთი აღვნიშნავთ ალბათობას იმ გამონათქვამისას, რომელიც მოთავსებული იქნება ფრჩხილებში. წაიკითხება ასე: „ალბათობა იმისა, რომ...“ და შემდეგ ფრჩხილებში ჩაწერილი გამოსახულება. ამ აღნიშვნებში ალბათობის ფუნქცია ასე შეიძლება ჩაიწეროს:

$$f(x) = P[X = x]$$

დისკრეტული შემთხვევითი სიდიდის ალბათობის ფუნქცია

წავიკითხება ასე: „ $f(x)$ არის ალბათობა იმისა, რომ (შემთხვევითი სიდიდე) X ღებულობს მნიშვნელობას x .“

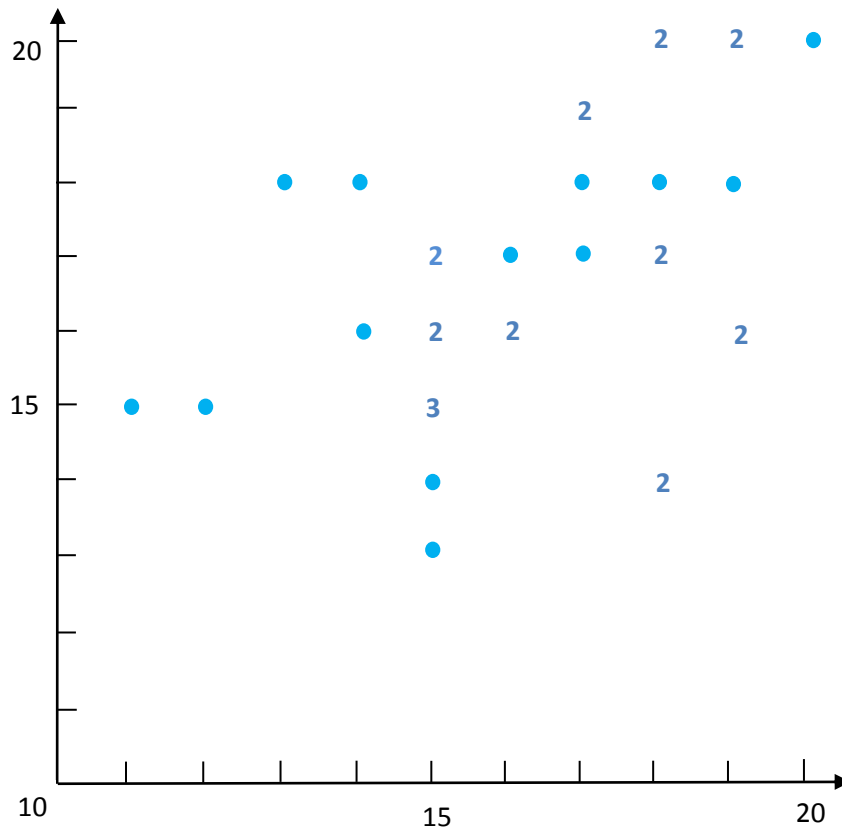
მაგალითი 3.1. ჭანჭიკის მოჭერის ბრუნვის მომენტი.

სტუდენტებმა: ბრინმა, ქრისტენსენმა და შნაიდერმა გაზომეს მძიმე ტექნიკის კომპონენტის წინა პანელზე არსებული 6 ჭანჭიკისთვის ბრუნვის მომენტები, რომლებიც საჭიროა მათ მოსახსნელად. 34 სხვადასხვა კომპონენტისთვის #3 და #4 ჭანჭიკებისთვის მონაცემები მოყვანილია შემდეგ ცხრილში.

ჭანჭიკი 3			ჭანჭიკი 4		
კომპონენტი	ბრუნვის მომენტი	ბრუნვის მომენტი	კომპონენტი	ბრუნვის მომენტი	ბრუნვის მომენტი
1	16	16	18	15	14
2	15	16	19	17	17
3	15	17	20	14	16
4	15	16	21	17	18
5	20	20	22	19	16
6	19	16	23	19	18
7	19	20	24	19	20
8	17	19	25	15	15
9	15	15	26	12	15
10	11	15	27	18	20
11	17	19	28	13	18
12	18	17	29	14	18
13	18	14	30	18	18
14	15	15	31	18	14
15	18	17	32	15	13
16	15	17	33	16	17
17	18	20	34	16	16

ცხრილი 3.1. ბრუნვის მომენტები, რომლებიც საჭიროა ჭანჭიკების მოსახსნელად. (საზომი ერთეულია (ftlb)).

მის შესაბამის გაბნევის დიაგრამას ექნება ასეთი სახე.



გრაფიკი 3.2. გაზნევის დიაგრამა #3 და #4 ჭანჭიკებისთვის.

შემოვიღოთ აღნიშვნა:

Z = ბრუნვის მომენტის შედეგი გაზომვის შედეგი #3 ჭანჭიკისთვის.

(დამრგვალებულია უახლოეს მთელ რიცხვამდე).

ჩავთვალოთ Z დისკრეტულ შემთხვევით სიდიდედ და ავაგოთ მისთვის ალბათობის ფუნქცია.

დავალება: დამოუკიდებლად გააკეთეთ ყველაფერი ანალოგიურად ჭანჭიკი #4 -სთვისაც!

ჭანჭიკი #3 -სთვის ვღებულობთ ფარდობით სიხშირეთა ასეთ განაწილებას:

z, ბრუნვის მომენტი(ft lb)	სიხშირე	ფარდობითი სიხშირე
11	1	$1/34 \approx .02941$
12	1	$1/34 \approx .02941$
13	1	$1/34 \approx .02941$
14	2	$2/34 \approx .05882$
15	9	$9/34 \approx .26471$
16	3	$3/34 \approx .08824$
17	4	$4/34 \approx .11765$
18	7	$7/34 \approx .20588$
19	5	$5/34 \approx .14706$
20	1	$1/34 \approx .02941$
34		1

ცხრილი 3.3. ფარდობით სიხშირეთა განაწილება.

ეს ცხრილი გვიჩვენებს, მაგალითად, რომ სტუდენტების მიერ მიღებული გაზომვებით 19 ft lb დაფიქსირდა გაზომვათა რაოდენობის 15% -ში. უპრიანი იქნება დავეყრდნოთ ცხრილი 3.4. -ის მონაცემებსაც, რადგან ისიც მოგვცემს წარმოდგენას Z - სთვის ალბათობის ფუნქციის აგებაზე, რომელის გაკეთებასაც ჩვენ შევეცდებით მიღებული ცხრილების დახმარებით. ანუ სხვანაირად რომ ვთქვათ, ალბათობის განაწილება შეიძლება განისაზღვროს შემდეგი ცხრილის საშუალებით.

ბრუნვის მომენტი	ალბათობა
Z	$f(z)$
11	.03
12	.03
13	.03
14	.06
15	.26
16	.09
17	.12
18	.20
19	.15
20	.03

ცხრილი 3.4. Z-ის ალბათობის ფუნქცია.

ფარდობით სიხშირეთა ცხრილიდან $f(z)$ -სთვის შემოთავაზებული მნიშვნელობა ალბათობის ფუნქციის ცხრილში შესაძლებელი იყო მიგვეღო ფარდობით სიხშირეთა ცხრილის მნიშვნელობათა ოდნავ დამრგვალებით. ეს ასე მოხდა კიდევ მეასედის სიზუსტით, მაგრამ გვახსოვდეს, დამრგვალება ყოველთვის ისე უნდა გაკეთდეს, რომ ალბათობების ჯამი უნდა გამოვიდეს ზუსტად 1.00.

აუცილებელია აგრეთვე, რომ $f(z)$ -ის მნიშვნელობები მოთავსებული იყვნენ $[0; 1]$ - შუალედში. უარყოფით ალბათობას და ერთზე მეტ ალბათობას არანაირი პრაქტიკული აზრი არ გააჩნია. ალბათობა ტოლია ერთის ნიშნავს, რომ მოვლენას რომელსაც ვაკვირდებით აუცილებლად მოხდება, ხოლო 0-ის ტოლია ნიშნავს, რომ მოვლენას არასდროს არა აქვს ადგილი. ის ფაქტი, რომ $f(z)$ -ის მნიშვნელობათა ჯამი უდრის 1 -ს, ნიშნავს რომ მნიშვნელობებიდან 11,12,13, 14, 15, 16, 17, 18, 19 და 20 (ft lb) ერთერთს აუცილებლად აქვს ადგილი.

ალბათობის ფუნქცია $f(x)$ გვაძლევს ალბათობას თითოეული ინდივიდუალური მნიშვნელობისთვის და აგრეთვე, რაღაც ნიშნით განსაზღვრული ელემენტების ერთობლიობისთვისაც X -დან. ამ უკანასკნელი შემთხვევის დროს უნდა შევკრიბოთ შერჩეული ელემენტების შესაბამისი ალბათობები.

მოდით, დავუბრუნდეთ ჭანჭიკების მაგალითს გამოვიყენოთ Z -ისალბათობის ფუნქციის ცხრილი და გამოვთვალოთ ასეთი ალბათობა:

$$P[Z > 17] = P[\text{შემდეგი ბრუნვის მომენტი მეტია 17 -ზე }].$$

ამისთვის, როგორც ვთქვით, უნდა შევკრიბოთ $f(z)$ -ის ყველა ის მნიშვნელობა, რომელიც შეესაბამება იმ შესაძლო მნიშვნელობებს, რომლებიც მეტია 17 (ft lb) -ზე,

$$P[Z > 17] = f(18) + f(19) + f(20) = .20 + .15 + .03 = .38,$$

ანუ, ალბათობა იმისა, რომ შემდეგი ბრუნვის მომენტი მეტია 17 -ზე, არის დაახლოებით 38%.

თუ მაგალითად, გვინდა, რომ გამოვთვალოთ ალბათობა ბრუნვითი მომენტებისა 16 ft lb -დან 21 ft lb -მდე, მაშინ ვიქცევით ასე:

$$\begin{aligned} P[16 \leq Z \leq 21] &= f(16) + f(17) + f(19) + f(20) + f(21) = \\ &= .09 + .12 + .20 + .15 + .03 + .00 = .59, \end{aligned}$$

შევნიშნოთ, რომ ამ შემთხვევაში $f(x)$ ფუნქცია მოცემული იყო ცხრილის სახით. სხვა შემთხვევაში ფუნქცია $f(x)$ შეიძლება მოცემული იყოს ფორმული სახით.

ზოგჯერ გამოიყენება დისკრეტული შემთხვევითი სიდიდის ალბათობის განაწილების შედგენის სხვა გზაც, რომელის არსიც მდგომარეობს იმაში, რომ მივუთითოთ (ან ავაგოთ) ე.წ. *დაგროვებითი ალბათობის ფუნქცია*. (ქართულ ენაზე არსებულ სახელმძღვანელოებში ამ ფუნქციას *განაწილების ფუნქციის* სახელით მოიხსენიებენ).

განმარტება 3.6. X დისკრეტული შემთხვევითი სიდიდის *დაგროვებითი ალბათობის ფუნქცია* ეწოდება ისეთ $F(x)$ ფუნქციას, რომელიც ყოველი x რიცხვისთვის განისაზღვრება როგორც ალბათობა იმისა, რომ X -ის მნიშვნელობები არ აღემატება x -ს. ეს ყველაფერი ფორმულით ასე გამოისახება:

$$F(x) = P[X \leq x]$$

რადგანაც დისკრეტული შემთხვევითი სიდიდისთვის ალბათობები გამოითვლება როგორც $F(x)$ მნიშვნელობათა ჯამი, ამიტომ გვექნება:

$$F(x) = \sum_{z \leq x} f(z)$$

X დისკრეტული შემთხვევითი სიდიდის დაგროვებითი ალბათობის ფუნქცია

(ჯამი აიღება ყველა იმ შესაძლო მნიშვნელობების მიმართ, რომლებიც არ აღემატებიან x -ს.) ამ დისკრეტულ შემთხვევაში $F(x)$ ფუნქციის გრაფიკი იქნება საფეხურა ტიპის ფუნქცია, რომელსაც ექნება ნახტომები იმ წერტილებში, რომლებიც წარმოადგენენ შესაძლო მნიშვნელობებს (ანუ შემთხვევით მნიშვნელობებს). ნახტომის სიდიდე ტოლია იმ ალბათობის, რომელიც დაკავშირებულია შესაბამის შესაძლო მნიშვნელობასთან.

მაგალითი 3.1. (გაგრძელება).

ორივე ფუნქციის, ალბათობის ფუნქციისა და დაგროვებითი ალბათობის ფუნქციის მნიშვნელობები ბრუნვითი მომენტის Z ცვლადისთვის მოცემულია შემდეგ ცხრილში. $F(z)$ -ის მნიშვნელობები სხვა z - ებისთვისაც ადვილად გამოითვლება. მაგალითად,

$$F(10.7) = P[Z \leq 10.7] = 0.$$

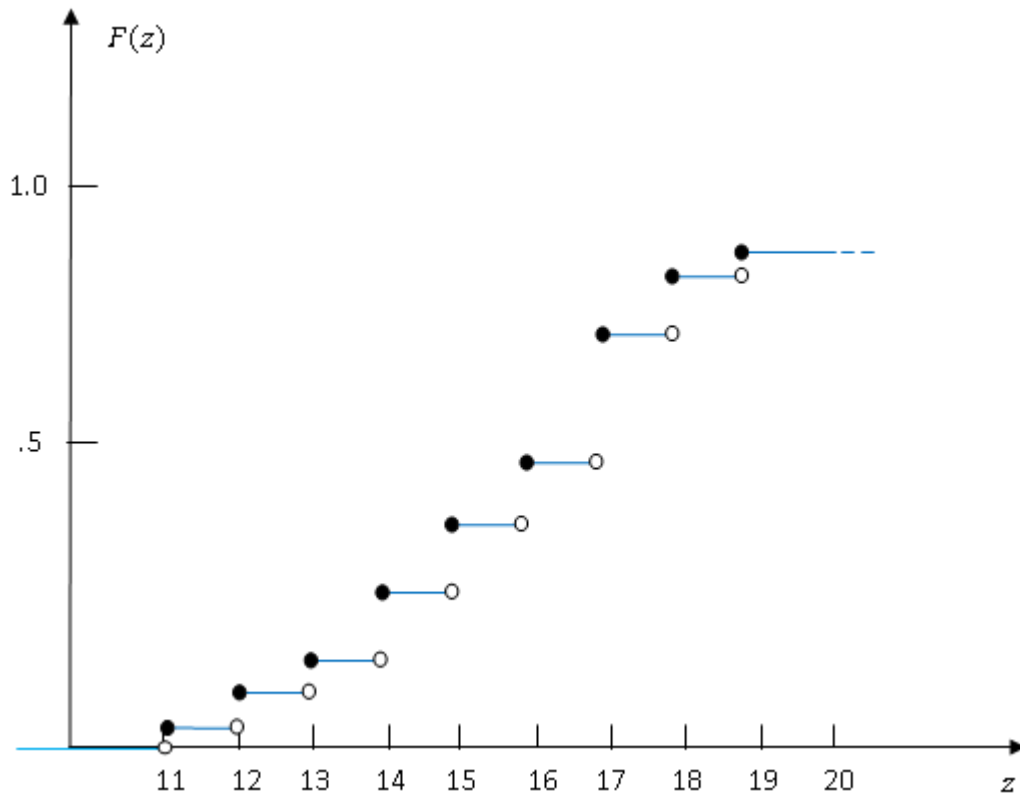
$$F(16.3) = P[Z \leq 16.3] = P[Z \leq 16] = F(16) = 0.50.$$

$$F(32) = P[Z \leq 32] = 1.00$$

z, ბრუნვის მომენტი	$f(z) = P[Z = z]$	$F(z) = P[Z \leq z]$
11	.03	.03
12	.03	.06
13	.03	.09
14	.06	.15
15	.26	.41
16	.09	.50
17	.12	.62
18	.20	.82
19	.15	.97
20	.03	1.00

ცხრილი 3.5. ალბათობის ფუნქციისა და დაგროვებითი ალბათობის ფუნქციის მნიშვნელობები Z -სთვის.

Z-ის დაგროვებითი ალბათობის ფუნქციის გრაფიკი კი როგორც უკვე ვთქვით, წარმოადგენს საფეხურა ფორმის ფუნქციას რაც დამახასიათებელია დისკრეტული შემთხვევითი სიდიდის დაგროვებითი ალბათობის ფუნქციებისთვის.



გრაფიკი 3.6. *z -ის დაგროვებითი ალბათობის ფუნქციის გრაფიკი.*

ინფორმაციას დისკრეტულ განაწილებაზე გვაძლევს დაგროვებითი ალბათობის ფუნქცია, რაც ექვივალენტურია იმის, რომ ეს ხორციელდება შესაბამისი ალბათობის ფუნქციის საშუალებით. ცხრილის შესადგენად ხანდახან დაგროვებით ვერსიას ამჯობინებენ ხოლმე, რადგანაც დამრგვალების პრობლემა უფრო მძიმეა როცა გვიწევს რამდენიმე $f(x)$ წევრის შეკრება, ვიდრე ის, რომ $F(x)$ -ის ორ მნიშვნელობას შორის სხვაობა რომ შევუსაბამოთ შესაძლო მნიშვნელობებს, როგორც ალბათობა.

3.3. დისკრეტული ალბათობის განაწილების შეჯამება.

თუ დისკრეტული შემთხვევითი სიდიდისთვის მნიშვნელობები თანაბარი მანძილითაა დაშორებული ერთმანეთისგან, მაშინ ალბათობის ჰისტოგრამა გვაძლევს სურათს, რომელსაც აქვს ცვლადების განაწილების ფორმა. იგი მიიღება ცენტრირებული სვეტოვანი დიაგრამის საშუალებით, ანუ იმ მართკუთხედებით, რომელთა სიმაღლეებია $f(x)$ მნიშვნელობები ყოველი შესაძლო მნიშვნელობების თავზე და ეს შესაძლო მნიშვნელობა წარმოადგენს მართკუთხედის ფუძის შუა წერტილს (ცენტრს). ასეთი ჰისტოგრამების ინტერპრეტაცია მსგავსია ფარდობითი სიხშირის ჰისტოგრამის ინტერპრეტაციის, იმ მცირე განსხვავებით, რომ ემპირიულად მიღებული ფარდობითი სიხშირის წილადების ნაცვლად წერია თეორიული ალბათობები.

შემდგომი ქმედებებისთვის სასარგებლოა, თუ დისკრეტული შემთხვევითი სიდიდის საშუალოსთვის (ანუ მისი ალბათობის განაწილებისთვის) შემოვიღებთ სპეციალურ აღნიშვნას. შემოვიღოთ ასეთი განმარტება.

განმარტება 3.7. X დისკრეტული შემთხვევითი სიდიდის საშუალო ან მოსალოდნელი მნიშვნელობა ან მათემატიკური ლოდინი (ზოგჯერ მას ალბათობის განაწილების საშუალოსაც უწოდებენ) ეწოდება რიცხვს

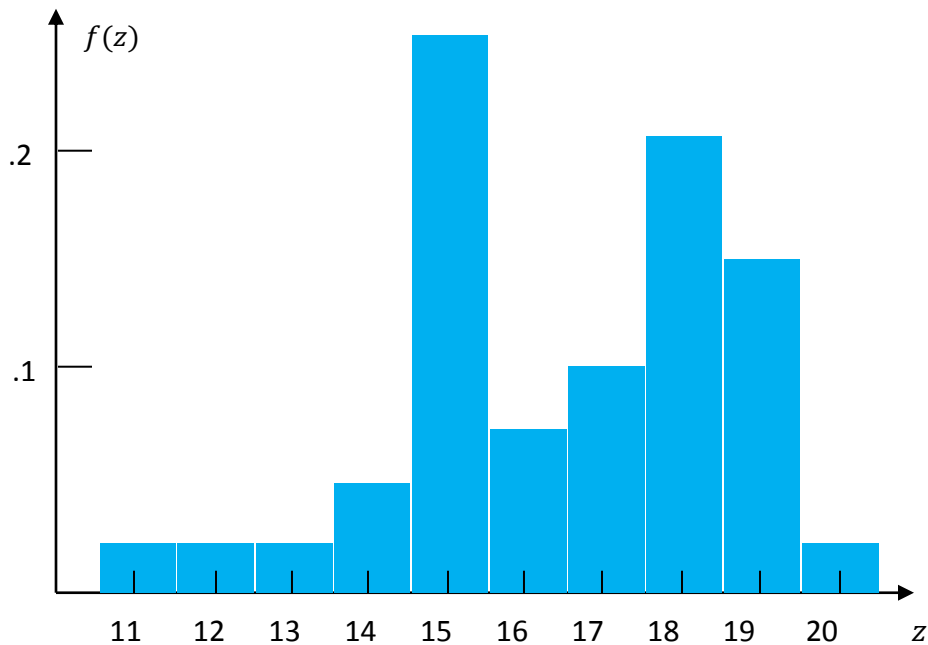
$$EX = \sum_x xf(x)$$

(3.1)

EX იკითხება ასე: „ X -ის მოსალოდნელი მნიშვნელობა ტოლია..“ ან „ X -ის მათემატიკური ლოდინი ტოლია...“ ზოგჯერ EX აღნიშვნის მაგიერ იყენებენ ბერძნულ ასოს μ -ს.

z შემთხვევითი ცვლადის ალბათობის ჰისტოგრამას კი აქვს ასეთი სახე:

Z-ის ალბათობის განაწილება



გრაფიკი 3.7. *Z-ის ალბათობის სპისტოგრამა (მაგალითი 1).*

მაგალითი 3.1. (გაგრძელება).

თუ დავუბრუნდებით ჭანჭიკის მოჭერის ბრუნვის მომენტის მაგალითს, მოსალოდნელი მნიშვნელობა (ანუ თეორიული საშუალო) ტოლია:

$$\begin{aligned}
 EZ &= \sum_z z f(z) = \\
 &= 11(.03) + 12(.03) + 13(.03) + 14(.06) + 15(.26) + \\
 &+ 16(.09) + 17(.12) + 18(.20) + 19(.15) + 20(.03) = \\
 &= 16.35 \text{ ftlb}
 \end{aligned}$$

ეს მნიშვნელობა მიღებულია ცხრილი 3.4 -ის მიხედვით. ამ შემთხვევაში EZ აღნიშვნის მაგიერ ცხადია, შეგვიძლია გამოვიყენოთ სიმბოლო μ .

დისკრეტული შემთხვევითი სიდიდის ალბათობის განაწილების საშუალოს აქვს წონასწორობის წერტილის შინაარსი, ისევე როგორც მონაცემთა სიმრავლის საშუალო არითმეტიკულს. თუ რიცხვით ღერძზე თითოეული შემთხვევითი მნიშვნელობის x წერტილზე განვათავსებთ მასებს, რომელთა სიდიდეები შესამისად არის $f(x)$ მნიშვნელობების ტოლი, მაშინ EX იქნება ამ განაწილების მასების ცენტრი.

(გავიხსენოთ, რომ ყველაფერს ვაკეთებთ ზუსტად ასევე ჭანჭიკი 4 -ის შემთხვევაშიც. დამოუკიდებლად!)

როგორც ცნობილია, სტატისტიკაში მნიშვნელოვანია მონაცემთა სიმრავლის (ან ფარდობით სიხშირეთა განაწილების) გაფანტულობის ანუ, *დისპერსიის* და *სტანდარტული გადახრის* ცოდნა. ამ შემთხვევაშიც მეტად მოსახერხებელი იქნება, თუ იგივე მიზეზის გამო შემოვიტანთ შესაბამის აღნიშვნებს დისკრეტული ალბათობის განაწილებისთვის.

განმარტება 3.8. X შემთხვევითი სიდიდის **დისპერსია** (ან განაწილების დისპერსია) არის:

$$\text{Var } X = \sum (x - EX)^2 f(x)$$

ან იგივეა, რაც

$$\left(\text{Var } X = \sum x^2 f(x) - (EX)^2 \right) \quad (5.2)$$

x -ის **სტანდარტული გადახრა** არის $\sqrt{\text{Var } X}$. ხშირად $\text{Var } X$ აღნიშვნის მაგიერ გამოიყენება σ^2 , და $\sqrt{\text{Var } X}$ -ის მაგიერ კი σ .

შემთხვეთი სიდიდის დისპერსია არის ცენტრიდან გადახრის კვადრატის ალბათობის განაწილების საშუალო (ან მოსალოდნელი მნიშვნელობა, ან მათემატიკური ლოდინი). σ^2 -ის გამოყენება პოპულაციის დისპერსიის და ალბათობის განაწილების დისპერსიის შემთხვევაშიც მოტივირებულია იგივე საფუძველზე, რომ ორივე ამ შემთხვევისთვის საშუალოსთვის აღნიშვნისთვის იხმარება μ .

მაგალითი 3.1. (გაგრძელება).

ჭანჭიკის დისპერსიისა და სტანდარტული გადახრის გამოსათვლელად შევადგინოთ ცხრილი

z	$f(x)$	$(z - 16.35)^2$	$(z - 16.35)^2 f(z)$
11	.03	28.6225	.8587
12	.03	18.9225	.5677
13	.03	11.2225	.3367
14	.06	5.5225	.3314
15	.26	1.8225	.4739
16	.09	.1225	.0110
17	.12	.4225	.0507
18	.20	2.7225	.5445
19	.15	7.0225	1.0534
20	.03	13.3225	.3997
Var Z = 4.6275			

ცხრილი 3.8. *Var Z-ის გამოთვლა*

ხოლო სტანდარტული გადახრა ტოლია:

$$\sigma = \sqrt{\text{Var } Z} = \sqrt{4.6275} = 2.15 \text{ ftlb}$$

დავალება: ჩაატარეთ Var X -ის ალტერნატიული გამოთვლა

$$\text{Var } Z = \sum z^2 f(z) - (EZ)^2 \text{ ფორმულის გამოყენებით ორივე ჭანჭიკისთვის.}$$

3.4. ბინომური და გეომეტრიული განაწილებები

ზოგიერთ შემთხვევაში დისკრეტული ალბათობის განაწილებები მიღებულია რაიმე კონკრეტულ ფიზიკურ მოვლენებზე ჩატარებული ექსპერიმენტების საფუძველზე (როგორც ეს იყო მაგალითი 1-ის შემთხვევაში). ზოგჯერ კი ასეთი განაწილების ასაგებად აიღება მარტივად სამოქმედო მათემატიკური დაშვებათა სიმრავლე, რომელმაც შეიძლება აღწეროს სხვადასხვა რეალური სიტუაცია, რომელთა ერთობლივად განხილვაც იქნება შესაძლებელი. მას შემდეგ რაც, მიღებული იქნება ასეთი განაწილება, იგი შეიძლება გამოყენებულ იქნას ბევრი სხვადასხვა მოვლენის აღსაწერად. ერთერთი ასეთ დაშვებათა სიმრავლეა, რომ მოვლენები ან ექსპერი-

მენტები უნდა იყვნენ დამოუკიდებელნი, ერთი და იგივე წარმატება-წარუმატებლობის თანაბარი შანსებით.

ბევრი საინჟინრო სიტუაცია შეიცავს ასეთი ტიპის მრავალჯერად გამეორებებს: „კი - არა“ („**GO - NO GO**“), ანუ წარმატება-წარუმატებლობა. მაგალითად, განვიხილოთ სცენარი, სადაც :

- 1) გვაქვს მუდმივი შანსი შედეგის **კი/წარმატება**. ცდათა ყოველ გამეორებაზე (ვუწოდოთ მას ალბათობა p).
- 2) ცდები არიან **დამოუკიდებელნი** იმ აზრით, რომ წინა შედეგები გავლენას არ ახდენენ შემდგომი შანსების შეფასებაზე.

(ასეთ ტიპის განაწილებას ვუწოდოთ **iid** განაწილება.)

ამ ტიპის ამოცანები მოიცავენ ყველა ელემენტის ერთობლოვ შემოწმებას, სადაც თითოეული შეიძლება კლასიფიცირდეს, როგორც „დადასტურდა“ ან არ „დადასტურდა“. მაგ. საგზაო საგუშაგოსთან ვაკვირდებით ავტომობილების გავლას, და ვინიშნავთ გაიარა დასაშვები სიჩქარით თუ ადგილი ჰქონდა სიჩქარის გადაჭარბებას.

ამ კონტექსტში გვაქვს ორი ტიპის შემთხვევითი სიდიდე, რომელთათვისაც შესაბამისი ალბათობების განაწილება უშუალოდ აიგება.

პირველი ტიპის შემთხვევაში გვაქვს ცდათა რაოდენობა n , დავითვლით რანდენჯერ ჰქონდა **კი/წარმატებას** ადგილი ანუ, განვიხილოთ ცვლადი

$X =$ **კი/წარმატებათა** რაოდენობა n დამოუკიდებელ ერთნაირ **წარმატება- წარუმატებლობა** ცდათა მიმდევრობაში.

ასეთ დროს ვამბობთ, რომ X შემთხვევითი სიდიდეს (**iid** -ს) აქვს **ბინომური (n , p) განაწილება**.

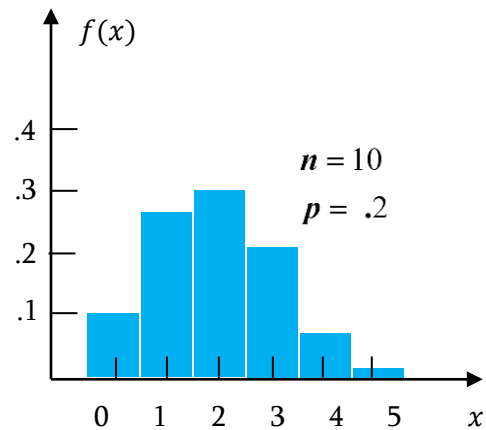
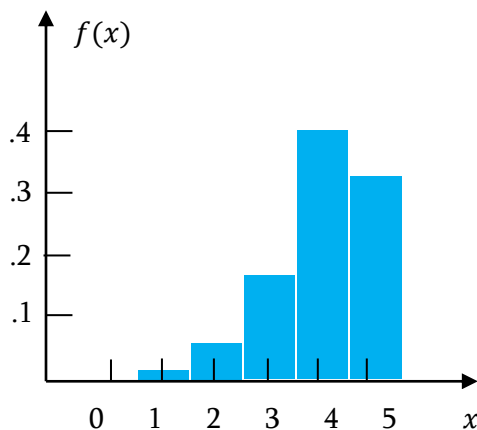
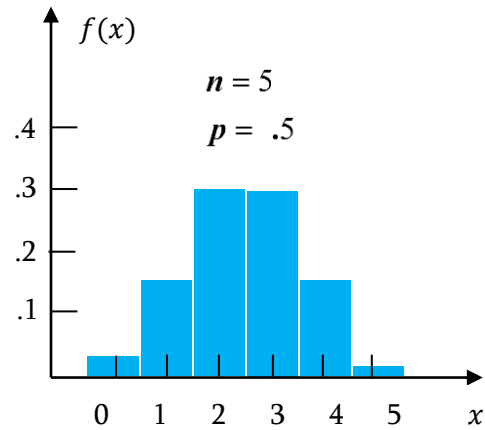
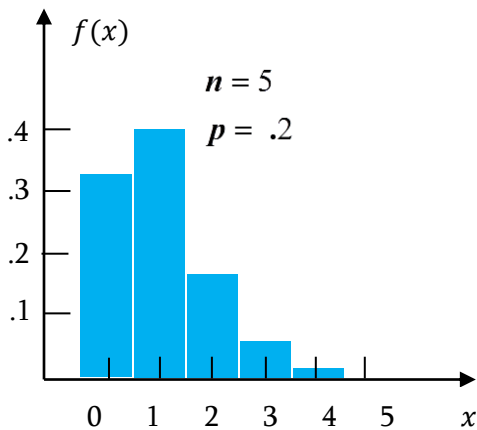
განმარტება 3.9. **ბინომური (n , p) განაწილება** არის დისკრეტული ალბათური განაწილება შემდეგი ალბათობის ფუნქციით

$$f(x) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & \text{როცა } x = 0, 1, \dots, n \\ 0 & \text{სხვა შემთხვევაში} \end{cases} \quad (3.3)$$

ნატურალური n რიცხვისთვის და $0 < p < 1$.

ფორმულა (3.3) სავსებით დამაჯერებელია. აქ გვაქვს ერთი თანამამრავლი p , რომელიც არის თითოეულ ცდაში კი/წარმატების შედეგიანობის ალბათობა და მეორე თანამამრავლი $(1 - p)$, რომელიც წარმოადგენს არა/წარუმატებლობის შედეგის ალბათობას. ხოლო $x! / n!(n - x)$ წევრი გამოსახავს n -ჯერ ჩატარებულ ცდაში დასაკვირვებელი მოვლენის x -ჯერ „კი/წარმატების“ გამოჩენის შესაძლო რაოდენობას. სახელი *ბინომური განაწილება* გამომდინარეობს იმ ფაქტიდან, რომ მნიშვნელობები $f(0), f(1), f(2), \dots, f(n)$ ($p + (1 - p)$) ^{n} ბინომის გაშლის წევრებია.

ამ დროს, თუ დავხაზავთ ალბათობის განაწილების ჰისტოგრამებს რამდენიმე სხვადასხვა ბინომური განაწილებისთვის, აღმოჩნდება რომ როცა $p < .5$, მაშინ ჰისტოგრამა დახრილია მარჯვნივ და როცა $p > .5$, ჰისტოგრამა დახრილია მარცხნივ. იმასთან, დახრილობა იზრდება რაც უფრო შორდება p -ს მნიშვნელობა $.5$ -ს, და დახრილობა მცირდება როცა n იზრდება. ბინომური ალბათობის ჰისტოგრამის შემდეგი ოთხი სურათი აღწერს ამ სიტუაციას.



გრაფიკი 3.9. ბინომური ალბათობის ჰისტოგრამები.

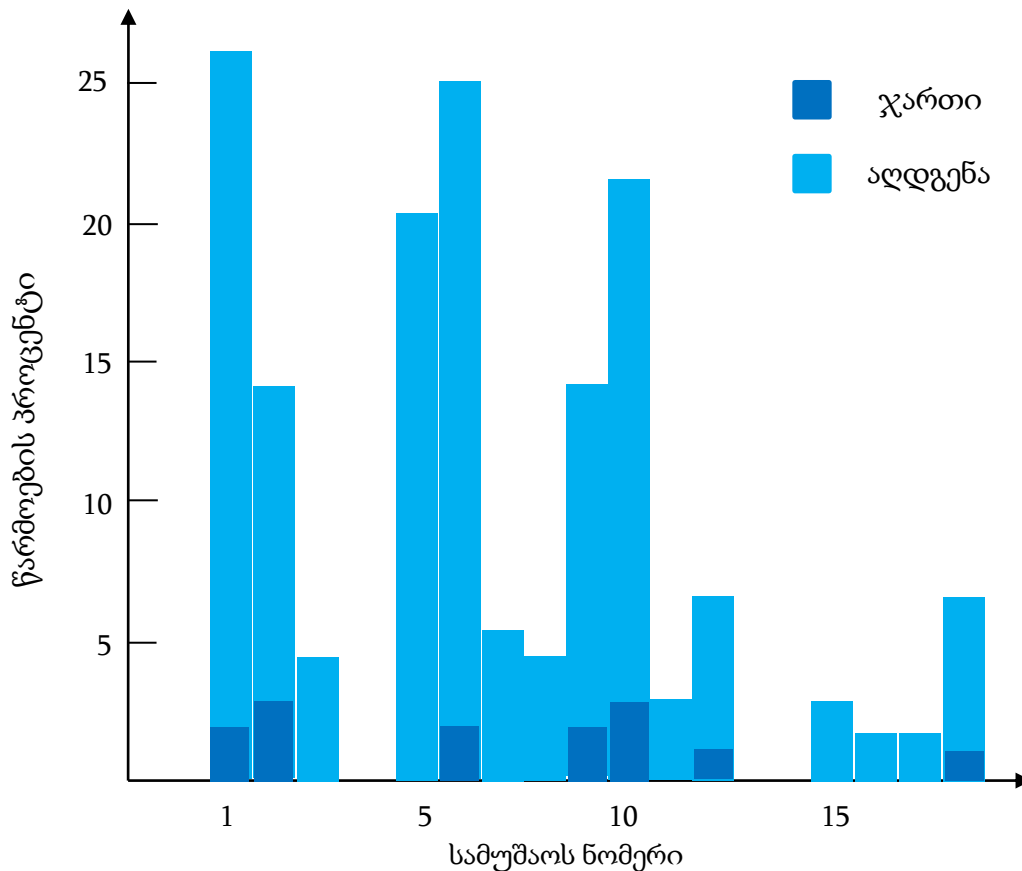
მაგალითი 3.2 ჯართი და ალდგენა განმეორებითი მუშაობისთვის.

სტატისტიკური ანალიზის თემაზე დაწერილ ერთერთ სტატიაში ჰ.როლვი აღწერს ფოლადის ღერძების ალდგენის პროცესში ხარიხის გაუმჯობესების სტატისტიკურად დაფუძნებულ გეგმას. ცხრილში წარმოდგენილია პროცენტული მაჩვენებლები ორი პარამეტრის მიხედვით (ჯართი, ალდგენა) იმ 18 ღერძისთვის, რომელიც წარმოდგენილი იყო შესწავლის პროცესში.

სამუშაოს ნომერი	ჯართის პროცენტი	ალდგენის პროცენტი
1	2	25
2	3	11
3	0	5
4	0	0
5	0	20
6	2	23
7	0	6
8	0	5
9	2	8
10	3	18
11	0	3
12	1	5
13	0	0
14	0	0
15	0	3
16	0	2
17	0	2
18	1	5

ცხრილი 3.10. პროცენტული მაჩვენებლები „ჯართი და ალდგენა განმეორებითი მუშაობისთვის“.

შემდეგ გრაფიკზე კი მოცემულია სამუშაოების ჰისტოგრამა დალაგებული ორივე პარამეტრის ყოფაქცევის მიხედვით. (სვეტის მთლიანი სიმაღლე წარმოადგენს ორივე პარამეტრის მნიშვნელობათა ჯამს). მე-10 და მე-11 ნომერ სამუშაოებს შორის შედეგების მკვეთრი შემცირება გამოწვეულია ერთერთი კომპანიის სახარატო დაზგების კაპიტალური რემონტის მიზეზით. შევნიშნოთ, რომ ეს მომენტი ადრევე საჭიროებს ყურადღებას მონაცემთა ანალიზის პროცესში.



გრაფიკი 3.11. სეგმენტებად მოცემული სვეტოვანი დიაგრამა ქართისა და აღდგენითი სამუშაოებისთვის.

ბინომური განაწილება და აღდგენადი ღერძების დათვლა.

ამ მაგალითში დაახლოებით ღერძების 20% აღდგენადია. ამიტომ რაელურია ჩავთვალოთ, რომაღბათობა იმისა, რომ მოცემული ღერძი იყოს აღდგენადი ტოლი იყოს $p = .2$. დავუშვათ აგრეთვე, რომ შესამოწმებლად შერჩეული ღერძების

რაოდენობა $n = 10$, და გამოვთვალოთ ალბათობა იმისა, რომ ორი მაინც ჩაითვლება აღდგენად.

ჩავთვალოთ, რომ ღერძის ვარგისიანობის დადგენის მოდელი არის დამო-უკიდებელი და წარმატება/წარუმატებლობის ერთნაირი შანსით.

U = აღდგენადი ღერძების რაოდენობა შერჩევაში, რომლის ზომა $n = 10$,

არის ბინომური შემთხვევითი სიდიდე, როცა $n = 10$ და $p = .2$. ამიტომ

$$\begin{aligned} P [\text{ორი ღერძი მაინც იქნება აღდგენადი}] &= P[U \geq 2] = \\ &= f(2) + f(3) + \dots + f(10) = 1 - (f(0) + f(1)) = \\ &= 1 - \left(\frac{10!}{0!10!} (.2)^0 (.8)^{10} + \frac{10!}{1!9!} (.2)^1 (.8)^9 \right) = .62 \end{aligned}$$

აქ გამოყენებული ტრიუკი საშუალებას გვაძლევს თავი ავარიდოთ ბინომური ალბათობის ფუნქციის ცხრაჯერ გამოთვლას. რადგან ყველას ჯამი 1 -ს ტოლია, ამიტომ გამითვლის ასეთი წესი უფრო მოკლე და მოსახერხებელია.

მნიშვნელობა .62 მაშინ არის კარგი, თუ მოდელი იძლევა დაშვებას, რომ ვაწარმოოთ ეს პროცესი. თუ დამუკიდებელი, ერთნაირ წარმატება/წარუმატებლობა ცდათა მიმდევრობაში, ღერძების წარმოებაში არის ჩავარდნები, მაშინ ეს იქნება არასასიამოვნო ფიზიკური რეალობა, თუმცა შესაძლებელია მნიშვნელობა .62 მათემატიკური თვალსაზრისით კარგად გამოიყურებოდეს. ცუდი აღწერით კი შეიძლება მოხდეს ის, რაც ვთქვით. მაგალითად, ვთქვათ ცვეთის გამო 40 ღერძი დაექვემდებარა შემოწმებას, შემდეგ 10 აღდგენილით მოხდა ინსტრუმენტის შეცვლა და ა.შ. ამ შემთხვევაში ბინომური განაწილება ძალიან მწირი აღწერაა U -სი და .62 მნიშვნელობას საერთოდ არა აქვს რაიმე ღირებულება.(ამ შემთხვევაში ცდათა დამოუკიდებლობის დაშვება შეუსაბამოა მოცემული სიტუაციისთვის.)

ბინომური განაწილება და მარტივი შემთხვევითი შერჩევა.

არსებობს ერთი მნიშვნელოვანი გარემოება; ზოგჯერ გვაქვს შემთხვევა, რომ მოდელი არის დამუკიდებელი, ერთნაირ „წარმატება/წარუმატებლობა“ ცდათა მიმდევრობით მაგრამ არც თუ მაინცა და მაინც შესაბამისი განაწილებით. ასეთ დროს შესაძლებელია ვისარგებლოთ ისეთიბინომური განაწილებით, რომელიც იქნება უფრო ადექვატური პრაქტიკული მიზნებისთვის. ასეთი არის, მაგალითად, ე.წ. დიქტომიური (ორად გაყოფადი) პოპულაციიდან აღებული მარტივი შემთხვევითი შერჩევის შედე-

გების აღწერა. დავუშვათ, რომ პოპულაცია, რომლის ზომა არის N , შეიცავს p ნაწილს A ტიპის საგნების და $(1 - p)$ ნაწილს B ტიპის საგნებისას. თუ გაკეთებულია მარტივი შემთხვევითი n ადგილიანი შერჩევა და

X = შერჩევაში A ტიპის საგნების რაოდენობა

მკაცრად შეგვიძლია ვთქვათ, რომ X არ არის ბინომური შემთხვევითი ცვლადი. მაგრამ თუ n მცირე ნაწილია N -ის, (ვთქვათ, ნაკლებია 10% -ზე), და p არ არის ძალიან ექსტრემალური (არ არის ძალიან ახლოს 0 -თან ან 1 -თან), X დაახლოებით იქნება ბინომური (n, p) .

მაგალითი 3.3. მარტივი შემთხვევითი შერჩევა უროტროპინის ბევრი გრანულებისთვის.

გრანულების დამამზადებელი მოწყობილობის ერთერთი ექსპერიმენტის დროს გრენიერმა, გრიმმა, ლარსონმა და ლუკომსკიმ მიაკვლიეს მანქანების განლაგების ისეთ კომბინაციას, რომელიც საშუალებას აძლევთ მათ აწარმოონ ახალი 66 გრანულა, რომელიც შეესაბამება ტექნიკური რეგლამენტის მოთხოვნებს 100 ცალიანი პარტიის სანაცვლოდ. ჩავთვალოთ, რომ 100 ცალი გრანულა არის ჩვენთვის საინტერესო პოპულაცია და განვიხილოთ მარტივი შემთხვევითი შერჩევა ზომით $n = 2$. თუ შემთხვევით ცვლადს განვსაზღვრავთ შემდეგნაირად,

V = ახალი გრანულების რაოდენობა შერჩევაში, რომლის ზომაა 2 ,

მაშინ V შემთხვევითი სიდიდისთვის ძალზე ბუნებრივად მიიღება ალბათობის განაწილება შემდეგნაირად: V -ს შესაძლო მნიშვნელობებია 0, 1 და 2 .

$$f(0) = P [V = 0] =$$

= P [პირველი ამოღებული გრანულა არ არის ახალი და შემდეგი მეორე გრანულაც აგრეთვე არ არის ახალი] .

$$f(2) = P [V = 2] =$$

= P [პირველი ამოღებული გრანულა არის ახალი და მეორე ამოღებული გრანულაც არის აგრეთვე ახალი] .

$$f(1) = 1 - (f(0) + f(2)) .$$

მაშინ ვფიქრობთ, „გრძელ ვადიან პერსპექტივაში მიიღება, რომ პირველი ამოღებული არა ახალი გრანულის მოსვლის შანსი იქნება 34 შესაძლებლობა 100 -დან. თუ განვიხილავთ მხოლოდ ასეთ შემთხვევას, მაშინ კვლავ გრძელ ვადიან პერსპექტივაში, ანუ როდესაც ცდათა რიცხვი ძალიან დიდია, შემდეგი ამოღებისას არა ახალი გრანულის მოსვლის შანსი იქნება 99 -დან 33.“ მაშინ უპრიანი იქნება, რომ $f(0)$ ასე გამოვთვალოთ:

$$f(0) = \frac{34}{100} \cdot \frac{33}{99} = .1133$$

მსგავსად,

$$f(2) = \frac{66}{100} \cdot \frac{65}{99} = .4333$$

და აგრეთვე

$$f(1) = 1 - (.1133 + .4333) = 1 - .5467 = .4533$$

ახლა, უკვე აღარ უნდა გვეგონოს, რომ V უქველად დამოუკიდებელი ცდებისგან არის წარმოშობილი. მაგალითად, თუ ვიცით, რომ პირველი ამოღებული გრანულა იყო ახალი, ხალხის უდიდესი ნაწილის შეფასებით, შანსი იმისა, რომ მეორედ ამოღებული გრანულაც იქნება დამადასტურებელი, არის $\frac{66}{100}$ -დან $\frac{65}{99}$ -მდე. მიუხედავად ამისა, უმეტესი წილი პრაქტიკული მიზნებისთვის უკეთესია ჩავთვალოთ, რომ V არის არსებითად ბინომური განაწილება, როცა $n=2$ და $p = .66$. ამის დასაანახად გამოვთვალოთ შემდეგი

$$\frac{2!}{0! 2!} (.34)^2 (.66)^0 = .1156 \approx f(0)$$

$$\frac{2!}{1! 1!} (.34)^1 (.66)^1 = .4488 \approx f(1)$$

$$\frac{2!}{2! 0!} (.34)^0 (.66)^2 = .4356 \approx f(2)$$

აქ n არის N -ის მცირე ნაწილი, ხოლო p არ არის ძალიან ექსტრემალური, მაშინ ბინომური განაწილება კარგად აღწერს ცვლადს, რომელიც წარმოიშობა მარტივი შემთხვევითი შერჩევისას.

ბინომური შემთხვევითი სიდიდის საშუალოს და დისპერსიის გამოთვლა ძალიან მარტივდება თუ (3.1) და (3.2) ფორმულებს გამოვიყენებთ ბინომური ალბათობების გამოსათვლელად რომელიც მოიცემა (3.3) ტოლობით. X ბინომური (n, p) შემთხვევითი სიდიდის საშუალოსთვის გვაქვს:

$$\mu = EX = \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = np \quad (3.4)$$

ბინომური (n, p) განაწილების საშუალო ანუ მათემატიკური ლოდინი

გარდა ამისა, დისპერსიისთვის გვექნება:

$$\sigma^2 = \text{Var}X = \sum_{x=0}^n (x - np)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = np(1-p) \quad (3.5)$$

ბინომური (n, p) განაწილების დისპერსია

მაგალითი 3.2. (გაგრძელება)

დავუბრუნდეთ ფოლადის ღერძების აღდგენის ამოცანას. კვლავ დავუშვათ, რომ გვაქვს ბინომური განაწილება $n = 10$ და $p = .2$, რომელიც შეესაბამება სიტუაციას:

U = აღდგენადი ღერძების რაოდენობა შერჩევაში, რომლის ზომა = 10.

მაშინ (5.4) და (5.5) ფორმულების თანახმად გვექნება:

$$EU = 10 \cdot (.2) = 2 \quad (\text{ღერძი})$$

$$\sqrt{\text{Var } U} = \sqrt{10 \cdot (.2) \cdot (.8)} = 1.26 \quad (\text{ღერძი})$$

შემთხვევითი სიდიდის **მეორე ტიპი**, რომელიც ასოცირდება დამოუკიდებელ ერთნაირ წარმატება - წარუმატებლობა ცდათა მიმდევრობასთან (iid), არის

X = ცდათა რაოდენობა, კი/წარმატების პირველ გამოჩენამდე.

ასეთ X განაწილებას ეწოდება **გეომეტრიული (p) განაწილება**.

განმარტება 3.10.

გეომეტრიული (p) განაწილება არის დისკრეტული ალბათური განაწილება შემდეგი ალბათობის ფუნქციით

$$f(x) = \begin{cases} p(1-p)^{x-1} & \text{როცა } x = 1, 2, \dots \\ 0 & \text{სხვა შემთხვევაში} \end{cases}$$

სადაც, $0 < p < 1$.

ფორმულა (3.6) -ს გააჩნია კარგი ინტუიციური შინაარსი. თუ X -ის დალაგებულ მნიშვნელობებში ავიღებთ მნიშვნელობას x -ს, მაშინ უნდა გვქონდეს მიყოლებით $x - 1$ არა/წარუმატებელი შედეგი, რომელსაც მოყვება კი/წარმატება. (შეიძლება ვიფიქროთ პირიქითაც: პირველ წარმატებას მოყვება $x - 1$ წარუმატებლობა). ფორმულა (3.6) -ში გვაქვს $x - 1$ ცალი $1 - p$ თანამამრავლი და ერთი p . მეორე მიზეზი, რომ დავინახოთ ფორმულა (3.6)-ის დამაჯერებლობა არის ის, რომ თუ X შემთხვევითი სიდიდის შემთხვევითი მნიშვნელობებია $x = 1, 2, \dots$ მაშინ

$$1 - F(x) = 1 - P[X \leq x] = P[X > x] =$$

$$= P[x \text{ არის არა/წარუმატებლობათა რაოდენობა } x \text{ ცდაში}],$$

თუ გამოვიყენებთ ბინომური (x, p) ალბათობის ფუნქციას, რომელიც მოიცემა (3.3) ფორმულით, მივიღებთ:

$$1 - F(x) = (1 - p)^x$$

ეს არის მარტივი დამოკიდებულება გეომეტრიულ (p) განაწილებასა და ალბათობის დაგროვების ფუნქციას შორის.

როცა $x = 2, 3, \dots$ გვაქვს,

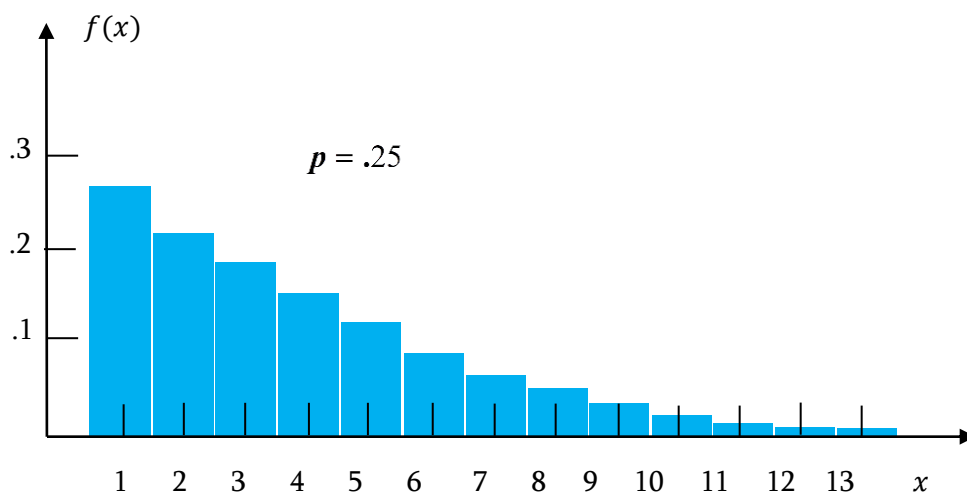
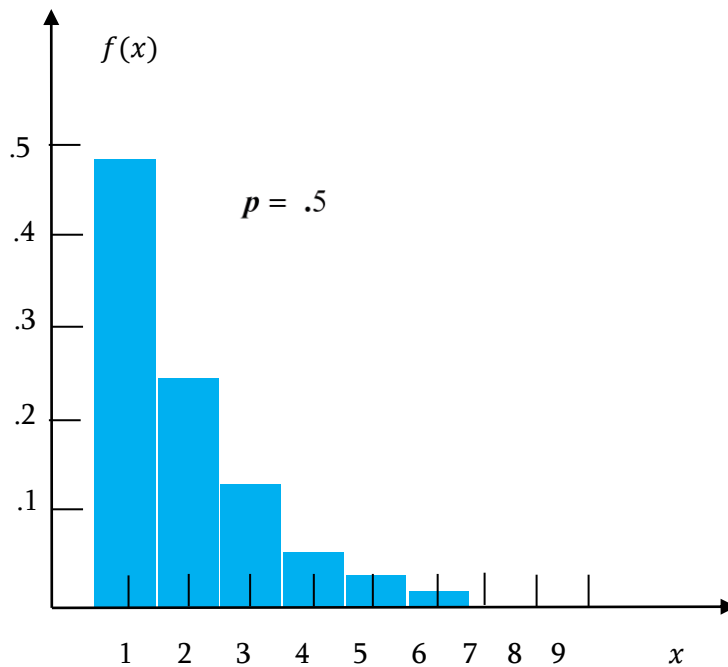
$$f(x) = F(x) - F(x-1) = -(1 - F(x)) + (1 - F(x-1)).$$

ეს კი (5.7) -თან ერთად გვაძლევს (5.6) -ს.

ტერმინი *გეომეტრიული* აღებულია იმ ფაქტის გამო, რომ მნიშვნელობები $f(1), f(2), f(3), \dots$ არიან უსასრულოდ კლებადი გეომეტრიული წევრები, რომლის ჯამია

$$p \cdot \frac{1}{1 - (1 - p)}$$

გეომეტრიული განაწილება არის ისეთი დისკრეტული განაწილება, რომლის ალბათობის განაწილების ჰისტოგრამა ექსპონენციალურად კლებულობს როცა x იზრდება. ორი სხვადასხვა გეომეტრიული ალბათობის ჰისტოგრამა გამოსახულია შემდეგ გრაფიკებზე.



გრაფიკი 3.12. ორი გეომეტრიული ალბათობის ჰისტოგრამა.

მაგალითი 3.4. გეომეტრიული განაწილება ნიკელ - კადმიუმის (NiCad) ბატარიებში და მოკლე შეერთებები.

ჟურნალ „*Journal of Quality Technology*, 1988” -ში გამოქვეყნებულ ერთერთ სტატიაში ოპირი, ელ-გადი და სნიდერი აღწერენ ექსპერიმენტების სერიას. იგი ტარდება იმ მიზნით, რომ შემცირდეს ბატარიის გარკვეული კომპონენტის წუნდებულობის ანუ ჯვართში ჩასაბარებელი რაოდენობის პროპორცია. წუნდებულობის მიზეზია ქარხნის მიერ გამოშვებული ბატარიების (აკუმულატორების) შიდა მოკლე ჩართვა. ეს ექსპერიმენტალური პროგრამა აღმოჩნდა წარმატებული. მან 1%-მდე შეამცირა ქარხნის მიერ წარმოებული იმ კომპონენტის წუნდებულობის მაჩვენებელი, რომლის მიზეზითაც ხდებოდა მოკლე ჩართვა.

დავუშვათ, რომ ტესტირება მიმდინარეობს ამ ქარხნის მთლიანი წარმოების ციკლზე. ვთქვათ,

T = ტესტირებათა რაოდენობა, როდესაც პირველად დაფიქსირდა მოკლე ჩართვა.

T -ს მოდელი (შესაბამისად თუ მიდრეკილია *iid* აღწერისკენ) არის გეომეტრიული განაწილება, რომლის $p = .01$. (p არის ალბათობა იმისა, რომ რაიმე კონკრეტული ტესტი აღმოაჩენს მოკლე ჩართვას.) მაშინ (5.6) ფორმულის გამოყენება გვაძლევს:

P [მოწყობილობის პირველი ან მეორე ტესტირების შედეგად

დაფიქსირდა მოკლე ჩართვა] =

$$= P[T = 1 \text{ ან } T = 2] = f(1) + f(2) = .01 + .01 \cdot (1 - .01) \approx .02 .$$

(5.7) ფორმულიდან კი გვექნება

P [სულ მცირე 50 -მა მოწყობილობამ გაიარა ტესტირება ისე, რომ მოკლე

ჩართვა არ აღმოჩენილა] =

$$= P[T > 50] = (1 - .01)^{50} = .61$$

ბინომური განაწილების მსგავსად გეომეტრიულ განაწილებასაც გააჩნია საშუალო (მათემატიკური ლოდინი) და დისპერსია, რომლებიც შეიძლება გამოისახოს როგორც p პარამეტრის მარტივი ფუნქციები. ე.ი. თუ X არის გეომეტრიული (p) განაწილება

$$\mu = EX = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{1}{p} \quad (3.8)$$

გეომეტრიული (p) განაწილების საშუალო (მათემატიკური ლოდინი)

$$\sigma^2 = \text{Var } X = \sum_{n=1}^{\infty} \left(x - \frac{1}{p}\right)^2 p(1-p)^{x-1} = \frac{1-p}{p^2} \quad (3.9)$$

გეომეტრიული (p) განაწილების დისპერსია.

მაგალითი 3.4. (გაგრძელება)

ბატარიების შემთხვევაში T -ს საშუალო და დისპერსია ტოლია:

$$ET = \frac{1}{.01} = 100 \quad \text{ბატარეა}$$

$$\sqrt{\text{Var } T} = \sqrt{\frac{(1 - .01)}{(.01)^2}} = 99.5 \quad \text{ბატარეა}$$

ფორმულა (3.8) ინტუიციურად გასაგები შედეგია, რადგან თუ 100 -დან ერთი შანსია იმისა, რომ ბატარეაში ტესტირების შედეგად აღმოჩნდება მოკლე ჩართვა, მაშინ აზრიანია ველოდეთ, რომ საშუალოდ 100 ტესტირებაში პრობლემა დაფიქსირდება პიველივე შემთხვევაში.

3.5. პუასონის განაწილება.

ხშირად ძალიან მნიშვნელოვანია მივაღწევოთ თვალყური მოვლენათა რაიმე გრძელ, ფარდობითად იშვიათ მოვლენათა თანმიმდევრობას. ანდა ხანგრძლივი დროის

განმავლობაში ვაკვირდებოდეთ მოვლენებს, რომელთაც პოტენციურად გააჩნიათ მსგავსი შედეგები.

ასეთ დროს, ალბათობის განაწილებები საჭიროებენ აღწეროთ ფარდობითად იშვიათი მოვლენების მოხდენათა შემთხვევითი რაოდენობა, განსაზღვრული რაიმე ტიპის ინტერვალის (მაგ., დროის, ან სივრცის) საშუალებით. ამ კონტექსტში დღესდღეობით ერთერთი ხშირად გამოყენებადი თეორიული განაწილება არის *პუასონის განაწილება*.

განმარტება 3.11. პუასონის (λ) განაწილება არის დისკრეტული შემთხვევითი სიდიდის ალბათობის განაწილება, რომელიც მოიცემა ალბათობის ფუნქციით

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{როცა } x = 0, 1, 2, \dots \\ 0 & \text{სხვა შემთხვევაში} \end{cases} \quad (5.10)$$

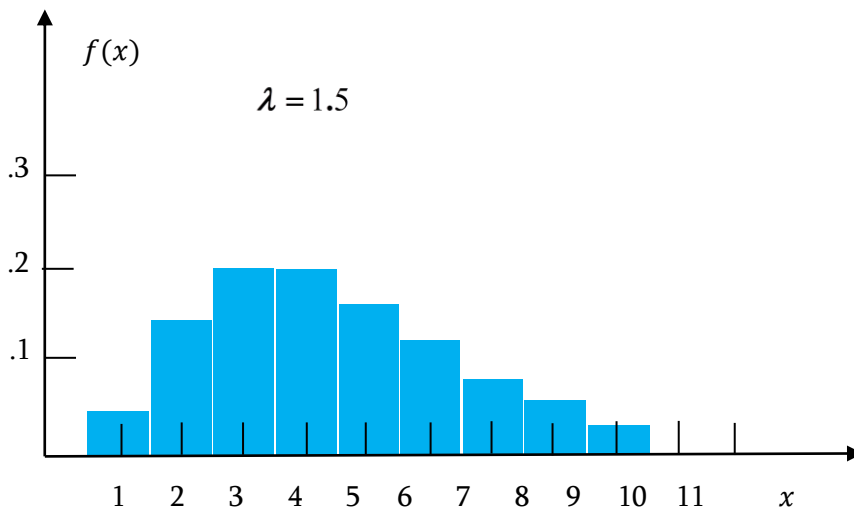
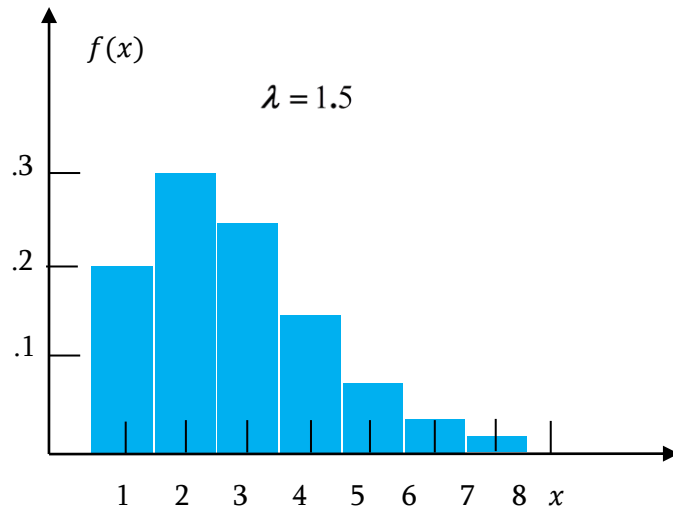
როცა $\lambda > 0$.

ეს ფორმულა თავიდან შეიძლება არ მოგვეჩვენოს მიმზიდველად. სინამდვილეში ეს არის ერთერთი ისეთი ფორმულა, რომელსაც გააჩნია დამაჯერებელი მათემატიკური საფუძველი. არის იოლად მოხმარებადი და გააჩნია მეტად სასარგებლო გამოყენებები „იშვიათი მოვლენების“ სიტუაციებში. ერთერთი მთავარი გარემოება, როდესაც უნდა მივმართოთ ამ ფორმულას, არის დამოუკიდებელ ცდათა ძალიან დიდი რიგი, სადაც წარმატების ალბათობათითოეულ ცდაში არის ძალიან მცირე და ცდათა რაოდენობის ნამრავლი წარმატების ალბათობაზე ტოლია λ -სი. ანუ $np = \lambda$. ამას მივყავართ კი ბინომური $\left(n, \frac{\lambda}{n}\right)$ განაწილებისკენ. თუმცა ფაქტია, რომ დიდი n -ებისთვის

ბინომური $\left(n, \frac{\lambda}{n}\right)$ ალბათობის ფუნქციის აპროქსიმაცია განისაზღვრება (3.10) ფორმულით. შეიძლება ვიფიქროთ, რომ პუასონის განაწილება არის იმ მიზნის შედეგი, როდესაც გვინდა, რომ დავითვალოთ ბევრ დამოუკიდებელ მოვლენათა მოხდენის ან არ მოხდენის ძალიან მცირე მსგავსი შანსები დროისა თუ სივრცის დიდშუალებდში.

პუასონის განაწილება არის მარჯვნივ დახრილი განაწილება. თუ $x = 0, 1, 2, \dots$ მნიშვნელობათა თავზე ავიღებთ ალბათობის ჰისტოგრამას, აღმოჩნდება რომ მისი

პიკი ახლოს იქნება λ -ს შესაბამის მნიშვნელობასთან. შემდეგ გრაფიკებზე წარმოდგენილია ორი განსხვავებული პუასონის განაწილების ჰისტოგრამა.



გრაფიკი 3.13. პუასონის ალბათობის ორი ჰისტოგრამა.

პუასონის (λ) განაწილებისთვის λ წარმოადგენს განაწილების საშუალოსაც და დისპერსიასაც. ანუ თუ X არის პუასონის (λ) განაწილება, მაშინ

$$\mu = EX = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \quad (3.11)$$

პუასონის (λ) განაწილების საშუალო

და

$$\text{Var } X = \sum_{x=0}^{\infty} (x - \lambda)^2 \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \quad (3.12)$$

პუასონის (λ) განაწილების დისპერსია

ფაქტია, რომ ფორმულა (3.11) ძალიან სასარგებლოა იმისთვის, რომ შევარჩიოთ პუასონის რომელი განაწილება აღწერს უფრო უკეთესად „იშვიათი მოვლენის“ სიტუაციას კონკრეტულ შემთხვევაში.

მაგალითი 3.5. პუასონის განაწილება და α -ნაწილაკების დათვლა.

რეზერფორდისა და გეიგერის კლასიკურ მონაცემთა სიმრავლე წარმოდგენილი იყო ჟურნალში *Philosophical Magazine*, 1910. ამოცანა იყო α -ნაწილაკების რაოდენობის დადგენა, რომელსაც ასხივებდა პოლონიუმის პატარა ღერო. ნაწილაკების შეჯახება ხდებოდა ღეროსთან ახლოს მდებარე ეკრანზე 2 608-ჯერ 8 წუთიანი პერიოდებით. რეზერფორდისა და გეიგერის ფარდობითი სიხშირის განაწილების საშუალოა 3.87 და აღსანიშნავია, რომ ფორმა დაახლოებით არის პუასონის ალბათობის განაწილების საშუალო $\lambda = 3.87$

რეზერფორდ - გეიგერის დუბლირებულ ექსპერიმენტში, მისაღებია ალბათობის ფუნქცია ამ პროცესის აღსაწერად,

$S = \alpha$ -ნაწილაკების ეკრანზე შეჯახებათა რაოდენობა 8 წუთიანი პერიოდებით

ფუნქცია იქნება ასეთი:

$$f(s) = \begin{cases} \frac{e^{-3.87} (3.87)^s}{s!} & s = 0, 1, 2, \dots \text{თვის} \\ 0 & \text{სხვა შემთხვევაში} \end{cases}$$

ამ მოდელის გამოყენებით ჩვენი მაგალითის შემთხვევაში გვექნება:

$$\begin{aligned}
 & P[\text{დარეგისტრირდა სულ მცირე 4 ნაწილაკი}] \\
 &= P[S \geq 4] = f(4) + f(5) + f(6) + \dots \\
 &= 1 - (f(0) + f(1) + f(2) + f(3)) \\
 &= 1 - \left(\frac{e^{-3.87} (3.87)^0}{0!} + \frac{e^{-3.87} (3.87)^1}{1!} + \frac{e^{-3.87} (3.87)^2}{2!} + \frac{e^{-3.87} (3.87)^3}{3!} \right) \\
 &= .54
 \end{aligned}$$

მაგალითი 3.6. ბიბლიოთეკაში შემომსვლელთა რაოდენობა.

სტროკი, ვოლსდორფი და მაკ-არტური აგროვებენ მონაცემებს უნივერსიტეტის ბიბლიოთეკაში შემომსვლელ სტუდენტთა რაოდენობის შესახებ კვირის სხადასხვა დღეს, დროის რაღაც პერიოდებში. მაგალითად, მათი მონაცემებით დაფიქსირდა, რომ 12:00 -დან 12:10 p.m. -მდე ორშაბათიდან ოთხშაბათამდე საშუალოდ ბიბლიოთეკა მოინახულა 125-მა სტუდენტმა. განვიხილოთ მოდელი

M = უნივერსიტეტის ბიბლიოთეკაში შემომსვლელ სტუდენტთა რაოდენობა 12:00 -დან 12:01 -მდე შემდეგ სამშაბათს.

თუ M -ის აღწერისთვის გამოვიყენებთ პუასონის განაწილებას, მაშინ მიასდება, რომ λ პარამეტრი განვსაზღვროთ შემდეგნაირად:

$$\lambda = \frac{125 \text{ სტუდენტი}}{10 \text{ წუთი}} (1 \text{ წუთი}) = 12.5 \text{ სტუდენტი.}$$

ამ დროს,

$$EM = \lambda = 12.5 \text{ სტუდენტი,}$$

$$\sqrt{\text{Var } M} = \sqrt{\lambda} = \sqrt{12.5} = 3.54 \text{ სტუდენტი}$$

და მაგალითად, ალბათობა იმისა, რომ 10 -დან 15 სტუდენტამდე (ჩათვლით) მოინახულეს 12:00 და 12:01 დროის შუალედში გამოითვლება შემდეგნაირად:

$$\begin{aligned}
 P[10 \leq M \leq 15] &= f(10) + f(11) + f(12) + f(13) + f(14) + f(15) \\
 &= \frac{e^{-12.5} (12.5)^{10}}{10!} + \frac{e^{-12.5} (12.5)^{11}}{11!} + \frac{e^{-12.5} (12.5)^{12}}{12!} \\
 &+ \frac{e^{-12.5} (12.5)^{13}}{13!} + \frac{e^{-12.5} (12.5)^{14}}{14!} + \frac{e^{-12.5} (12.5)^{15}}{15!} \\
 &= .60
 \end{aligned}$$

სავარჯიშოები:

1. X დისკრეტული შემთხვევითი სიდიდე მოცემულია ალბათობის ფუნქციით

x	2	3	4	5	6
$f(x)$	.1	.2	.3	.3	.1

ა) შეადგინეთ ალბათობის ჰისტოგრამა X -სთვის. დახაზეთ გრაფიკი $F(x)$ დაგროვილი ალბათობის ფუნქციისთვის.

ბ) იპოვეთ X -ის საშუალო და სტანდარტული გადახრა.

2. წარმოვიდგინოთ, რომ მცირე პოპულაცია ზომით $N = 6$ და მნიშვნელობებით 2, 3, 4, 4, 5, 6.

ა) დახაზეთ ამ პოპულაციისთვის ფარდობითი სიხშირის ჰისტოგრამა და გამოითვალეთ პოპულაციის საშუალო μ და სტანდარტული გადახრა σ .

ბ) ვთქვათ, $X =$ ამოვირჩიოთ შემთხვევით ერთი რიცხვი პოპულაციიდან. დახაზეთ მისთვის ალბათობის ჰისტოგრამა და გამოითვალეთ EX და $\text{Var } X$.

გ) განვიხილოთ მოცემული პატარა პოპულაციიდან მარტივი შემთხვევითი ამორჩევა ზომით $n = 2$. შეადგინეთ ცხრილი, რომელიც გამოსახავს ალბათობის განაწილებას შემთხვევითი სიდიდეებისთვის:

$\bar{X}$ = შერჩევის საშუალო

s^2 = შერჩევის დისპერსია

(აქ გვექნება 15 შესაძლებლობა არადალაგებული ამორჩევებისა ზომით 2 პოპულაციიდან, რომელიც შეიცავს 6 ელემენტს. ყოველი შერჩევა ამ 15 -დან ტოლ-შესაძლებელია, რომ იქნეს ამორჩეული. მათ აქვთ თავიანთი $\bar{x}$ და s^2 .) გამოიყენეთ ცხრილი და ააგეთ ალბათობის ჰისტოგრამა. გამოთვალეთ $E \bar{X}$ და $\text{Var } \bar{X}$. შეადარეთ ეს სიდიდეები μ -ს და σ^2 -ს.

3. დახაზეთ ალბათობის ჰისტოგრამა ბინომური განაწილებისთვის, რომლის $n = 5$ და $p = .1, .3, .5, .7$ და $.9$. თითოეულ ჰისტოგრამაზე მონიშნეთ საშუალო და მიუთითეთ სტანდარტული გადახრა.
4. დახაზეთ ალბათობის ჰისტოგრამა პუსაონის განაწილებისთვის, რომლის საშუალოები $\lambda = .5, 1.0, 2.0$, და 4.0 . თითოეულ ჰისტოგრამაზე მონიშნეთ საშუალო და მიუთითეთ სტანდარტული გადახრა.
5. რა განსხვავებაა ფარდობითი სიხშირის განაწილებასა და ალბათობის განაწილებას შორის?

4. უწყვეტი შემთხვევითი სიდიდე

ძალიან ხშირად მოსახერხებელია ვიგულისხმოდ, რომ შემთხვევითი სიდიდე დისკრეტული კი არ არის, არამედ უწყვეტია, იმ თვალსაზრისით, რომ გვაქვს მთლიანი (უწყვეტი) ინტერვალი შემთხვევითი სიდიდის შესაძლო მნიშვნელობათა სიმრავლისთვის. წინა თავისგან განსხვავებით, უწყვეტი ალბათობის განაწილება იყენებს კომპიუტერულ ტექნოლოგიებს. ჩვენი პირველი ამოცანაა აუწყვეტი შემთხვევითი სიდიდისთვის შემოვიტანოთ ალბათობის სიმკვრივის ფუნქციის ცნება და აღნიშვნა (ქართულ სამეცნიერო ლიტერატურაში ეს ფუნქცია *განაწილების სიმკვრივის* სახელითაა ცნობილი) და ვაჩვენოთ მისი დამოკიდებულება დაგროვებით ალბათობის ფუნქციასთან (რომელიც ქართულ სამეცნიერო ლიტერატურაში ცნობილია *განაწილების ფუნქციის* სახელით). ვაჩვენოთ აგრეთვე, თუ როგორ გამოვთვალოთ უწყვეტი განაწილების საშუალო და დისპერსია. ამის შემდეგ განვიხილავთ რამდენიმე სტანდარტულ უწყვეტ განაწილებას, რომელთაც აქვთ მეტად საჭირო გამოყენებები საინჟინრო ამოცანების გადაწყვეტისას. ასეთია მაგალითად, ნორმალური (ანუ გაუსის) მაჩვენებლიანი და ვეიბულის განაწილებები.

4.1. ალბათობის სიმკვრივის ფუნქციები და დაგროვებითი ალბათობის ფუნქციები

მეთოდებს, რომლებიც გამოიყენება ალბათობის განაწილებების განსაზღვრისა და აღწერისთვის, გააჩნიათ პარალელები მექანიკაში. როდესაც განვიხილავთ ალბათობის უწყვეტ განაწილებებს, ანალოგიები მექანიკასთან ხდება განსაკუთრებულად სასარგებლო. მექანიკაში მასის უწყვეტი განაწილების თვისებები დაკავშირებულია სიმკვრივის შესაძლო ცვლილებასთან მისი განთავსების ადგილებში. მასის რაოდენობა კონკრეტულ არეებში მიიღება სიმკვრივიდან მისი ინტეგრებით.

ცნება „ალბათობის სიმკვრივე“ ალბათობათა თეორიაში შეესაბამება მასის სიმკვრივის ცნებას მექანიკიდან. ალბათობის უწყვეტი განაწილების განსაზღვრისთვის საჭიროა აღვწეროთ „რამდენად მსხვილია“ ალბათობა შესაძლო

მნიშვნელობათა სიმრავლის სხვადასხვა ნაწილში. მისი ფორმალური განსაზღვრება ასეთია:

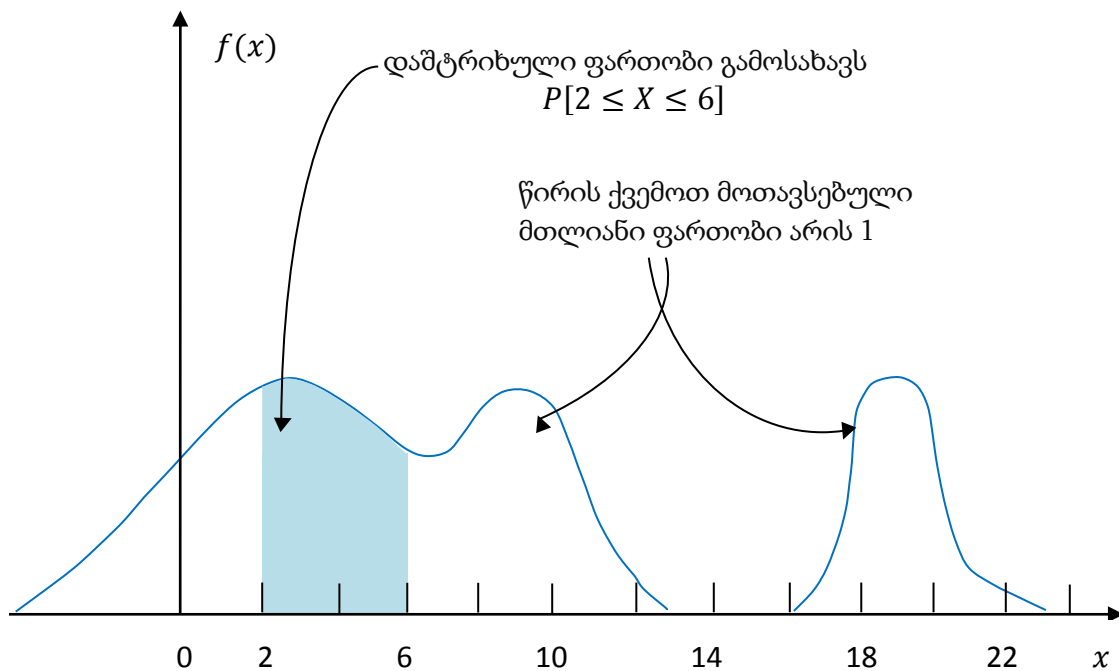
განმარტება 4.1. უწყვეტი X შემთხვევითი სიდიდისთვის ალბათობის სიმკვრივის ფუნქცია არის ისეთი არაუარყოფითი $f(x)$ ფუნქცია, რომლისთვისაც სრულდება პირობა

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad (4.4)$$

და ყოველი $a \leq b$ რიცხვებისთვის $P[a \leq X \leq b]$ განისაზღვრება ტოლობით

$$P[a \leq X \leq b] = \int_a^b f(x) dx \quad (4.2)$$

ზოგადად, ალბათობის სიმკვრივის ფუნქცია გამოსახულია შემდეგ გრაფიკზე:



გრაფიკი 4.1. ალბათობის სიმკვრივის ზოგადი ფუნქცია

ძალაში დარჩნენ (4.1) და(4.2) ფორმულები. გრაფიკი არ ჩამოდის x ღერძის ქვემოთ, აგრეთვე $y = f(x)$ გრაფიკის ქვევით მოთავსებული მთლიანი ფიგურის ფართობი ტოლია 1-ის და ფიგურის ფართობი ყოველ კონკრეტულ მნიშვნელობებს შორის გვაძლევს ამ ინტერვალების შესაბამის ალბათობებს.

ალბათობის სიმკვრივის მექანიკური ანალოგი

პირდაპირი ანალოგი იმისა, რაც კეთდება მექანიკაში, არის ის, რომ $f(x)$ მართლაც არის „ალბათობის სიმკვრივე“ x -ის მიდამოში, მაშინ ალბათობა dx ზომის მცირე ინტერვალში დაახლოებით არის $f(x)dx$. (მექანიკაში, თუ $f(x)$ გამოსახავს *მასის* სიმკვრივეს x -ის მიდამოში, მაშინ *მასა* dx ზომის მცირე ინტერვალში დაახლოებით არის $f(x)dx$). მაშინ a და b -ს შორის ალბათობის მისაღებად გვჭირდება ასეთი $f(x)dx$ -ების მნიშვნელობათა შეჯამება. $\int_a^b f(x)dx$ კი არის ზუსტად ზღვარი $\sum f(x)dx$ -მნიშვნელობისა, როცა dx ღებულობს მცირე მნიშვნელობებს. (მექანიკაში, $\int_a^b f(x)dx$ არის *მასა* a და b -ს შორის). ამიტომ (4.2) ფორმულა „გონივრულია“.

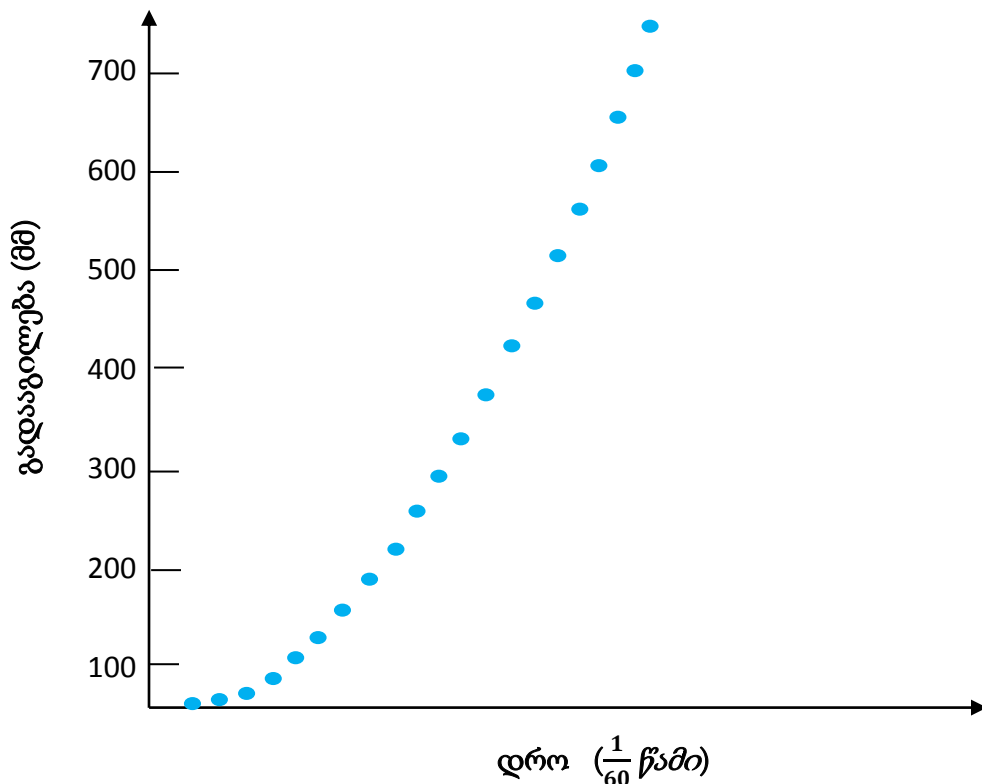
მაგალითი 8. სხეულის თავისუფალი ვარდნა

ტრადიციულად, კოლეჯების ფიზიკის ლაბორატორიებში ხშირად ტარდება ექსპერიმენტები თავისუფლად ვარდნილი სხეულის g აჩქარების გამოთვლაზე. ქვემოთ მოგვყავს ერთ-ერთი ასეთი ექსპერიმენტის ჩატარების დროს მიღებული გაზომვის შედეგები. გაიზომა სხეულის ვერტიკალური გადაადგილება $\frac{1}{60}$ წამის დროის ინტერვალში.

პუნქტის ნომერი	გადაადგილება (მმ)	პუნქტის ნომერი	გადაადგილება (მმ)
1	.8	13	223.8
2	4.8	14	260.0
3	10.8	15	299.2
4	20.1	16	340.5
5	31.9	17	385.0
6	45.9	18	432.2
7	63.3	19	481.8
8	83.1	20	534.2
9	105.8	21	589.8
10	131.3	22	647.7
11	159.5	23	708.8
12	190.5		

ცხრილი 4.2. თავისუფალი ვარდნის გადაადგილების გაზომვები

ამ მონაცემების გრაფიკზე მონიშვნა გვაძლევს პარაბოლის ფორმის ფიგურას, რომელიც გვიჩვენებს რომ არსებობს დაახლოებით კვადრატული დამოკიდებულება გავლილ მანძილსა და დროს შორის.



გრაფიკი 4.3. სხეულის მდებარეობა თავისუფალი ვარდნის დროს

წარმოდგენილი მონაცემებიდან ორჯერ გაწარმოება საშუალებას იძლევა დავასკვნათ, რომ აჩქარება გამოწვეული გრავიტაციით, უხეშად რომ ვთქვათ, არის მუდმივი. შეიძლება ვიფიქროთ, რომ შემთხვევითი სიდიდე

$Y =$ მოძრაობაზე დახარჯული დრო (წამებში)

არის უწყვეტი, რომლის შესაძლო მნიშვნელობათა სიმრავლეა $(0; \frac{1}{60})$.

რა იქნება დამატებული ალბათობის სიმკვრივის ფუნქცია Y -სთვის? იგი შეიძლება ავიღოთ მუდმივი $(0; \frac{1}{60})$ ინტერვალის შიგნით და 0 ამ ინტერვალის გარეთ. ანუ, ნებისმიერი ორი y_1 და y_2 მნიშვნელობებისთვის $(0; \frac{1}{60})$ ინტერვალიდან ალბათობა იმისა, რომ Y ღებულობს მნიშვნელობებს y_1 -ის შემცველი მცირე ინტერვალიდან,

რომლის სიგრძეა dy (ე.ი. $f(y_1)dy$ დაახლოებით) იქნება იგივე, რაც ალბათობა იმისა, რომ Y ღებულობს მნიშვნელობებს y_2 -ის შემცველი მცირე ინტერვალიდან, რომლის სიგრძე იგივე dy (ე.ი. $f(y_2)dy$ დაახლოებით). ამიტომ $f(y_1) = f(y_2)$, და ალბათობის სიმკვრივე $(0; \frac{1}{60})$ ინტერვალზე უნდა იყოს მუდმივი.

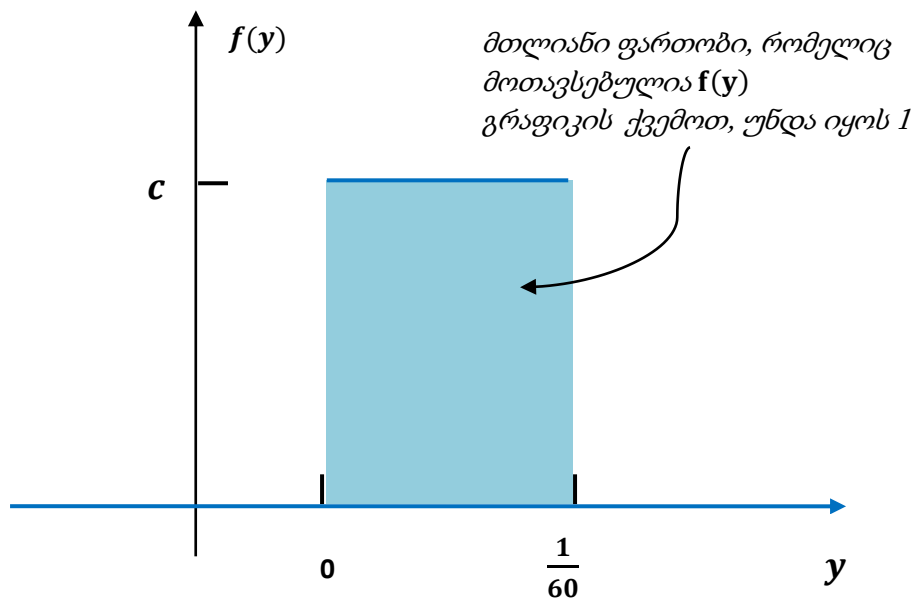
ახლა თუ $f(y)$ -ს აქვს ფორმა

$$f(y) = \begin{cases} c & 0 < y < \frac{1}{60} - \text{თვის} \\ 0 & \text{სხვა შემთხვევაში} \end{cases}$$

რადგან c მუდმივისთვის (ანუ, რაც წარმოდგენილია შემდეგ გრაფიკზე) ფორმულა (4.1) -ს ძალით ეს უნდა იყოს

$$1 = \int_{-\infty}^{+\infty} f(y)dy = \int_{-\infty}^0 0 \cdot dy + \int_0^{1/60} c \, dy + \int_{1/60}^{+\infty} 0 \cdot dy = \frac{c}{60},$$

ანუ $c = 60$ და $f(y)$ -ს აქვს ასეთი სახე:



გრაფიკი 4.4. ალბათობის სიმკვრივის ფუნქცია Y -სთვის.

$$f(y) = \begin{cases} 60, & 0 < y < \frac{1}{60} - \text{თვის} \\ 0, & \text{სხვა შემთხვევაში} \end{cases} \quad (4.3)$$

თუ ამ ფორმულით განსაზღვრული ფუნქცია მიღებულია ალბათობის სიმკვრივის ფუნქციად Y -თვის, მაშინ, მაგალითად, შესაძლებელია გამოვთვალოთ

$$P\left[Y \leq \frac{1}{100}\right] = \int_{-\infty}^{1/100} f(y)dy = \int_{-\infty}^0 0 \cdot dy + \int_0^{1/100} 60dy = .6$$

უწყვეტი შემთხვევითი სიდიდისთვის $P[X = a] = 0$

არის ერთი მნიშვნელოვანი მომენტი უწყვეტი ტიპის ალბათობის განაწილებების შესწავლისას. პირველი შეხედვით გვეჩვენება, თითქოს გვაქვს ინტუიციური პრობლემა. იგი უკავშირდება უწყვეტი შემთხვევითი სიდიდის მიერ წინასწარ დასახელებული მნიშვნელობის (ვთქვათ, a) მიღებას. ისევე როგორც, მასის უწყვეტი განაწილების დროს, მისი მნიშვნელობა რაიმე ერთ კონკრეტულ წერტილში არის 0, აქაც ასევე, უწყვეტი ტიპის X შემთხვევითი სიდიდის დროს გვაქვს, რომ $P[X = a] = 0$. ეს გამომდინარეობს (4.3) ფორმულიდანაც, რადგან

$$P[a \leq X \leq b] = \int_a^b f(x)dx = 0$$

მათემატიკური კურიოზის კიდევ ერთი შედეგია ის, რომ როდესაც მუშაობთ უწყვეტ შემთხვევით სიდიდეებთან, არ გჭირდებათ ზრუნვა იმაზე, თუ როგორი უტოლობა წერია, მკაცრი თუ არამკაცრი. ანუ, თუ X არის უწყვეტი,

$$P[a \leq X \leq b] = P[a < X \leq b] = P[a \leq X < b] = P[a < X < b].$$

წინა თავში მოცემული განმარტება 4.1 ნამდვილად არის შემთხვევითი სიდიდის დაგროვების ალბათობის ფუნქციის სრულყოფილი განმარტება (რომელიც განსაზღვრული იყო დისკრეტული შემთხვევითი სიდიდის შემთხვევაში). ამ შემთხვევაში ტოლობა (4.2) შეიძლება გამოყენებულ იქნეს უწყვეტი შემთხვევითი სიდიდის დაგროვების ალბათობის ფუნქციის გამოსახვისთვის ალბათობის სიმკვრივისა და ინტეგრალის ტერმინებში. ანუ, თუ X არის უწყვეტი, რომლის ალბათობის სიმკვრივეა $f(x)$, მაშინ

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(t)dt \quad (4.4)$$

უწყვეტი შემთხვევითი სიდიდის დაგროვების ალბათობის ფუნქცია

$F(x)$ მიღებულია $f(x)$ -სგან ინტეგრებით და კალკულუსის ძირითადი თეორემის გამოყენებით. წინა ტოლობიდან კი ვღებულობთ, რომ

$$\frac{d}{dx} F(x) = f(x) \quad (4.5)$$

მეორენაირი დამოკიდებულება $F(x)$ და $f(x)$ შორის.

ანუ, $f(x)$ მიიღება $F(x)$ -დან დიფერენცირებით.

მაგალითი 8. (გაგრძელება)

დაგროვების ალბათობის ფუნქცია Y -სთვის იოლად მიიღება (4.3) ტოლობიდან. იმ შემთხვევაში, როცა $y \leq 0$,

$$F(y) = P[Y \leq y] = \int_{-\infty}^y f(t)dt = \int_{-\infty}^y 0dt = 0$$

და $0 < y \leq \frac{1}{60}$ - სთვის,

$$F(y) = P[Y \leq y] = \int_{-\infty}^y f(t)dt = \int_{-\infty}^y 0dt + \int_0^y 60dt = 0 + 60y = 60y$$

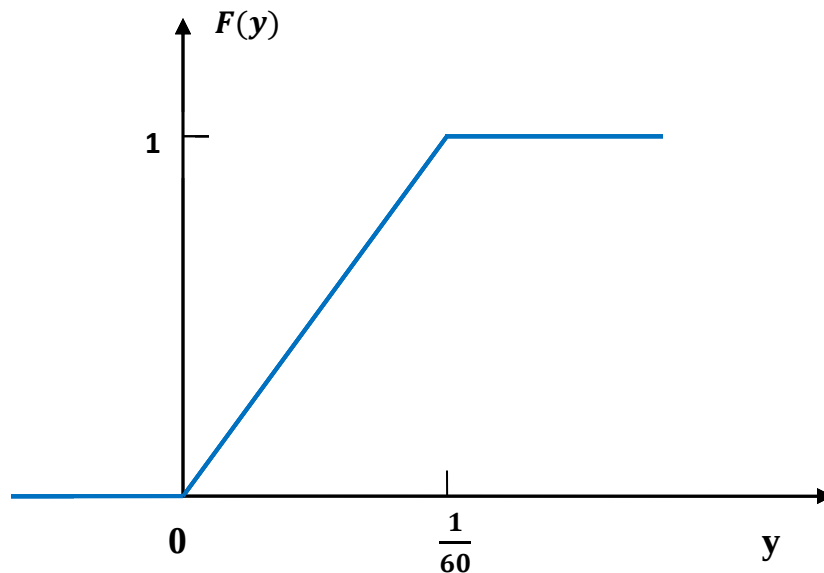
და $y > \frac{1}{60}$ - სთვის,

$$F(y) = P[Y \leq y] = \int_{-\infty}^y f(t)dt = \int_{-\infty}^0 0dt + \int_0^{1/60} 60dt + \int_{1/60}^y 0dt = 1.$$

საბოლოოდ:

$$F(y) = \begin{cases} 0 & \text{თუ } y \leq 0 \\ 60y & \text{თუ } 0 < y \leq 1/60 \\ 1 & \text{თუ } \frac{1}{60} < y \end{cases}$$

$F(y)$ -ის გრაფიკი მოცემულია შემდეგ ნახაზზე. შეადარეთ იგი გრაფიკს 4.4 და ნახეთ, რომ $F(y)$ -ის დახრილობა მართლაც არის ნული, როცა $y < 0$ და $y > \frac{1}{60}$, და დახრილობაა 60, როცა $0 < y < \frac{1}{60}$. ეს კი თავის მხრივ ნიშნავს, რომ $f(y)$ არის $F(y)$ ფუნქციის წარმოებული, რასაც გვპირდებოდა (4.5) ფორმულა.



გრაფიკი 4.5. დაგროვებითი ალბათობის ფუნქცია Υ -სთვის.

ეს გრაფიკი არის ტიპური დაგროვებითი ალბათობის ფუნქცია უწყვეტი ტიპის შემთხვევითი სიდიდეებისთვის. ასეთი დაგროვებითი ალბათობის ფუნქციის ყველა გრაფიკი არის უწყვეტი, რაც ჩანს კიდევ ნახაზზე.

4.2. უწყვეტი განაწილების საშუალო და დისპერსია

ალბათობის სიმკვრივის $f(x)$ -ის გრაფიკი არის იდეალიზებული ჰისტოგრამის სახის. მის ვიზუალურ ინტერპრეტაციას აქვს იგივე სახე, რაც ფარდობითი სიხშირის ჰისტოგრამას და ალბათობის ჰისტოგრამას. ამის შემდგომ შესაძლებელია

განისაზღვროს საშუალო და დისპერსია ალბათობის უწყვეტი განაწილებისთვისაც. ისევე როგორც დისკრეტული განაწილების შემთხვევაში, აქაც საშუალო და დისპერსია გამოიყენება მონაცემთა სიმრავლეების აღწერისთვის და მათი რიცხვითი შეჯამებებიც მიმდინარეობს ისევე როგორც დისკრეტულ შემთხვევაში.

განმარტება 4.2. X უწყვეტი შემთხვევითი სიდიდის საშუალო ან მოსალოდნელი მნიშვნელობა (ზოგჯერ ვუწოდებთ ალბათობის განაწილების საშუალოს) არის

$$EX = \int_{-\infty}^{\infty} xf(x)dx \quad (4.6)$$

ისე როგორც დისკრეტულ შემთხვევაში, EX -ის მაგიერ გამოიყენება აღნიშვნა μ .

(ქართულ ენაზე არსებულ ლიტერატურაში ამ სიდიდეს უწოდებენ *მათემატიკურ ლოდინს*. მთარგმნელის შენიშვნა).

ეს უკანასკნელი ფორმულა სავსებით დამაჯერებელია სულ მცირე ორი მიზეზის გამო. პირველი, ალბათობა x -ის მცირე მიდამოში (ინტერვალში), რომლის სიგრძეა dx , დაახლოებით არის $f(x)dx$. თუ მას გავამრავლებთ x -ზე და შევკრებთ მივიღებთ $\sum x f(x)dx$, ხოლო ფორმულა (4.6) არის ზუსტად ამ გამოსახულების ზღვარი, როცა dx უსასრულოდ მცირდება. მეორე, მექანიკაში მასის უწყვეტი განაწილების დროს მასის ცენტრი მოიცემა (4.6) ფორმულით, იმ განსხვავებით, რომ გაყოფა ხდება მთლიან მასაზე, რაც ალბათობის განაწილების შემთხვევაში არის 1.

მაგალითი 8. (გაგრძელება)

თუ დავაკვირდებით გრაფიკს 4.4, ალბათობის სიმკვრივე არის იდეალიზებული ჰისტოგრამა და თუ ვიფიქრებთ, რომ საშუალო გამოსახავს წონასწორობის წერტილს, მაშინ ცხადია, რომ EY აღმოჩნდება $\frac{1}{120}$ გასული დროის ცვლადისთვის. საბედნიეროდ, იმავეს გვაძლევენ (4.6) და (4.3) ტოლობებიც:

$$\mu = EY = \int_{-\infty}^{+\infty} y f(y) dy =$$

$$\int_{-\infty}^0 y \cdot 0 dy + \int_0^{1/60} y \cdot 60 dy + \int_{1/60}^{+\infty} y \cdot 0 dy =$$

$$= 30y^2 \Big|_0^{1/60} = \frac{1}{120} \text{ წამი}$$

დისკრეტული შემთხვევითი სიდიდის დისპერსიის ფორმულის „გაგრძელებადობას“ მიყვავართ უწყვეტი შემთხვევითი სიდიდის დისპერსიის ცნებამდე.

განმარტება 4.3. X უწყვეტი შემთხვევითი სიდიდის დისპერსია (ზოგჯერ მას ალბათობის განაწილების დისპერსიასაც უწოდებენ) არის

$$Var X = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx \quad (= \int_{-\infty}^{+\infty} x^2 f(x) dx - (EX)^2) \quad (4.7)$$

X -ის სტანდარტული გადახრა არის $\sqrt{Var X}$. ხშირად $Var X$ -ს აღნიშნავენ σ^2 -თ, ხოლო $\sqrt{Var X}$ -ს σ -თი.

მაგალითი 8. (გაგრძელება)

დავუბრუნდეთ სხეულის თავისუფალი ვარდნის მაგალითს და Y შემთხვევით სიდიდეს. (4.7) ფორმულის გამოყენებით, და Y შემთხვევით სიდიდის ალბათობის სიმკვრივიდან ვღებულობთ, რომ

$$\begin{aligned} \sigma^2 = Var Y &= \int_{-\infty}^0 \left(y - \frac{1}{120} \right)^2 \cdot 0 dy + \int_0^{1/60} \left(y - \frac{1}{120} \right)^2 \cdot 60 dy + \\ &+ \int_{1/60}^{\infty} \left(y - \frac{1}{120} \right)^2 \cdot 0 dy = \frac{60 \left(y - \frac{1}{120} \right)^3 \Big|_0^{1/60}}{3} = \frac{1}{3} \left(\frac{1}{120} \right)^2 \end{aligned}$$

ანუ Y -ის სტანდარტული გადახრა არის

$$\sigma = \sqrt{\text{Var } Y} = \sqrt{\frac{1}{3} \left(\frac{1}{120} \right)^2} = .0048 \text{ წმ}$$

4.3. ალბათობის ნორმალური განაწილება

არსებობს მთელი რიგი დისკრეტული სტანდარტული განაწილებების, რომლებიც ხშირად გამოიყენებიან საინჟინრო ამოცანების გადაწყვეტისას. დისკრეტულთან ერთად არსებობს აგრეთვე ბევრი უწყვეტი ტიპის სტანდარტული განაწილებებიც, რომლებიც ძალიან საჭირო დახმარებას გვიწევენ საინჟინრო საქმეში. შესავალი სიტყვები მიგვანიშნებს, რომ გვინდა შევისწავლოთ ნორმალური ანუ გაუსის განაწილებები და გამოვსახოთ მათი თვისებები გრაფიკების საშუალებით. ახლა კი დროა, რომ შემოვიტანოთ მისი ფორმალური განსაზღვრება.

განმარტება 4.4. ნორმალური ანუ გაუსის (μ, σ^2) განაწილება ეწოდება ალბათობის ისეთ უწყვეტ განაწილებას, რომლის სიმკვრივე გამოისახება ფორმულით:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (4.8)$$

ყოველი x -სთვის და $\sigma > 0$ -სთვის.

ფორმულა (4.8) აკმაყოფილებს ალბათობის სიმკვრივის ყველა მოთხოვნას, კერძოდ, ის რომ $y = f(x)$ გრაფიკის ქვემოთ მოთავსებული ფიგურის ფართობი არის 1, თუმცა ეს ფაქტი არ არის თვალსაჩინო. ამის გარდა, აგრეთვე გვაქვს, რომ

ნორმალური განაწილების საშუალო და დისპერსია არის

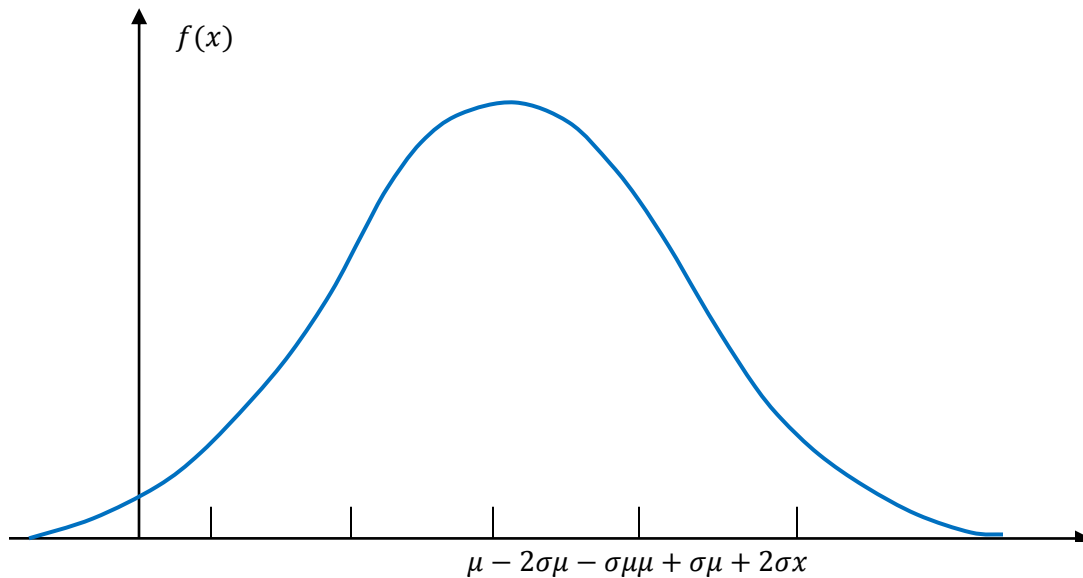
$$EX = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \mu$$

და

$$\text{Var } X = \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2$$

სადაც μ და σ^2 ზუსტად ის პარამეტრებია, რომლებიც გამოყენებული იყო განმარტებაში 4.4. შესაბამისად, განაწილების საშუალო და დისპერსია განსაზღვრულია ისე, როგორც განმარტებებში 4.6 და 4.7.

ქვემოთ მოცემულ ნახაზზე მოცემულია ალბათობის სიმკვრივის გრაფიკი, რომელიც განსაზღვრულია (4.8) ფორმულით.



ნახაზი 4.6. ნორმალური განაწილების ალბათობის სიმკვრივის ფუნქციის გრაფიკი.

როგორც ვხედავთ, ზარის - ფორმის წირის გრაფიკი სიმეტრიულია $x = \mu$ -ს მიმართ და აქვს გადაღუნვის წერტილები $\mu - \sigma$ და $\mu + \sigma$ მნიშვნელობებში. (4.8) ფორმულის ზუსტი სახის დადგენას სჭირდება მრავალი თეორიული წანამძღვარი. ეს ფორმულა აღმოჩნდა აგრეთვე ძალიან სასარგებლო ემპირიული თვალსაზრისითაც, ვინაიდან მას გააჩნია მრავალი სხვადასხვა გამოყენება.

თეორიაში, ნორმალური განაწილების ალბათობა შესაძლებელია უშუალოდ გამოვთვალოთ ინტეგრებით, (4.8) ფორმულის გამოყენებით. მართლაც, მკითხველს შეუძლია იმ კალკულატორის გამოყენებით, რომელშიც ჩატვირთულია რიცხვითი ინტეგრების პროგრამა, ნახოს ინსტრუქცია და შეამოწმოს მისი მუშაობა შემდგომში

ზოგიერთ მაგალითზე (4.2) და (4.8) ფორმულების პირდაპირ გამოყენებაზე. მაგრამ პირველი კურსის კალკულუსის მეთოდების გამოყენებით, ინტეგრალის გამოთვლა პირველადი ფუნქციების საშუალებით, წარმატებას ვერ მოგვიტანს ნორმალური სიმკვრივისთვის. ამ ფუნქციებს ელემენტარულ ფუნქციათა კლასში არ გააჩნიათ პირველადი ფუნქციები. ამიტომ ინტეგრების მაგიერ ვიყენებთ ე.წ. ნორმალური ალბათობების ცხრილებს, რომლებსაც როგორც წესი, გააჩნიათ ფართო გამოყენებები.

ცხრილების გამოყენება ნორმალური ალბათობების გამოსათვლელად დამოკიდებულია შემდეგ თანაფარდობაზე. თუ X არის ნორმალურად განაწილებული შემთხვევითი სიდიდე, რომლის საშუალოა μ და დისპერსია σ^2 ,

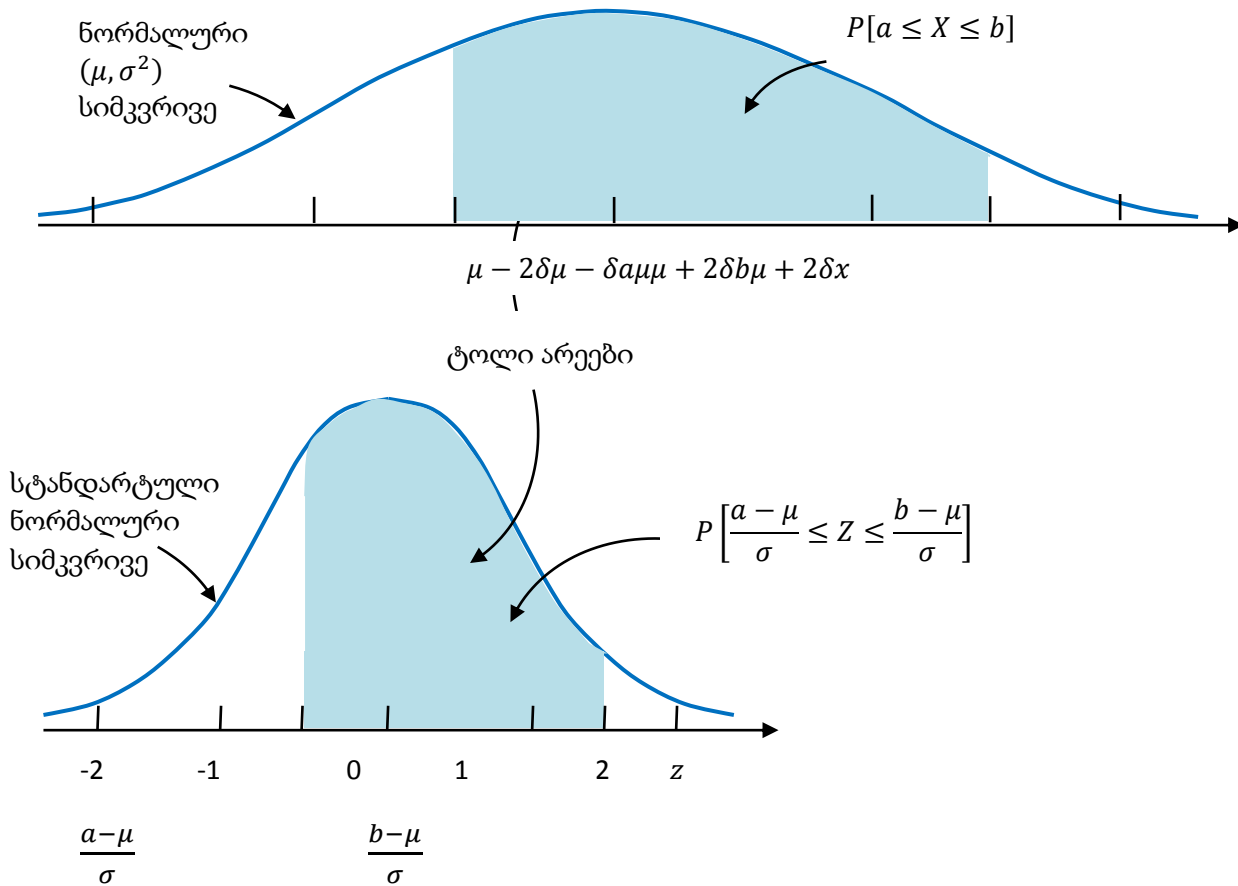
$$P[a \leq X \leq b] = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

სადაც მეორე ტოლობა გამომდინარეობს ცვლადის გარდაქმნით ანუ ჩასმოდან $z = \frac{x-\mu}{\sigma}$. ტოლობა (4.9) შეიცავს ინტეგრალს ნორმალური განაწილების სიმკვრივიდან, როცა $\mu = 0$ და $\sigma = 1$. ეს კი გვეუბნება, რომ ნებისმიერი ნორმალური ალბათობის გამოთვლა შეიძლება დავიყვანოთ სპეციალური სახის ნორმალური ალბათობის გამოთვლაზე.

განმარტება 4.4. ნორმალურ განაწილებას $\mu = 0$ და $\sigma = 1$ სიდიდეებით ეწოდება სტანდარტული ნორმალური განაწილება.

დამოკიდებულება ნორმალურ (μ, σ^2) ალბათობებსა და სტანდარტულ ნორმალურ ალბათობებს შორის

დამოკიდებულება ნორმალურ (μ, σ^2) ალბათობებსა და სტანდარტულ ნორმალურ ალბათობებს შორის გამოსახულია შემდეგ ნახაზზე.



ნახაზი 4.7. ნორმალური (μ, σ^2) ალბათობებისა და სტანდარტულ ნორმალურ ალბათობებს შორის დამოკიდებულების გრაფიკული ილუსტრაცია

პირველი მათგანი არის რეალიზაცია ყველა ნორმალური განაწილების ალბათობების, რომელთა დატაბულებაც შესაძლებელია სტანდარტული ნორმალური განაწილებით. გაცილებით მარტივი საკითხია, გამოვიყენოთ რიცხვითი ინტეგრების ტექნიკა და შევადგინოთ სტანდარტული ნორმალური განაწილების ცხრილი. ცხრილს, რომელსაც ჩვენ გამოვიყენებთ, მოცემულია დანართში **ცხრილი B.3**. (არსებობს სხვა ფორმის ცხრილებიც).

ეს არის ცხრილი სტანდარტული ნორმალური განაწილების დაგროვებითი ალბათობის ფუნქციის მნიშვნელობათა გამოსათვლად. ამ დროს z -ით აღნიშნულია ინტეგრალის ცვლადი ზედა საზღვარი, ხოლო თვით ინტეგრალი კი ასე გამოიყურება:

$$\Phi(z) = F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

(Φ ასოთი ზოგადად აღინიშნება სტანდარტული ნორმალური განაწილების დაგროვებითი ალბათობის ფუნქცია, უფრო ზოგადი $F(z)$ ფუნქციის მაგიერ).

მაგალითი 9. სტანდარტული ნორმალური ალბათობები

წარმოვიდგინოთ, რომ Z არის სტანდარტული ნორმალური შემთხვევითი სიდიდე. ჩვენ გამოვთვლით ზოგიერთ ალბათობას Z შემთხვევითი სიდიდისთვის B.1 ცხრილის გამოყენებით.

მოდით, ყველაფერი მოვძებნოთ ცხრილში,

$$P[Z < 1.76] = \Phi(1.76) = .96$$

(ცხრილის მნიშვნელობაა .9608, მაგრამ თუ კვლავ ალბათობის საბოლოო მნიშვნელობაში შევინახავთ მხოლოდ ორ ათწილად ნიშანს, ცხრილის მნიშვნელობის დამრგვალებით ვღებულობთ .96-ს.) ცხრილის სხვადასხვა მნიშვნელობათა გამოკლებით კი მივიღებთ, რომ

$$\begin{aligned} P[.57 < Z < 1.32] &= P[Z < 1.32] - P[Z \leq .57] = \\ &= \Phi(1.32) - \Phi(.57) = .9066 - .7157 = .19 \end{aligned}$$

თუ ერთხელ კიდევ გამოვიყენებთ ცხრილს, ვნახავთ, გამოკლების კიდევ ერთ გამოყენებას მარჯვენა - კუდის ალბათობის გამოსათვლელად, შემდეგნაირად:

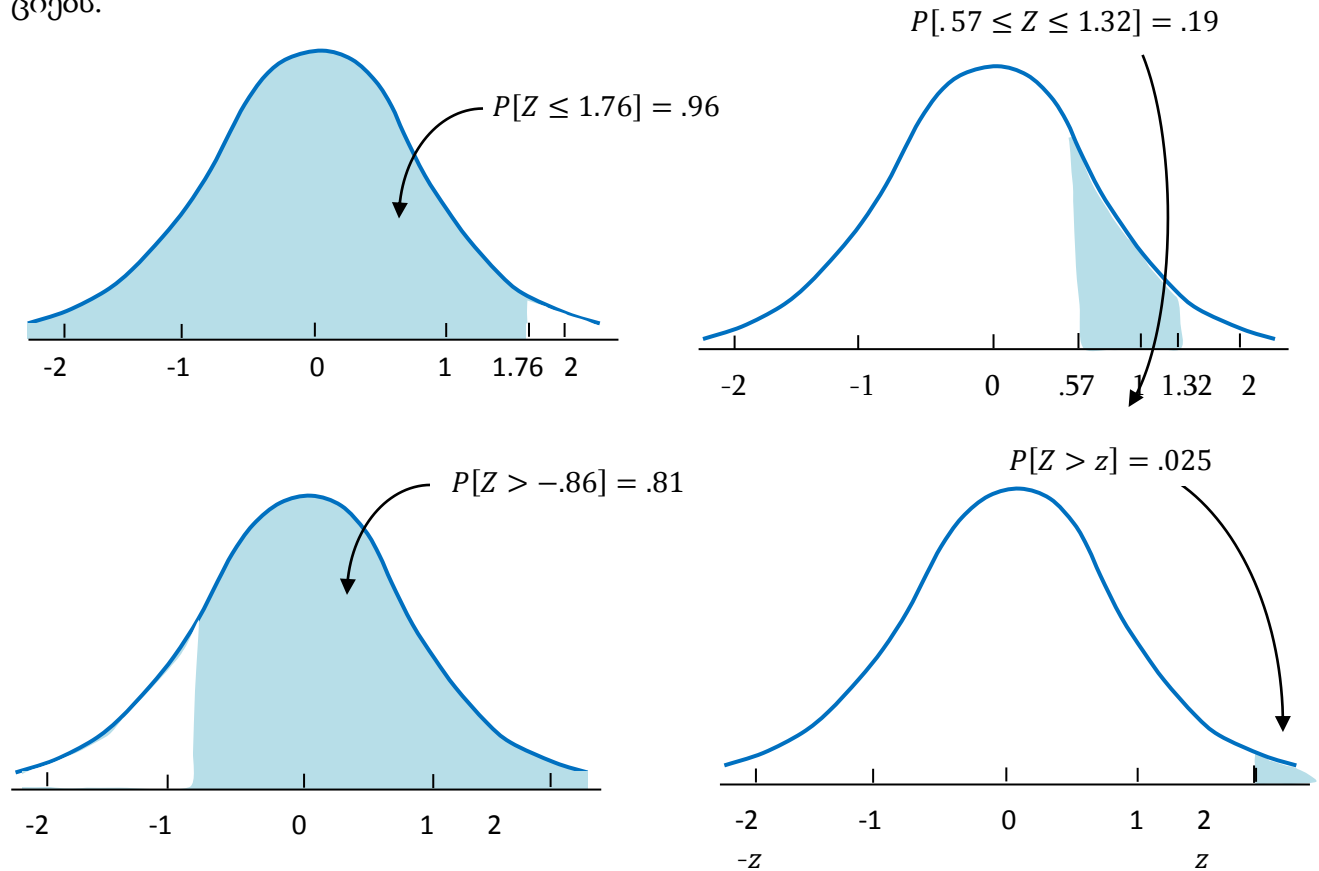
$$P[Z > -.89] = 1 - P[Z \leq -.89] = 1 - .1867 = .81$$

ამ სამ მაგალითში ცხრილი გამოვიყენეთ z -ის მნიშვნელობების ალბათობების გამოსათვლელად. z -ის ეს მნიშვნელობები განთავსებულია ცხრილის კიდეში ხოლო ალბათობები უნდა ვეძებოთ ცხრილის სიღრმეში. პროცესი შესაძლოა წარიმართოს შექცეულადაც. ანუ, ალბათობებით, რომლებიც განთავსებულნი არიან ცხრილის სიღრმეში, შეიძლება განვსაზღვროთ z -ის ეს მნიშვნელობები ცხრილის კიდეში. მაგალითად, განვსაზღვროთ z -ის მდებარეობა რომლისთვისაც

$$P[-z < Z < z] = .95$$

მაშინ z მიიღებს მნიშვნელობას $\frac{1-.95}{2} = .025$ სტანდარტული ნორმალური განაწილების მარჯვენა კუდში. მაშინ $\Phi(z) = .975$. თუ მოვნახავთ ცხრილში .975 -ს, ვნახავთ, რომ იგი შეესაბამება $z = 1.96$ მნიშვნელობას.

შემდეგი გრაფიკები წარმოადგენენ ჩატარებული გამოთვლების ნათელ ილუსტრაციებს.



ნახაზი 4.8. სტანდარტული ნორმალური ალბათობები მაგალითი 9-სთვის.

მაგალითი 9-ის ბოლო გამოთვლები ტოლფასია .975 კვანტილის პოვნისა სტანდარტული ნორმალური განაწილებისთვის. მკითხველს უკვე შეუძლია გაიგოს ცხრილი B.2-ის წარმოშობა. სტანდარტული ნორმალური კვანტილები ნაპოვნია ცხრილი B.1-დან (სიღრმეში) შესაბამისი ალბათობებით და შესაბამისი z -ებით, რომლებიც მოთავსებულნი არიან ცხრილის კიდეში.

მათემატიკურ სიმბოლოებში სტანდარტული ნორმალური დაგროვებითი ალბათობის $\Phi(z)$ ფუნქციისა და სტანდარტული ნორმალურ კვანტილის ფუნქციას z შორის

$$\begin{cases} \Phi(Q_z(p)) = p \\ Q_z(\Phi(z)) = z \end{cases}^{(4.10)}$$

დამოკიდებულება ნიშნავს, რომ Q_z და Φ არიან შერეული ფუნქციები. (ეს დამოკიდებულება $Q = \Phi^{-1}$ არ არის მოვლენა მხოლოდ სტანდარტული ნორმალური შემთხვევისთვის, არამედ ჭეშმარიტია ზოგადად უწყვეტი განაწილებებისთვის, ანუ სამართლიანია $Q = F^{-1}$.)

დამოკიდებულება (4.9) გვიჩვენებს, როგორ გამოვიყენოთ სტანდარტული ნორმალური დაგროვებითი ფუნქცია ზოგადი ნორმალური ალბათობების საპოვნელად. X ნორმალური (μ, σ^2) შემთხვევითი სიდიდისთვის x შემთხვევითი მნიშვნელობა, რომელიც დაკავშირებულია X -სთან, შეიძლება გარდავქმნათ სტანდარტულ ნორმალურ შემთხვევით სიდიდედ საშუალოსა სტანდარტული გადახრის საშუალებით შემდეგნაირად:

$$z = \frac{x - \mu}{\sigma} \quad (4.11)$$

ეს არის z-მნიშვნელობა ნორმალური (μ, σ^2) შემთხვევითი სიდიდის x შემთხვევითი მნიშვნელობისთვის.

ამის მერე შეგვიძლია მივმართოთ სტანდარტულ ნორმალურ ცხრილს და x -ის მაგიერ გამოვიყენოთ z .

მაგალითი 10. ბავშვის საკვების სუფთა წონა ქილებში

ჯ. ფიშერი, თავის სტატიაში „სუფთა წონის კონტროლი კომპიუტერის დახმარებით“ (Quality Progress, June 1983), განიხილავს ქილის ბავშვის საკვებით შევსების წონას. სტატიაში ნათქვამია, რომ გონივრულია ზარის-ფორმის ჰისტოგრამა ინდივიდუალური სუფთა წონებისთვის, მაგალითად, ქლიავის ფაფისთვის. წონების საშუალო მნიშვნელობა ტოლია დაახლოებით 137.2 გრამის, ხოლო სტანდარტული გადახრა კი - დაახლოებით 1.6 გრ. ქილის ეტიკეტზე აღნიშნული წონაარის 135.0 გრ.

დავუშვათ, რომ ამ ყველაფრის შესატყვისი მოდელია

$W =$ შემდეგ ქილაში ქლიავის ფაფის წონა

ნორმალური განაწილება, რომლის $\mu = 137.2$ და $\sigma = 1.6$. შემდგომ, ალბათობა იმისა, რომ მომდევნო ქილაში აღმოჩნდება ეტიკეტზე აღნიშნულზე ნაკლები წონა,

აღვნიშნოთ $P[W < 135.0]$ -თი და სწორედ ეს წარმოადგენს ჩვენს ინტერესს. (4.11) ფორმულის გამოყენებით $w = 135.0$ გარდავქმნათ სტანდარტული გადახრის ერთეულებში ისე, როგორც ეს გავაკეთეთ μ -სთვის, როცა ის გარდავქმენით z -ცვლადად, ასე

$$z = \frac{135.0 - 137.2}{1.6} = -1.38,$$

მაშინ, ცხრილი B.1 -ის გამოყენებით გვაქვს:

$$P[W < 135.0] = \Phi(-1.38) = .08$$

ეს მოდელი უშვებს შესაძლებლობას, რომ ეტიკეტზე აღნიშნულზე ნაკლები წონის ქილის შეხვედრის შანსი არის 8%.

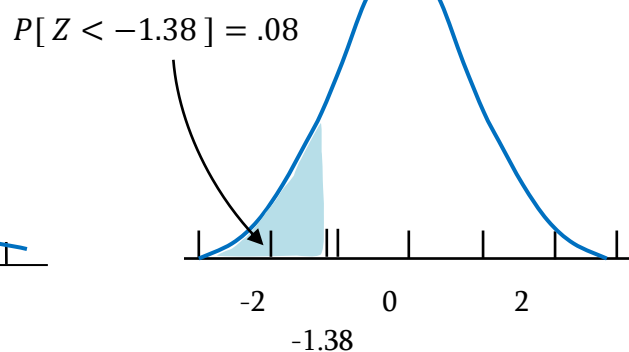
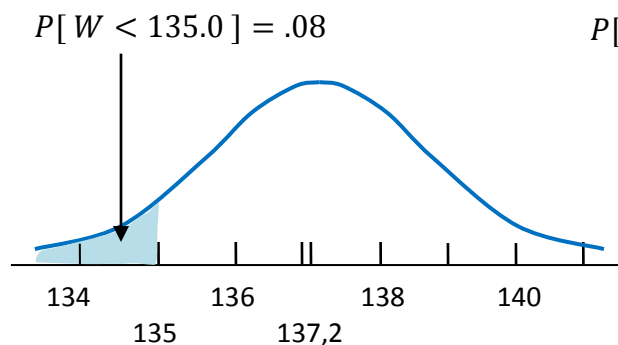
მეორე მაგალითად განვიხილოთ, ალბათობა იმისა, რომ W არის ნომინალიდან 1 გრამით გადახრილი (ანუ, $P[134.0 < W < 136]$). ფორმულა (4.11) გამოყენებით შესაძლებელია $w_1 = 134.0$ და $w_2 = 136.0$ გარდაიქმნას z -ცვლადად ანუ სტანდარტული გადახრის ერთეულებში, ისე როგორც ეს ზემოთ გავაკეთეთ საშუალოსთვის

$$z_1 = \frac{134.0 - 137.2}{1.6} = -2.0$$

$$z_2 = \frac{136.0 - 137.2}{1.6} = -.75$$

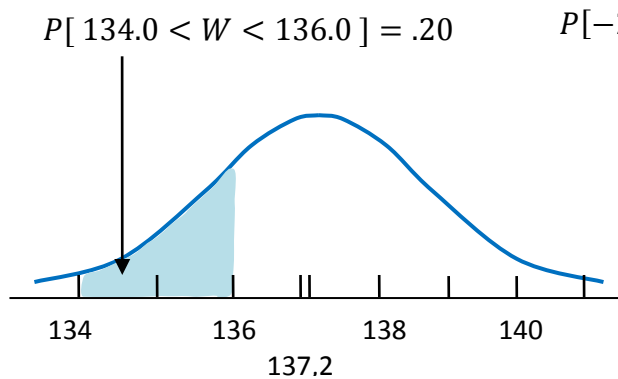
ნორმალური $\mu = 137.2$, $\sigma = 1.6$ სიმკვრივე

სტანდარტული ნორმალური
სიმკვრივე

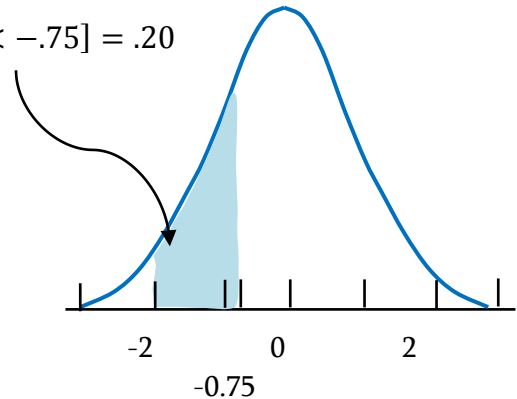


ნორმალური $\mu = 137.2$, $\sigma = 1.6$ სიმკვრივე

სტანდარტული ნორმალური
სიმკვრივე



$P[-2.0 < Z < -0.75] = .20$



ნახაზი 4.9. ნორმალური ალბათობები მაგალითი 10-სთვის

მაშინ,

$$P[134.0 < W < 136.0] = \Phi(-.75) - \Phi(-2.00) = .2266 - .0228 = .20$$

წინა ორი ალბათობა და მათი სტანდარტული ნორმალური „კოლეგები“ ნაჩვენებია ნახაზზე.

როგორც ვნახეთ, ამ მაგალითების გამოთვლები ეყრდნობოდა (4.11) ფორმულას და ცხრილი B.1-ის კიდურა სვეტიდან გადავინაცვლეთ მის სიღრმეში W -ს ალბათობების საპოვნელად. მნიშვნელოვანი პროცესი იყო აგრეთვე ისიც, რომ ცხრილის შუაგულიდან გადავინაცვლეთ კიდეში, როცა გვინდოდა z -ის მიღება. შემდეგ სამი კვანტილიდან, მხოლოდ მოცემული ორის საშუალებით ვიპოვეთ მესამე.

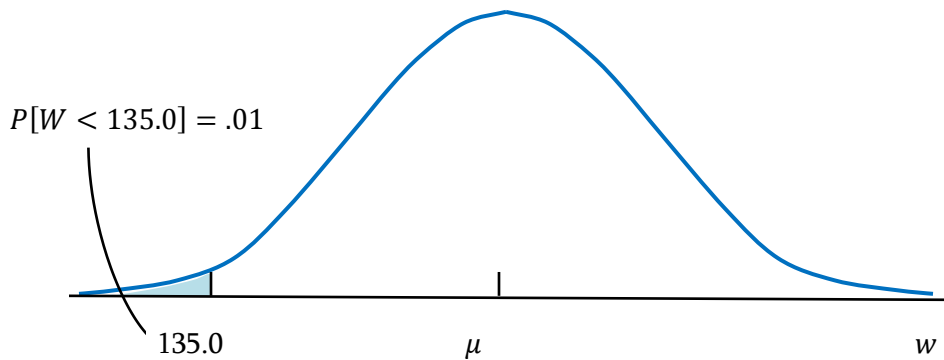
წარმოიდგინოთ, რომ ადვილია ვარეგულიროთ სიტუაცია და შევცვალოთ ქილის შევსების პროცესი (ანუ, W -ს საშუალო μ). გვინდა, შემცირდეს ალბათობა იმისა, რომ შემდგომი ქილის წონა იქნება ნაკლები მითითებულ წონაზე, 135.0 გრამზე იყოს .01. ამის მიღწევა შესაძლებელია μ -ს გაზრდით. რა არის μ -ს მინიმალური მნიშვნელობა,

როდესაც შესაძლებელია ამის მიღწევა? (ვიგულისხმობთ, რომ σ რჩება იგივე, ანუ 1.6 გრ.)

ნახაზი 4.9 გვიჩვენებს რა უნდა ვქნათ. μ უნდა ავირჩიოთ ისე, რომ $w = 135.0$ ხდება .01 კვანტილი იმ ნორმალური განაწილებისა, რომლის საშუალო μ და სტანდარტული გადახრა $\sigma = 1.6$. შემდეგ ცხრილი B.2-ის ან B.1-ის გამოყენებით ადვილია განვსაზღვროთ .01 კვანტილი სტანდარტული ნორმალური განაწილებისთვის. იგი არის

$$z = Q_z(.01) = -2.33$$

ნორმალური სიმკვრივე $= \mu, \sigma = 1.6$



ნახაზი 4.10. ნორმალური განაწილება და $P[W < 135.0] = .01$

ასე, რომ (4.11) ტოლობის შუქზე გვინდა, რომ გვქონდეს

$$-2.33 = \frac{135.0 - \mu}{1.6}$$

ანუ,

$$\mu = 138.7 \text{ გრ.}$$

ამგვარად, შევსების დონის გაზრდა $138.7 - 137.2 = 1.5$ გრამით, ნიშნავს, რომ მიზანი მიღწეული იქნება.

პრაქტიკული ტერმინებით რომ ვისაუბროთ, $P[W < 135.0]$ ალბათობის შემცირება ნიშნავს, რომ ბავშვის საკვების შესაძენი ფასი იზრდება, რადგან საშუალო ფასი

დამოკიდებულია ქილის შევსების დონეზე, ვინაიდან იგი შეიცავს გაცილებით მეტს, ვიდრე მისი ნომინალური შემცველობაა. ზოგიერთ შემთხვევებში ასეთნაირად ფასის დადება შეზღუდულია. არსებობს სხვა მიდგომა, ამჯერად ინჟინრული. ეს არის, რომ უნდა შემცირდეს ქილის შევსების დონის ცვალებადობა, რომელიც თავის მხრივ საჭიროებს უფრო მაღალი სიზუსტის შემავსებელი მოწყობილობის შეძენას. კვლავ (4.11) ფორმულაზე დაყრდნობით μ -ს ზრდის მაგიერ შეგვიძლია განვიხილოთ ფასის დამოკიდებულება σ -ს შემცირებაზე. ვიმედოვნებთ, მკითხველი ხალისით შეამოწმებს, რომ σ -ს შემცირება 0.94-მდე აგრეთვე გვაძლევს $P[W < 135.0] = .01$ -ს μ -ს ყოველგვარი ცვლილების გარეშე.

როგორც მაგალითი 10 გვიჩვენებს, ტოლობა (4.11) არის ფუნდამენტური დამოკიდებულება, რომელიც გამოიყენება ამოცანებში, სადაც მონაწილეობს ნორმალური განაწილება. თუ (4.11) ფორმულაში მონაწილე ოთხი მონაცემიდან ამა თუ იმ გზით განსაზღვრულია სამი მათგანი, მაშინ შეგვიძლია მეოთხის პოვნაც.

4.4. მაჩვენებლიანი განაწილება (ნებაყოფლობით)

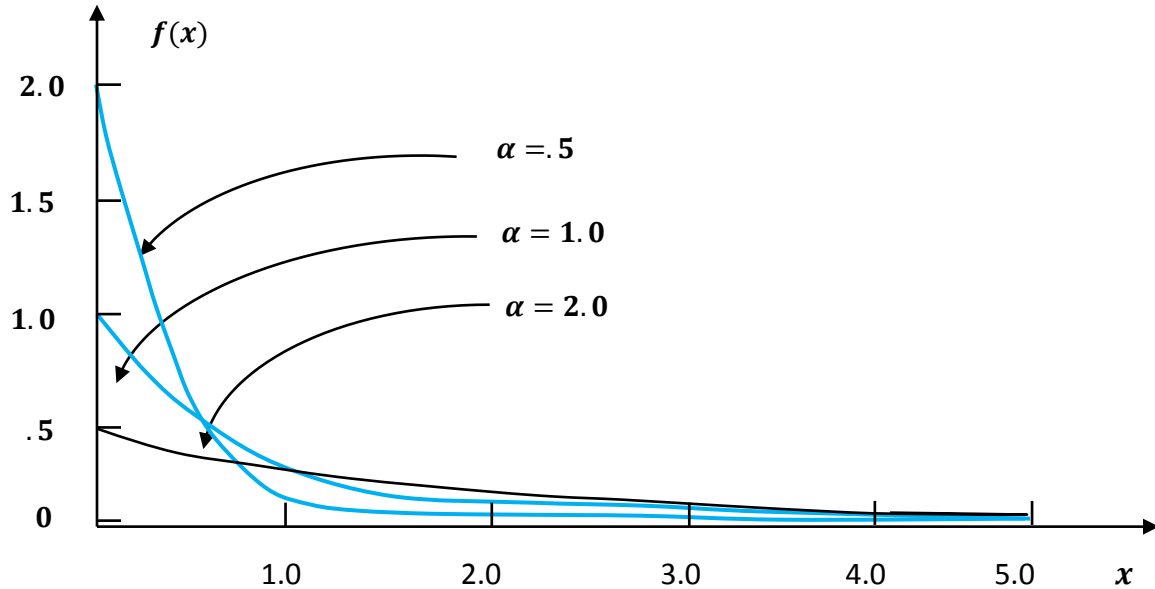
როგორც გვახსოვს, დისკრეტული შემთხვევითი სიდიდეების შესწავლისას, პუასონის განაწილება გამოიყენებოდა ისეთი მოდელების შესასწავლად, როდესაც ხდომილების მოხდენათა რაოდენობა არის შედარებით იშვიათი მოვლენა დროის განსაზღვრულ ინტერვალში. იგივე მათემატიკური თეორია, რომელიც თვლიდა პუასონის განაწილების მიზნშეწონილობას იმ კონტექსტში, ახლაც მიზნშეწონილად თვლის, რომ მეტად სასარგებლოა მაჩვენებლიანი (ან რაც იგივეა - ექსპონენციალური) განაწილება მოვლენის მოხდენამდე ლოდინის (waiting) დროის აღწერისთვის.

განმარტება 4.6. მაჩვენებლიანი (α) განაწილება არის უწყვეტი ალბათობის განაწილება, რომლის ალბათობის სიმკვრივეა

$$f(x) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha}, & \text{როცა } x > 0 \\ 0, & \text{როცა } x \leq 0 \end{cases} \quad (4.12)$$

$\alpha > 0$ -თვის.

შემდეგ ნახაზზე გამოსახულია $f(x)$ ფუნქციის გრაფიკები α -ს სამი სხვადასხვა მნიშვნელობისათვის



ნახაზი (4.11) სამი მაჩვენებლიანი ალბათობის სიმკვრივე.

გამოსახულება (5.12) განსაკუთრებულად მოსახერხებელია, თანაც არაა ძნელი ვაჩვენოთ, რომ α არის ორივე: საშუალოც და სტანდარტული გადახრაც მაჩვენებლიანი (α) განაწილებისთვის. ანუ

$$\mu = EX = \int_0^{+\infty} x \frac{1}{\alpha} e^{-x/\alpha} dx = \alpha$$

მაჩვენებლიანი (α) განაწილების საშუალო

და

$$\sigma^2 = \text{Var}X = \int_0^{+\infty} (x - \alpha)^2 \frac{1}{\alpha} e^{-x/\alpha} dx = \alpha^2$$

მაჩვენებლიანი (α) განაწილების დისპერსია

გარდა ამისა, მაჩვენებლიან (α) განაწილებას გააჩნია ძალიან მარტივი დაგროვებითი ალბათობის ფუნქცია

$$F(x) = \begin{cases} 0, & \text{თუ } x \leq 0 \\ 1 - e^{-\frac{x}{\alpha}}, & \text{თუ } x > 0 \end{cases} \quad (4.13)$$

მაჩვენებლიანი (α) განაწილების დაგროვებითი ალბათობის ფუნქცია

მაგალითი 11. (დაბრუნება მე-7 მაგალითზე). **მაჩვენებლიანი განაწილება და უნივერსიტეტის ბიბლიოთეკაში შემსვლელთა რაოდენობა**

გავიხსენოთ, რომ სტროკი, ვოლსდორფი და მაკ-არტური აგროვებდნენ მონაცემებს უნივერსიტეტის ბიბლიოთეკაში შემომსვლელ სტუდენტთა რაოდენობის შესახებ 12:00-დან 12:10 p.m.-მდე. აღმოჩნდა, რომ წუთში საშუალოდ 12.5 სტუდენტი შევიდა ბიბლიოთეკაში. თუ გადავიყვანთ საშუალო დროში, რომელიც საჭიროა სტუდენტის გამოჩენამდე, მივიღებთ $1/12.5 = .08$ წუთს.

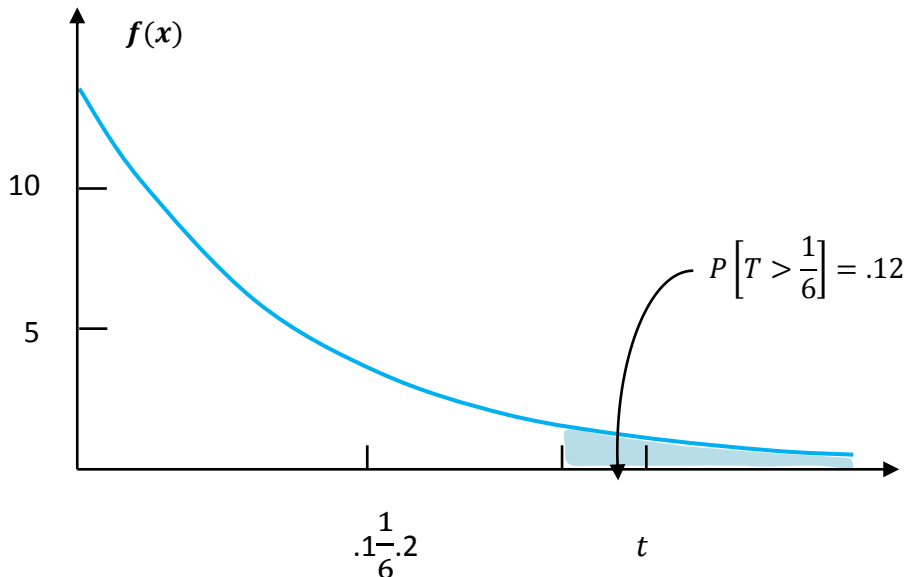
განვხილოთ ბიბლიოთეკაში შემსვლელებზე დაკვირვების პროცესი. დავუშვათ, რომ შემდეგ სამშაბათს შესვლა დაიწყო ზუსტად შუადღეს (12:00). განვსაზღვროთ შემთხვევითი სიდიდე

T = დალოდების დრო (წუთებში), სანამ პირველი სტუდენტი გაივლის კარებში

T -ს შესაძლო მოდელი არის მაჩვენებლიანი განაწილება, რომლის $\alpha = .08$. გამოვთვალოთ ალბათობა იმისა, რომ დალოდების დრო პირველ შემომსვლელამდე მეტია ვიდრე 10 წამი ($1/6$ წთ),

$$P \left[T > \frac{1}{6} \right] = 1 - F \left(\frac{1}{6} \right) = 1 - \left(1 - e^{-\frac{1}{6(.08)}} \right) = .12$$

ეს შედეგი გამოსახულია შემდეგ ნახაზზე:



ნახაზი 4.12 მაჩვენებლიანი ალბათობა მაგალითი 11-სთვის

გეომეტრიული და მაჩვენებლიანი განაწილებები

მაჩვენებლიანი განაწილება არის გეომეტრიული განაწილების უწყვეტი ანალოგი რამდენიმე თვალსაზრისით. პირველი: გეომეტრიული განაწილების ალბათობის ფუნქციაც და მაჩვენებლიანი განაწილების ალბათობის სიმკვრივეც კლებულობს ექსპონენციალურად არგუმენტის მიმართ. მეორე: ორივე პროცესი არის ე.წ. „მეხსიერების გარეშე“ ტიპის.

4.5. ვეიბულის განაწილება (ნებაყოფლობით)

ვეიბულის განაწილებები არამაჩვენებლიანი განაწილებების განზოგადებებია. ისინი უფრო მოქნილი არიან გრაფიკების ფორმების თვალსაზრისით. ეს განაწილებები მეტად პოპულარული არიან სანუირო საეშეში, როდესაც განიხილება, მაგალითად,

მასალის გამძლეობის თვისებები ან დამზადებული მოწყობილობის სიცოცხლის ხანგრძლივობა. ამ განაწილებათა განსაზღვრა, მათივე დაროვნებითი ალბათობის ფუნქციის საშუალებით, არის მათი განმარტების ყველაზე უფრო ბუნებრივი გზა.

განმარტება 4.6. ვეიბულის (α, β) განაწილება არის უწყვეტი ალბათობის განაწილება, რომლის დაგროვებითი ალბათობის ფუნქციაა

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}, & x \geq 0 \end{cases} \quad (4.14)$$

სადაც პარამეტრები $\alpha > 0$ და $\beta > 0$.

ამ ფორმულიდან გამომდინარე, შეიძლება განიმარტოს ვეიბულის განაწილებების თვისებები. (4.14) ფორმულის გაწარმოება გვაძლევს ვეიბულის (α, β) განაწილების ალბათობის სიმკვრივეს

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}, & x > 0 \end{cases} \quad (4.15)$$

აღმოჩნდა, რომ შესაძლებელია ვაჩვენოთ, რას უდრის საშუალო და დისპერსია. ანუ ვეიბულის (α, β) განაწილების საშუალოა

$$\mu = EX = \alpha \Gamma\left(1 + \frac{1}{\beta}\right) \quad (4.16)$$

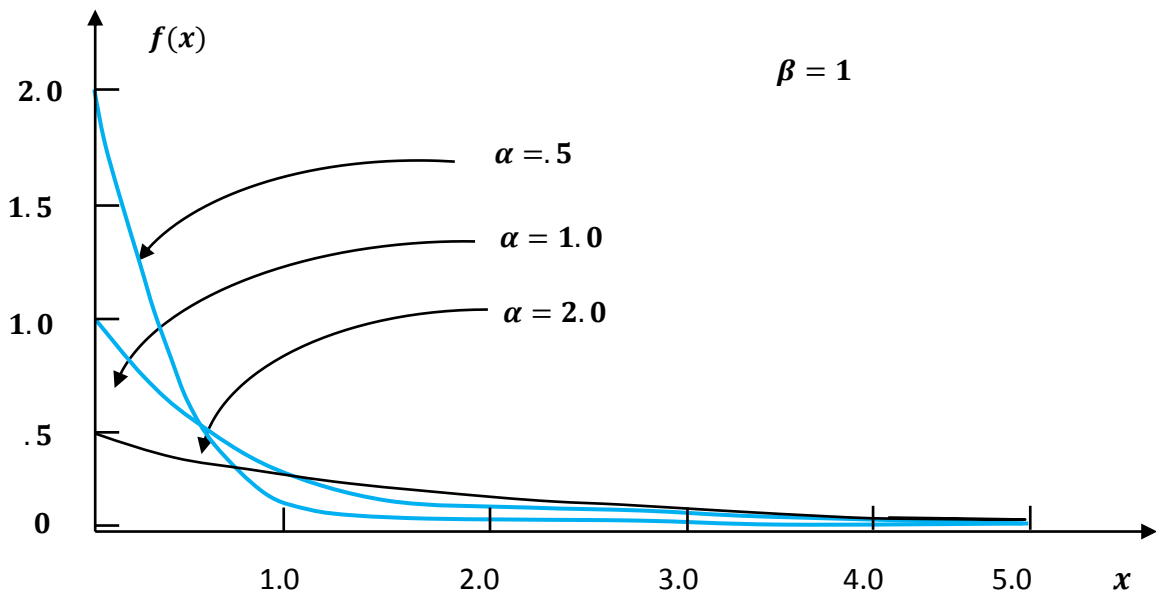
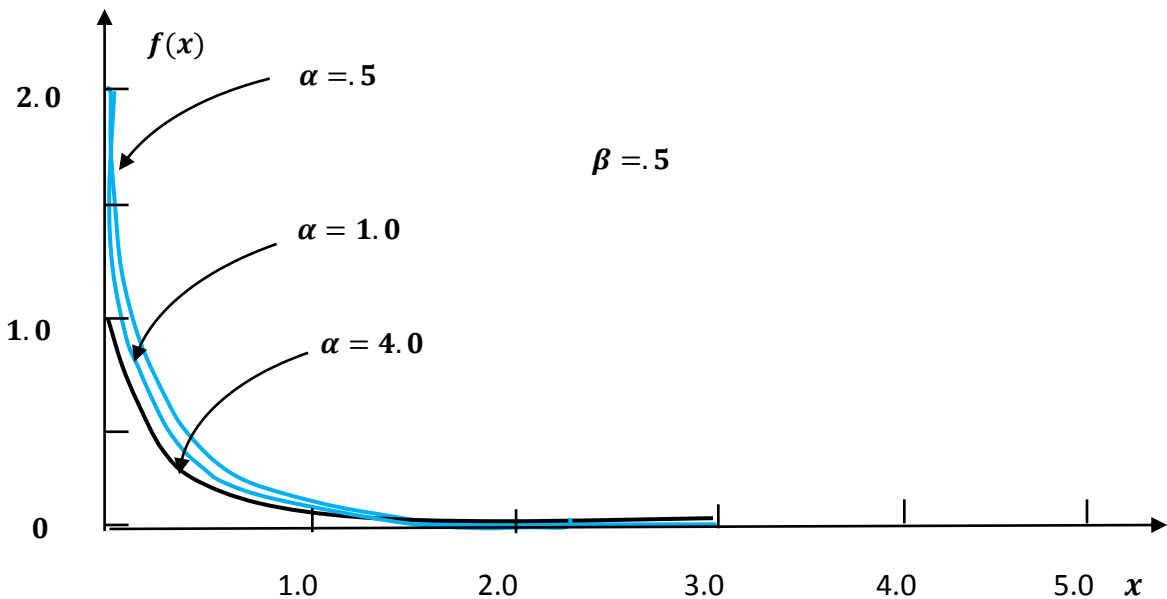
ვეიბულის (α, β) განაწილების დისპერსია

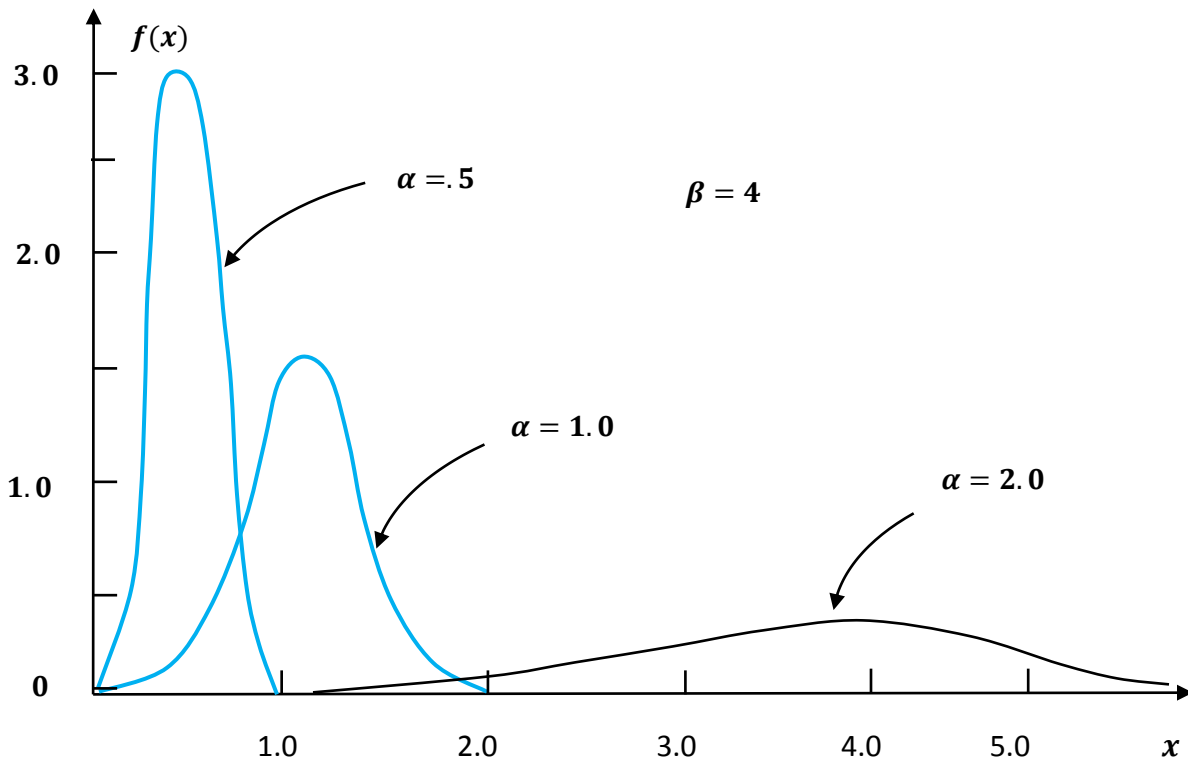
$$\sigma^2 = \text{Var } X = \alpha^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right) \right)^2 \right] \quad (4.17)$$

სადაც

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

არის *გამა-ფუნქცია* თანამედროვე (გაფართოებული) კალკულუსიდან. n -ის მთელი მნიშვნელობებისთვის $\Gamma(n) = (n-1)!$. ეს ფორმულები $f(x)$ -სთვის, μ -სა და σ^2 -სთვის არ არის მაინც და მაინც თვალსაჩინო. ამიტომ შესაძლოა უფრო გამოსაყენებელი იყოს მისი მარტივი რეალიზაციები, სადაც β აკონტროლებს ვეიბულის განაწილების ფორმას, ხოლო α აკონტროლებს სკალას (მასშტაბს). ქვემოთ ნაჩვენებია $f(x)$ -ის გრაფიკები რამდენიმე (α, β) წყვილისთვის.





ნახაზი 5.16. ვეიბულის ალბათობის ცხრა სიმკვრივის გრაფიკი.

შევნიშნოთ, რომ $\beta = 1$ იძლევა მაჩვენებლიანი განაწილებების სპეციალურ შემთხვევას. როცა β მცირეა, განაწილებები მარჯვნივ - დახრილია, მაგრამ როცა β მეტია დაახლოებით 3.6-ზე, რეალურად ხდებიან მარცხნივ - გადახრილნი. რაც შეეხება განაწილების ადგილმდებარეობას, განაწილების საშუალოს ფორმა, რომელიც განისაზღვრება (4.16) ფორმულით, არ არის საკმარისად გამოკვეთილი. შესაძლოა უფრო მეტად სასარგებლო იყოს ვიცოდეთ ვეიბულის (α, β) განაწილების მედიანა

$$Q(.5) = \alpha e^{-\frac{.3665}{\beta}}$$

(4.18)

მაგალითად, ფორმის პარამეტრის β -ს დიდი მნიშვნელობებისთვის ვეიბულის მედიანა ძირითად α -ს ტოლია. ფორმულები (4.16) და (4.18) გვიჩვენებენ, რომ ფიქსირებული β -სთვის ვეიბულის საშუალო, მედიანა და სტანდარტული გადახრა ყველა პროპორციულია α მასშტაბური პარამეტრის.

მაგალითი 12. ვეიბულის განაწილება და კერამიკული მასალის გამძლეობა

ე. ლენოუს მოხსენებაში „სემინარის მიმოხილვა თემაზე: კერამიკის დიზაინი, ანალიზი და საიმედოობის პროგნოზირება - ნაწ. II” (*Office of Naval Research Far East Scientific Bulletin 1987*) წარმოდგენილია კვლევა .95 მმ ღეროს შემოწმების შესახებ გამძლეობის ზღვარზე. იგი შეიძლება აღიწეროს ვეიბულის განაწილებით, რომლის $\beta = 8.8$ და მედიანაა 428. ვთქვათ,

S = შემდეგი ღეროს დაჭიმულობის ძალა

იმ დაშვებით, რომ S -ის მოდელირება შესაძლებელია ვეიბულის განაწილებით, წარმოდგენილ მახასიათებლებში, დავუშვათ, რომ გვჭირდება $P[S \leq 400]$ -ის გამოთვლა. (4.18) ტოლობის გამოყენებით,

$$428 = \alpha e^{-\frac{.3665}{8.8}}$$

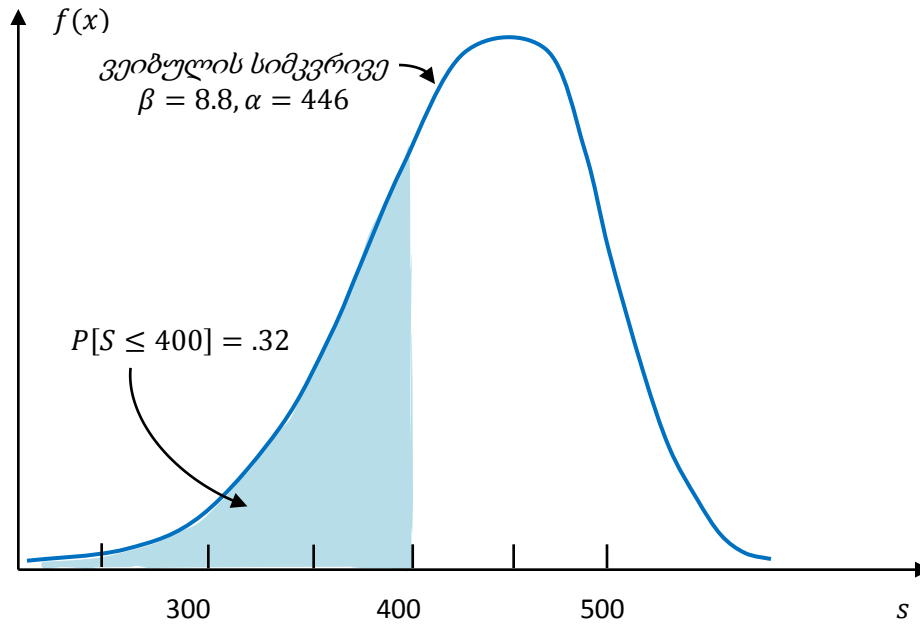
ანუ, ვეიბულის მასშტაბური პარამეტრი არის,

$$\alpha = 446$$

შემდგომ (4.14) ფორმულის გამოყენებით

$$P[S \leq 400] = 1 - e^{\left(-\frac{400}{446}\right)^{8.8}} = .32$$

შემდეგი გრაფიკი არის ალბათობის გამოთვლის ილუსტრაცია.



ნახაზი 5.17. ვეიბულის სიმკვრივე და $P[S \leq 400]$.

სავარჯიშოები:

1. ვიგულისხმობთ, რომ Z არის სტანდარტული ნორმალური შემთხვევითი ცვლადი. გამოთვალეთ შემდეგი ალბათობები, რომლებიც შეიცავენ Z -ს:

- ა) $P[Z < -.62]$ ბ) $P[Z > 1.06]$ გ) $P[-.37 < Z < .51]$
 დ) $P[|Z| \leq .47]$ ე) $P[|Z| > .93]$ ვ) $P[-3.0 < Z < 3.0]$

ახლა იპოვეთ ისეთი $\#$ რიცხვები, რომ შემდეგი დებულებები, რომლებიც შეიცავენ Z -ს, იყოს ჭეშმარიტი:

- ზ) $P[Z \leq \#] = .90$ თ) $P[|Z| < \#] = .90$ ი) $P[|Z| > \#] = .90$

2. ვიგულისხმობთ, რომ X არის ნორმალური შემთხვევითი ცვლადი, რომლის საშუალოა 43.0 და სტანდარტული გადახრა 3.6. გამოითვალეთ შემდეგი ალბათობები, რომლებიც შეიცავენ X -ს:

$$\text{ა) } P[X < 45.2]$$

$$\text{ბ) } P[X \leq 41.7]$$

$$\text{გ) } P[43.8 < X \leq 47.0]$$

$$\text{დ) } P[|X - 43.0| \leq 2.0]$$

$$\text{ე) } P[|X - 43.0| > 1.7]$$

ახლა იპოვეთ ისეთი $\#$ რიცხვები, რომ შემდეგი დებულებები, რომლებიც შეიცავენ X -ს, იყოს ჭეშმარიტი:

$$\text{ვ) } P[X < \#] = .95$$

$$\text{ზ) } P[X \geq \#] = .30$$

$$\text{თ) } P[|X - 43.0| > \#] = .05$$

5. ერთობლივი განაწილებები და დამოუკიდებლობა

საინჟინრო სტატისტიკაში ძალიან ხშირად ვხვდებით შემთხვევებს, როდესაც გვჭირდება ალბათობის გამოყენება არა მხოლოდ ერთი, არამედ ორი ან მეტი შემთხვევითი ცვლადის პირობებში. ზოგჯერ კი არის მრავალი ცვლადის შემთხვევებიც. ასეთ დროს უფრო მეტად, ვიდრე ერთი ცვლადის შემთხვევაში, უნდა გავითვალისწინოთ შემთხვევითი მომენტების გავლენები ალბათობის გამოთვლაზე, რომელიც თავის მხრივ ასოცირებული უნდა იყოს შემთხვევითი ცვლადების (სიდიდეების) კომბინაციებთან. მაგალითად, თუ განვიხილავთ, საკისრების აწყობის პროცესს, შიდა რგოლების ნორმალური დიამეტრი ვთქვათ, უნდა იყოს 1.00 in., ხოლო ღეროს ნორმალური დიამეტრი, რომელიც მასში უნდა გავუყაროთ არის .99 in. თუ

X = საკისრის რგოლის შიგა დიამეტრი,

Y = ღეროს დიამეტრი,

მაშინ ჩვენთვის საინტერესო ფაქტი, რომ

$$P[X < Y] = P[\text{ხელშემშლელი ფაქტორია აწყობისთვის}],$$

მოიცავს ორივე ცვლადს.

საინჟინრო გამოყენებებში თითქმის ყოველთვის გამოიყენება შერჩევები, რომელთა n ზომები ზოგადად მეტია 1-ზე.

5.1. დისკრეტულ შემთხვევით სიდიდეთა ერთობლობა

რამდენიმე დისკრეტული შემთხვევითი სიდიდის ალბათობის განსაზღვრისთვის გამოიყენება *ერთობლივი ალბათობის ფუნქცია*. მოვიყვანოთ მისი განმარტება ორი ცვლადის შემთხვევაში.

განმარტება 5.1. X და Y დისკრეტულ შემთხვევით სიდიდეთა ერთობლივი ალბათობის ფუნქცია არის არაუარყოფითი $f(x, y)$ ფუნქცია, რომელიც ალბათობას

განსაზღვრავს (ერთდროულად), როდესაც X ღებულობს x მნიშვნელობებს და Y ღებულობს y მნიშვნელობებს, ანუ

$$f(x, y) = P [X = x \text{ და } Y = y]$$

მაგალითი 5.1. ერთობლივი ალბათობის განაწილება: ორი ჭანჭიკის მოჭერის ბრუნვის მომენტი. (დაბრუნება მაგალით 3.1-ზე). დავუბრუნდეთ ბრინის, ქრისტენსენისა და შნაიდერის ექსპერიმენტს. მონაცემები შედგენილი გვაქვს ბრუნვის მომენტების უახლოესი მთელი რიცხვებით.

X = ბრუნვის მომენტი ჩაწერილი #3 ჭანჭიკისთვის

Y = ბრუნვის მომენტი ჩაწერილი #4 ჭანჭიკისთვის

მონაცემების საშუალებით, რომლებიც წარმოდგენილია ცხრილში 3.1 და გრაფიკით 3.2, აზრიანია დავასკვნათ, რომ მაგალითად, $P[X = 18 \text{ და } Y = 18]$ შეიძლება იყოს $\frac{1}{34}$, რომელიც არის წყვილის ფარდობითი სიხშირე მონაცემების მიხედვით. მსგავსად შეიძლება განისაზღვროს

$$P[X = 18 \text{ და } Y = 17] = \frac{2}{34}$$

და

$$P[X = 14 \text{ და } Y = 9] = 0,$$

რომლებიც შეესაბამება დაკვირვებების ფარდობით სიხშირეებს.

ამ პირობებში შესაძლებელია მივიღოთ სტუდენტების მიერ მოგროვილი მონაცემების ფარდობით სიხშირეთა სრული სურათი ორი X და Y შემთხვევითი სიდიდეებისთვის. ამისთვის მოსახერხებელია, რომ მონაცემები დავიტანოთ ორგანოზომილებიან ცხრილზე და განვსაზღვროთ ერთობლივი ალბათობის ფუნქცია X და Y ცვლადებისთვის. ყოველივე ეს გამოსახულია შემდეგ ცხრილში. (ცხრილის გადატვირთვისგან თავის არიდების მიზნით 0-ების მაგიერ ცარიელი ადგილებია დატოვებული).

$y \backslash x$	11	12	13	14	15	16	17	18	19	20
20								2/34	2/34	1/34
19							2/34			
18			1/34	1/34			1/34	1/34	1/34	
17					2/34	1/34	1/34	2/34		
16				1/34	2/34	2/34			2/34	
15	1/34	1/34			3/34					
14					1/34			2/34		
13					1/34					

ცხრილი 5.1. $f(x,y)$ ჭანჭიკების მოჭერის ამოცანისთვის.

ერთობლივი ალბათობის ფუნქციის თვისებები: ერთობლივი ალბათობის ფუნქციას გააჩნია ორი თვისება რომელიც სავალდებულოა, რომ შეთანხმებული იყოს მათემატიკურ წარმოდგენებთან. ესენია: $f(x,y)$ -ის ყველა მნიშვნელობა უნდა ვარდებოდეს $[0; 1]$ შუალედში და მათი ყველას ჯამი უნდა იყოს 1. (x,y) -ის ზოგიერთი მნიშვნელობების შეკრებისას ალბათობები ასოცირდება X და Y ცვლადების მიერ მიღებულ საინტერესო კონფიგურაციებთან ანუ გრაფიკზე გამოსახულ ფორმებთან.

განვიხილოთ ერთობლივი განაწილების ზედა ცხრილი და გამოვთვალოთ

$$P[X \geq Y], \quad P[|X - Y| \leq 1] \quad \text{და} \quad P[X = 17].$$

პირველი $P[X \geq Y]$ ნიშნავს, რომ X ანუ #3 ჭანჭიკის ბრუნვის მომენტები არ არის ნაკლები #4 ჭანჭიკის ბრუნვის მომენტებზე - Y -ზე. ქვემოთ მოცემულ გრაფიკზე ვარსკვლავებით მონიშნულია ის უჯრები, რომლებიც შეესაბამებიან ამ სიტუაციას.

$x \backslash$	11	12	13	14	15	16	17	18	19	20
20										*
19									*	*
18								*	*	*
17							*	*	*	*
16						*	*	*	*	*
15					*	*	*	*	*	*
14				*	*	*	*	*	*	*
13			*	*	*	*	*	*	*	*

გრაფიკი 5.2. #3 და #4 ჭანჭიკების კომბინაციები, როდესაც $X \geq Y$

თუ ერთიანად შევკრებთ ვარსკვლავებით მონიშნული უჯრების შესაბამის მნიშვნელობებს ცხრილი 5.1 -დან, მივიღებთ:

$$\begin{aligned}
 P[X \geq Y] &= f(15, 13) + f(15, 14) + f(15, 15) + f(16, 16) + f(17, 17) + \\
 &+ f(18, 14) + f(18, 17) + f(18, 18) + f(19, 16) + f(19, 18) + f(20, 20) = \\
 &= \frac{1}{34} + \frac{1}{34} + \frac{3}{34} + \frac{2}{34} + \dots + \frac{1}{34} = \frac{17}{34} = \frac{1}{2}
 \end{aligned}$$

ანალოგიურად, შეიძლება გამოვთვალოთ $P[|X - Y| \leq 1]$ — ალბათობა იმისა, რომ #3 და #4 ჭანჭიკების ბრუნვის მომენტები 1 (ft lb) -ზე მეტად არ არიან განსხვავებულები. ეს დამოკიდებულება შემდეგ გრაფიკზეა წარმოდგენილი.

$x \backslash$	11	12	13	14	15	16	17	18	19	20
20									*	*
19								*	*	*
18							*	*	*	
17						*	*	*		
16					*	*	*			
15				*	*	*				
14			*	*	*					
13		*	*	*						

გრაფიკი 5.3. #3 და #4 ჭანჭიკების კომბინაციები, როდესაც $|X - Y| \leq 1$

კვლავ თუ შევკრებთ კომბინაციების შესაბამის ალბათობის მნიშვნელობებს, მივიღებთ:

$$\begin{aligned}
 P[|X - Y| \leq 1] &= f(15, 14) + f(15, 15) + f(15, 16) + f(16, 16) + f(16, 17) + \\
 &+ f(17, 17) + f(17, 18) + f(18, 17) + f(18, 18) + f(19, 18) + \\
 &+ f(19, 20) + f(20, 20) = \frac{18}{34}
 \end{aligned}$$

ბოლოს $P[X = 17]$ არის ალბათობა იმისა, რომ #3 ჭანჭიკის ბრუნვის მომენტი არის 17 ft-lb. იგი მიიღება შესაბამისი მნიშვნელობების შეკრებით როცა $x = 17$, ცხრილი 5.1-დან და გვაქვს:

$$\begin{aligned}
 P[X = 17] &= f(17, 17) + f(17, 18) + f(17, 19) = \\
 &= \frac{1}{34} + \frac{1}{34} + \frac{2}{34} = \frac{4}{34}
 \end{aligned}$$

ზღვრული ალბათობის ფუნქციის მოძებნა ბივარიანტული ერთობლივი ალბათობის ფუნქციის საშუალებით

ბივარიანტულ ამოცანებში, მაგალითად, ისეთებში, როგორიც ახლა წარმოვადგინეთ, შესაძლებელია შევკრიბოთ ცხრილში წარმოდგენილი $f(x, y)$ ალბათობები სვეტების მიხედვით და მივიღოთ X შემთხვევითი სიდიდის $f_X(x)$ ალბათობის ფუნქცია. ანალოგიურად, შეგვიძლია იგივე ცხრილში შევკრიბოთ ალბათობები სტრიქონების გასწვრივ და მივიღოთ Y შემთხვევითი სიდიდის $f_Y(y)$ ალბათობის ფუნქცია. შეგვიძლია სვეტებისა და სტრიქონების მიხედვით შეჯამების შედეგები მივუწეროთ ცხრილს კიდევში. გასაკვირი არაა, რომ ალბათობის განაწილებები ინდივიდუალური შემთხვევითი ცვლადებისთვის გამოითვლება მათი ერთობლივი განაწილებიდან და ეწოდებათ *ზღვრული (მარგინალური) განაწილებები*. ფორმალურად ეს ტერმინი ორი დიკრეტული ცვლადის შემთხვევისთვის ასე ჩამოყალიბდება:

განმარტება 5.2. X და Y დისკრეტული შემთხვევითი სიდიდეების ერთობლივი ალბათობის $f(x, y)$ ფუნქციის ინდივიდუალურ ალბათობის ფუნქციებს ეწოდებათ *ზღვრული ალბათობის ფუნქციები*. ისინი მიიღებიან $f(x, y)$ ფუნქციის ყველა შესაძლო მნიშვნელობათა შეკრებით სხვა ცვლადის მიმართ. (თუ $f_X(x)$ გვინდა, ვკრიბავთ y -ების და თუ $f_Y(y)$ გვინდა — x -ების მიმართ. ზღვრული ალბათობის ფუნქცია x -სთვის ფორმულით ასე გამოისახება:

$$f_X(x) = \sum_y f(x, y)$$

ხოლო ზღვრული ალბათობის ფუნქცია y -სთვის ასეთია:

$$f_Y(y) = \sum_x f(x, y)$$

შემდეგი ცხრილი თითქმის იგივეა რაც ცხრილი 5.1, მაგრამ გადიდებული X და Y -ის ზღვრული ალბათობებით.

$y \setminus x$	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20								2/34	2/34	1/34	5/34
19							2/34				2/34
18			1/34	1/34			1/34	1/34	1/34		5/34
17					2/34	1/34	1/34	2/34			6/34
16				1/34	2/34	2/34			2/34		7/34
15	1/34	1/34			3/34						5/34
14					1/34			2/34			3/34
13					1/34						1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	

ცხრილი 5.2. ერთობლივი და ზღვრული ალბათობები X და Y -ისთვის

აქედან შესაძლებელია გაკეთდეს სეპარაცია ანუ ცალკე გამოიყოს ზღვრული ალბათობის ფუნქცია მხოლოდ Y -ისთვის. ეს გაკეთებულია შემდეგ ცხრილში.

y	$f_Y(y)$
13	1/34
14	3/34
15	5/34
16	7/34
17	6/34
18	5/34
19	2/34
20	5/34

ცხრილი 5.3. ზღვრული ალბათობის ფუნქცია Y -ისთვის

ერთობლივი ალბათობის ფუნქციდან ზღვრული (მარგინალური) ალბათობის მიღების შესაძლებლობა ბადებს ძალიან ბუნებრივ კითხვას: შესაძლებელია თუ არა პროცესის შებრუნება, ანუ, თუ $f_X(x)$ და $f_Y(y)$ ცნობილია, გვაქვს თუ არა ერთი არჩევანი $f(x,y)$ -სთვის? ანუ შესაძლებელია თუ არა, რომ $f(x,y)$ ცალსახად განისაზღვროს? ამ კითხვაზე პასუხი არის „არა“.

შემდეგ ცხრილები გვიჩვენებენ, რომ არსებითად განსხვავებულ ბივარიანტულ ერთობლივი ალბათობის ფუნქციებს შესაძლებელია ჰქონდეთ ერთი იმავე ზღვრული ალბათობის ფუნქციები.

$y \backslash x$	1	2	3	
3	.4	0	0	.4
2	0	.4	0	.4
1	0	0	.2	.2
	.4	.4	.2	

$x \backslash y$	1	2	3	
3	.16	.16	.08	.4
2	.16	.16	.08	.4
1	.08	.08	.04	.2
	.4	.4	.2	

ცხრილი 5.3. *ორი განსხვავებული ერთობლივი განაწილება ერთი და იმავე ზღვრული განაწილებებით*

5.2. პირობითი ალბათობა და დამოუკიდებლობა დისკრეტული შემთხვევითი სიდიდებისთვის

როდესაც გვაქვს რამდენიმე შემთხვევითი სიდიდე, კარგია თუ ხშირად დავფიქრდებით იმაზე, თუ რას უნდა ველოდეთ ერთი რომელიმე ცვლადისგან, როდესაც ვიცით, დავუშვათ, სხვა ყველა დანარჩენების მნიშვნელობები. ალბათობის თეორია საშუალებას იძლევა „ვიპოვოთ ერთი ცვლადის განაწილება, თუ ვიცით სხვა ცვლადების მნიშვნელობები“. ეს კეთდება პირობითი ალბათობის დახმარებით. ორი ცვლადის შემთხვევისთვის ეს ასე განიმარტება:

განმარტება 5.3. X და Y დისკრეტული შემთხვევითი სიდიდეებისთვის რომელთა ერთობლივი ალბათობის ფუნქციაა $f(x, y)$, პირობითი ალბათობის ფუნქცია X -ის მიმართ, როცა $Y = y$, არის x -ის ფუნქცია და

$$f_{X|Y}(x | y) = \frac{f(x, y)}{\sum_x f(x, y)}$$

პირობითი ალბათობის ფუნქცია Y -ის მიმართ, როცა $X = x$, არის y -ის ფუნქცია და

$$f_{Y|X}(y | x) = \frac{f(x, y)}{\sum_y f(x, y)}$$

ამ და წინა ფორმულების შედარება გვაძლევს:

X -ის პირობითი ალბათობის ფუნქცია, როცა $Y = y$, არის

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}$$

და

Y -ის პირობითი ალბათობის ფუნქცია, როცა $X = x$, არის

$$f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)}$$

ეს ტოლობები მიგვანიშნებენ, რომ თუ გავყვებით მხოლოდ სტრიქონს (პირველ ფორმულაში), რომელიც განსაზღვრულია ტოლობით $Y = y$, მაშინ შესაბამისი (პირობითი) განაწილება X -ის მიმართ ტოლია სტრიქონში მოცემული ალბათობები ($f(x, y)$ -ის მნიშვნელობები) გაყოფილი მათ ჯამზე (ანუ $\sum_x f(x, y)$ -ზე). ამ დროს

მომხდარია მათი რენორმალიზება, ანუ ალბათობების ჯამი ტოლია ერთის. მსგავსად იქნება მეორე ტოლობისთვისაც. ამ შემთხვევაში უნდა გავყვეთ სვეტს, რომელიც განსაზღვრულია ტოლობით $X = x$. შესაბამისი პირობითი განაწილება Y -სთვის მოიცემა სვეტში მოცემული ალბათობების გაყოფით მათ ჯამზე.

მაგალითი 5.1. (გაგრძელება). ამ ფორმულების საილუსტრაციოდ განვიხილოთ რამდენიმე პირობითი ალბათობა, ასოცირებული #3 და #4 ჭანჭიკების ერთობლივ განაწილებასთან. დავიწყეთ მაგალითად Y -ის პირობითი განაწილებით, როდესაც მოცემულია, რომ $X = 15$,

$$f_{Y|X}(y | 15) = \frac{f(15, y)}{f_X(15)}$$

დავუბრუნდეთ ცხრილს 5.2. ზღვრული ალბათობა, რომლიც ასოცირებულია $X=15$ მნიშვნელობასთან არის $\frac{9}{34}$. მაშინ, თუ $X = 15$ სვეტის მნიშვნელობებს გავყოფთ $\frac{9}{34}$ -ზე, მივიღებთ პირობით განაწილებას Y -ისთვის, რაც მოცემულია შემდეგ ცხრილში და იძლევა ისეთ მნიშვნელობებს, რომლებიც თანხვედრაშია ინტუიციასთან.

y	$f_{Y X}(y 15)$
13	$\left(\frac{1}{34}\right) \div \left(\frac{9}{34}\right) = \frac{1}{9}$
14	$\left(\frac{1}{34}\right) \div \left(\frac{9}{34}\right) = \frac{1}{9}$
15	$\left(\frac{3}{34}\right) \div \left(\frac{9}{34}\right) = \frac{3}{9}$
16	$\left(\frac{2}{34}\right) \div \left(\frac{9}{34}\right) = \frac{2}{9}$
17	$\left(\frac{2}{34}\right) \div \left(\frac{9}{34}\right) = \frac{2}{9}$

ცხრილი 5.4. Y -ის პირობითი ალბათობის ფუნქცია მოცემული $X = 15$ პირობით

ახლა განვიხილოთ $f_{Y|X}(y | 18)$, რომელიც განსაზღვრულია ტოლობით:

$$f_{Y|X}(y | 18) = \frac{f(18, y)}{f_X(18)}$$

კვლავ განვიხილავთ ცხრილს 5.2 და პირობით განაწილებას Y -ისთვის, როცა $X=15$ ექნება ასეთი სახე:

y	$f_{Y X}(y 18)$
14	2/7
17	2/7
18	1/7
20	2/7

ცხრილი 5.5. Y -ის პირობითი ალბათობის ფუნქცია მოცემული $X = 18$ პირობით

ამ ცხრილების (5.4 -ის და 5.5 -ის) შედარებით ვნახავთ, რომ Y -ის პირობითი განაწილებები $X=15$ და $X=18$ -სთვის სრულიად განსხვავებულები არიან. მაგალითად, $X=18$ -ის შემთხვევაში მთლიანობაში უნდა ველოდოთ, რომ მისი მნიშვნელობები იქნებიან მეტნი ვიდრე $X=15$ -ის შემთხვევაში.

ცხადია, შეგვიძლია გავაკეთოთ იგივე და განვიხილოთ პირობითი განაწილება #3 ჭანჭიკისთვის (X) თუ მოცემულია #4 ჭანჭიკის ბრუნვის მომენტი (მაგ. $Y = 20$). ამ სიტუაციაში გვაქვს, რომ

$$f_{X|Y}(x | 20) = \frac{f(x, 20)}{f_Y(20)}$$

X -ის პირობითი ალბათობა, როცა $Y = 20$, ნიშნავს რომ სტრიქონის ელემენტები უნდა გაიყოს $Y = 20$ ზღვრულ მნიშვნელობაზე. ეს $f_{X|Y}(x | 20)$ მნიშვნელობები მოცემულია შემდეგ ცხრილში:

x	$f_{X Y}(x 20)$
18	$\left(\frac{2}{34}\right) \div \left(\frac{5}{34}\right) = \frac{2}{5}$
19	$\left(\frac{2}{34}\right) \div \left(\frac{5}{34}\right) = \frac{2}{5}$
20	$\left(\frac{3}{34}\right) \div \left(\frac{5}{34}\right) = \frac{1}{5}$

ცხრილი 5.6. X -ის პირობითი ალბათობის ფუნქცია მოცემული $Y = 20$ პირობით

შემოვიტანოთ ორი ცვლადისთვის დამოუკიდებლობის ცნება.

განმარტება 5.4. X და Y დისკრეტულ შემთხვევით სიდიდეებს ეწოდებათ **დამოუკიდებლები**, თუ ერთობლივი ალბათობის ფუნქცია $f(x, y)$ წარმოადგენს შესაბამისი ზღვრული ალბათობების ფუნქციების ნამრავლს, ანუ ფორმულით დამოუკიდებლობა ნიშნავს შემდეგ ტოლობას:

$$f(x, y) = f_X(x) \cdot f_Y(y),$$

ყოველი x და y - სთვის.

თუ ეს ფორმულა არაა ძალაში, მაშინ ვამბობთ, რომ X და Y ცვლადები **დამოუკიდებლები** არიან.

განმარტებიდან გამომდინარეობს, რომ დამოუკიდებლობის შემთხვევაში პირობითი განაწილება ტოლია შესაბამისი ზღვრული განაწილების (მარგინალის). ამიტომ ეს განმარტება შეესაბამება „დამოუკიდებლობის“ ცნების შემოტანის მოტივაციას.

მაგალითში განხილული X და Y ცვლადები დამოუკიდებლები არიან. მაგრამ თუ დავუბრუნდებით ცხრილს 5.3 და დავაკვირდებით ერთობლივი განაწილების ორ ცხრილს, შევამჩნევთ, რომ პირველ მათგანში გვაქვს თვალნათლივი დამოკიდებულება, ხოლო მეორე შემთხვევაში ცვლადები არიან დამოუკიდებლები, თუმცა გააჩნიათ ერთი და იგივე მარგინალები.

განმარტება 5.5. $X_1, X_2, \dots, X_n$ შემთხვევით სიდიდეებს, რომელთაც აქვთ ერთი და იგივე ზღვრული განაწილებები და არიან დამოუკიდებლები ეწოდებათ **iid**, ანუ **დამოუკიდებელი და ერთნაირად განაწილებული** შემთხვევით სიდიდეები.

როდის შეიძლება რომ დაკვირვება მოდელირდეს როგორც iid?

სტანდარტული სტატისტიკური მაგალითები **iid** შემთხვევით სიდიდეებზე არიან თანმიმდევრული საზომები, აღებული სტაბილური პროცესებიდან და ერთგვაროვანი პოპულაციის შემთხვევითი შერჩევებიდან, დაბრუნების გარეშე. კითხვა: როდის არის **iid** მოდელი შესაბამისი კონკრეტული სტატისტიკური

პრობლემისთვის? პასუხი დამოკიდებულია იმაზე, თუ მონაცემთა გენერირების მექანიზმი, რომლითაც მიმდინარეობდა შესწავლა, რამდენადაა მისი (*iid*-ის) კონცეპტუალურად ეკვივალენტური.

5.3. მრავალი ცვლადის ფუნქციები

ამოცანა, რომელიც ამ ნაწილში უნდა განვიხილოთ, მდგომარეობს შემდეგში. მოცემული გვაქვს $X, Y, \dots, Z$ ცვლადების ერთობლივი განაწილება და ფუნქცია $g(x, y, \dots, z)$, რომელიც აკეთებს

$$U = g(X, Y, \dots, Z)$$

პროგნოზს $X, Y, \dots, Z$ შემთხვევით სიდიდეთა შესახებ. თუმცა აქ ფრთხილად უნდა ვიყოთ იმ გაგებით, რომ არ უნდა ველოდოთ, რომ გამოსავალი ფუნქცია იქნება უფრო მეტად რეალური, ვიდრე დაშვებები.

ზოგიერთ სპეციალურ, მარტივ შემთხვევაში შესაძლებელია ზუსტად აღვწეროთ, თუ რა მემკვიდრეობა მიიღო U განაწილებამ $X, Y, \dots, Z$ ცვლადებისგან.

მაგალითი 5.2. შემთხვევითად განსაზღვრული განზომილების ორ შეუღლებულ დეტალს შორის ღრიჩოთა განაწილება.

წარმოვიდგინოთ, რომ ფოლადის ფირფიტა, რომლის ნორმალური სისქეა .15 (in.), ზედაპირის მექანიკური დამუშავებისთვის მოთავსებულია ფოლადის ბლოკში, სიგანით .155 (in.). დამზადებულია ბევრი ფირფიტა და გაზომილია მათი სისქე. შედეგები წარმოდგენილია ფარდობით სიხშირეთა ცხრილების სახით:

ფირფიტის სისქე(in.)	ფარდობითი სიხშირე
.148	.4
.149	.3
.150	.3

ცხრილი 5.7. ფირფიტის სისქის ფარდობით სიხშირეთა განაწილება

ბლოკის სიგანე(in.)	ფარდობითი სიხშირე
.153	.2
.154	.2
.155	.4
.156	.2

ცხრილი 5.8. ბლოკის სიგანის ფარდობით სიხშირეთა განაწილება

თუ ფირფიტა შერჩეულია შემთხვევითად და ბლოკიც განცალგებულად შემთხვევითად არის შერჩეული, მაშინ შემთხვევითი სიდიდეების

X = ფირფიტის სისქე

Y = ბლოკის სიგანე

დამოუკიდებლობა არიან და ერთობლივი განაწილებაში X -ის მარგინალური განაწილება მოცემულია ცხრილი 5.7-ით, ხოლო Y -ის ცხრილი 5.8 -ით. შემდეგი ცხრილი კი წარმოგვიდგენს ერთ-ერთ სავარაუდო ერთობლივი ალბათობის ფუნქციას X და Y ცვლადებისთვის.

$y \setminus x$	.148	.149	.150	$f_Y(y)$
.156	.08	.06	.06	.2
.155	.16	.12	.12	.4
.154	.08	.06	.06	.2
.153	.08	.06	.06	.2
$f_X(x)$	.4	.3	.3	

ცხრილი 5.9. ზღვრული და ერთობლივი ალბათობები X და Y ცვლადებისთვის

ცვლადი, რომელიც მიიღება X და Y ცვლადებისგან და რომლის მიმართაც გვაქვს მნიშვნელოვანი პოტენციური ინტერესი, არის ღრიჭე, რომელიც გამოისახება ასე

$$U = Y - X .$$

შევნიშნოთ, რომ ცხრილი 5.7 და ცხრილი 5.8 -დან თუ ავიღებთ ექსტრემალურ წარმომადგენლებს, გარანტირებულად შეგვიძლია ვთქვათ, რომ არ შეიძლება იყოს $.153 - .150 = .003$ in.-ზე ნაკლები და მეტი $.156 - .148 = .008$ in.-ზე მეტი. შეგვიძლია ვთქვათ კიდევ მეტი რამ. მაგ, თუ დავაკვირდებით ცხრილი 5.9 -ს, დავინახავთ, რომ დიგონალზე (ქვედა მარცხენა კუთხიდან ზედა მარჯვენამდე კუთხემდე) შეესაბამება იგივე $Y - X$ მნიშვნელობას. თუ შევკრებთ ალბათობებს აღნიშნული დიაგონალის პარალელურ დიაგინალებზე, მივიღებთ U -ს განაწილების შემდეგ ცხრილს:

u	$f(u)$
.003	.06
.004	.12 = .06 + .06
.005	.26 = .08 + .06 + .12
.006	.26 = .08 + .12 + .06
.007	.22 = .16 + .06
.008	.08

ცხრილი 5.10. ალბათობის ფუნქცია $U = Y - X$ დრიჭესთვის

ეს მაგალითი შეიცავს ძალიან მარტივ ერთობლივ დისკრეტულ განაწილებას და ძალიან მარტივ g ფუნქციას, მაგალითად, $g(x, y) = y - x$. უნდა ითქვას, რომ ზოგადად ამ პრობლემის ზუსტი და სრულყოფილი ამოხსნა არ არსებობს, ანუ პრაქტიკულად შეუძლებელია $U = g(X, Y, \dots, Z)$ განაწილების პოვნა. თუმცა საბედნიეროდ ისიც უნდა ითქვას, რომ ალბათობისა და სტატისტიკის მრავალ ინჟინრულ გამოყენებებში მიახლოებითი და/ან კერძო ამონახნები სავსებით საკმარისია იმისთვის, რომ ვუპასუხოთ კონკრეტული ინტერესის მქონე კითხვას. ამისათვის არსებობს სპეციალური მეთოდები, რომლებიც წარმოგვიდგენს U ფუნქციის მიახლოებით და/ან კერძო აღწერას.

5.4. საშუალო და დისპერსია შემთხვევით სიდიდეთა წრფივი კომბინაციებისთვის

საინჟინრო მიზნებისთვის, ხშირად $U = g(X, Y, \dots, Z)$ განაწილებისათვის სრულიად საკმარისია ვიცოდეთ მისი საშუალო და დისპერსია (იმის მაგივრად, რომ ვიცოდეთ U -ს მთლიანი განაწილება). როდესაც g არის წრფივი ფუნქცია, მაშინ არსებობენ ზუსტი ფორმულები.

თეორემა 5.1. თუ $X, Y, \dots, Z$ არიან n რაოდენობის დამოუკიდებელი შემთხვევითი სიდიდეები და $a_0, a_1, a_2, \dots, a_n$ მუდმივებია $n + 1$ ცალი, მაშინ $U = a_0 + a_1X + a_2Y + \dots + a_nZ$ შემთხვევითი სიდიდის საშუალოა (მათემატიკური ლოდინი):

$$EU = a_0 + a_1EX + a_2EY + \dots + a_nEZ$$

და დისპერსიაა:

$$Var U = a_1^2 \cdot Var X + a_2^2 \cdot Var Y + \dots + a_n^2 \cdot Var Z$$

შევნიშნოთ, რომ საშუალოს ფორმულა ძალაშია ყოველთვის, ანუ იმისდა მიუხედავად $X, Y, \dots, Z$ შემთხვევითი სიდიდეები დამოუკიდებლები არიან თუ დამოკიდებულნი, მაშინ როცა დისპერსიის ფორმულა მართებულია მხოლოდ დამოუკიდებლობის შემთხვევაში. მიუხედავად ამისა, ზოგიერთ კერძო შემთხვევაში არსებობს მისი განზოგადება. მაგალითად, როცა $a_0 = 0$ და სხვა a_i -ები ტოლია ერთის ან მინუს ერთის.

მაგალითი 5.2. (გაგრძელება). დავუბრუნდეთ ფოლადის ფირფიტებისა და ბლოკების სიტუაციას. თუ X, Y და U შესაბამისად არიან ფირფიტის სისქე, ბლოკის

სიგანე და ღრეჭო, მაშინ საშუალო და დისპერსია ადვილად გამოითვლებიან ზემოთ მოყვანილი ცხრილების საშუალებით. მკითხველს ვთხოვთ, გადაამოწმოს, რომ

$$EX = .1489 \text{ და } Var X = 6.9 \cdot 10^{-7}$$

$$EY = .1546 \text{ და } Var Y = 1.04 \cdot 10^{-6}$$

ვინაიდან, $U = Y - X = (-1) \cdot X - 1 \cdot Y$, თეორემა 5.1 -ის გამოყენებით შეიძლება გამოვთვალოთ, რომ

$$EU = -1 \cdot EX + 1 \cdot EY = -.1489 + .1546 = .0057 \text{ (in.)}$$

$$Var U = (-1)^2 \cdot 6.9 \cdot 10^{-7} + 1^2 \cdot 1.04 \cdot 10^{-6} = 1.73 \cdot 10^{-6}$$

ასე, რომ

$$\sqrt{Var U} = .0013 \text{ (in.)}$$

ღირს ძალისხმევად შევამოწმოთ შემდეგი: ღრეჭოს საშუალო და სტანდარტული გადახრა იმ ფორმით, როგორც იყო ნათქვამი თეორემაში, სრულ თანხმობაშია იმასთან, რაც მივიღეთ U -ს განაწილების 5.10 ცხრილიდან.

სხვა მნიშვნელოვანი გამოყენება თეორემა 5.1-ის ეხება n რაოდენობის *iid* შემთხვევით ცვლადებს, სადაც თითოეული a_i არის $\frac{1}{n}$. ეს ის შემთხვევაა, როდესაც $X_1, X_2, \dots, X_n$ შემთხვევითი ცვლადები კონცეპტუალურად ეკვივალენტურნი არიან შემთხვევითი ამორჩევების მიმართ, (დაბრუნების გარეშე), თითოეული პოპულაციიდან. თეორემა 5.1 ამბობს, რომ ასეთი შემთხვევითი სიდიდის

$$\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n$$

საშუალო და დისპერსია დამოკიდებულია პოპულაციის μ და σ^2 პარამეტრებზე. დამოუკიდებელი $X_1, X_2, \dots, X_n$ ცვლადებისთვის, საერთო μ საშუალოთი და σ^2 დისპერსიით, თეორემის მიხედვით გვაქვს, რომ

n ცალი iid შემთხვევითი სიდიდის მათემატიკური ლოდინების საშუალო არითმეტიკული ანუ საშუალოების საშუალო

$$E\bar{X} = \frac{1}{n}EX_1 + \frac{1}{n}EX_2 + \dots + \frac{1}{n}EX_n = n\left(\frac{1}{n}\mu\right) = \mu$$

და n ცალი iid შემთხვევითი სიდიდის საშუალოების დისპერსია

$$\begin{aligned} \text{Var } \bar{X} &= \left(\frac{1}{n}\right)^2 \text{Var } X_1 + \left(\frac{1}{n}\right)^2 \text{Var } X_2 + \cdots + \left(\frac{1}{n}\right)^2 \text{Var } X_n = \\ &= n \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

იმის გამო, რომ $\frac{\sigma^2}{n}$ კლებულობს n -ის ზრდასთან ერთად, ეს ფორმულები იძლევიან იმედის მომცემ სურათს იმისას, რომ $\bar{X}$ -ს გააჩნია ალბათობის განაწილება, რომლის ცენტრია პოპულაციის μ საშუალო. გაფანტულობა კი მცირდება ისე, როგორც იზრდება შერჩევის ზომა.

5.5. ცენტრალური ზღვარითი ეფექტი

საინჟინრო გამოყენებებში ყველაზე ხშირად ხმარებადი სტატისტიკა არის შერჩევის საშუალო. მოყვანილი ფორმულები აკავშირებენ ალბათობის განაწილების საშუალოსა და დისპერსიას მათ ერთჯერად დაკვირვებებთან, როდესაც *iid* არის შესატყვისი მოდელი. გამოყენებითი ალბათობის ერთერთი უმნიშვნელოვანესი ფაქტი არის ის, რომ თუ შერჩევის ზომა საკმარისად დიდია, მაშინ აგრეთვე შესაძლებელია მოვახდინოთ $\bar{X}$ -ის ალბათობის განაწილების ფორმის აპროქსიმაცია დამოუკიდებელ განაწილების ფორმებთან, რომლებიც საფუძვლად უდევს ინდივიდუალური დაკვირვებებს, ანუ სამართლიანია შემდეგი ფაქტი:

ცენტრალური ზღვარითი თეორემა

თუ $X_1, X_2, \dots, X_n$ არიან *iid* შემთხვევითი სიდიდეები, რომელთა საშუალოა μ და დისპერსიაა σ^2 , მაშინ საკმარისად დიდი n -სთვის $\bar{X}$ მიახლოებით არის ნორმალური განაწილება ანუ თუ $n \rightarrow \infty$, მაშინ $X_1 + X_2 + \cdots + X_n$ ჯამის განაწილება მიისწრაფვის

ნორმალური განაწილებისკენ ე.ი. შესაძლებელია $\bar{X}$ -ის მიახლოებითი ალბათობების გამოთვლა ნორმალური განაწილების საშუალებით, რომლის საშუალოა μ და დისპერსიაა σ^2/n .

სავარჯიშოები:

1. ახსენით რაოდენობრივ ტერმინებში რას ნიშნავს ორი X და Y ცვლადის დამოუკიდებლობა. რა უპირატესობა გააჩნია იმას, რომ X და Y ცვლადები შეიძლება აღიწეროს როგორც დამოუკიდებლები.
2. ხარისხის აუდიტის ჩანაწერებში ინახება ე.წ. დიდი და მცირე წარუმატებლობის სქემის პაკეტი, ელექტრონული საკომუნიკაციო მოწყობილობის მწყობრიდან გამოსვლამდე. ისინი მიუთითებენ, რომ კონკრეტული მოწყობილობის ტიპისთვის ეს ყველაფერი მიახლოებით შეიძლება აღიწეროს შემდეგი შემთხვევითი ცვლადებით:

X = დიდი წარუმატებლობის რაოდენობა

და

Y = მცირე წარუმატებლობის რაოდენობა,

რომელთაც გააჩნიათ დაახლოებით ასეთი თანმხლები ერთობლივი განაწილება

$y \setminus x$	0	1	2
0	.15	.05	.01
1	.10	.08	.01
2	.10	.14	.02
3	.10	.08	.03
4	.05	.05	.03

(ა) იპოვეთ ზღვრული ალბათობის ფუნქციები X და Y შემთხვევითი სიდიდეებისთვის, ანუ $f_X(x)$ და $f_Y(y)$.

(ბ) არიან თუ არა X და Y დამოუკიდებლები? ახსენით.

(გ) იპოვეთ საშუალო და დისპერსია $X - EX$ -სთვის და $Var X$.

(დ) იპოვეთ საშუალო და დისპერსია $Y - EY$ -სთვის და $Var Y$.

(ე) იპოვეთ Y -ის პირობითი ალბათობის ფუნქცია $X = 0$ პირობით, ანუ ეს ნიშნავს, რომ არ გვაქვს დიდი წარუმატებლობა, ანუ იპოვეთ $f_{Y|X}(y|x)$.

რა არის ამ პირობითი განაწილების საშუალო?

3. დამზადებულია მანქანის საკისრების ღეროები და რგოლები. ღერო უნდა მოერგოს რგოლს. მისი მოდელი

X = ღეროს დიამეტრი,

Y = რგოლის დიამეტრი

დამოუკიდებელი შემთხვევითი სიდიდეებია, X -ის უნიფორმა არის (1.97; 2.02), ხოლო Y -ის (2.00; 2.06). ($f_X(x) = 1/.05$, როცა $1.97 < x < 2.02$, ხოლო სხვა შემთხვევაში $f_X(x) = 0$. ანალოგიურად, $f_Y(y) = 1/.06$, როცა $2.00 < y < 2.06$, ხოლო სხვა შემთხვევაში $f_Y(y) = 0$). ამ მოდელისთვის შეასრულეთ შემდეგი:

(ა) დაწერეთ ერთობლივი ალბათობის სიმკვრივე X და Y -სთვის. (იგი იქნება დადებითი მხოლოდ მაშინ, როცა $1.97 < x < 2.02$ და $2.00 < y < 2.06$).

(ბ) გამოთვალეთ $P[Y - X < 0]$, ალბათობა იმისა, რომ კომპლექტში შეუთავსებლობაა.

4. ნომინალური სისქის $\frac{3}{4}$ დუიმიანი ფანერა შედგება ხუთი ფენისგან. ფენების სისქე უხეზად რომ აღვწეროთ, შეიძლება ჩავთვალოთ, რომ დამოუკიდებელი შემთხვევითი სიდიდეებია. მათი საშუალოები და სტანდარტული გადახრები არის შემდეგი:

ფენა	საშუალო(in.)	სტანდარტული გადახრა(in.)
1	.094	.001
2	.156	.002
3	.234	.002
4	.172	.002
5	.094	.001

იპოვეთ მთლიანი სისქის საშუალო და სტანდარტული გადახრა, რომელიც ასოცირებულია ამ ინდივიდუალურ სიდიდეებთან.

6. ფორმალური სტატისტიკური დასკვნები

6.1. დიდი შერჩევის ნდობის ინტერვალი საშუალოსთვის

სტატისტიკის უამრავი საინჟინრო გამოყენებები მორგებულია შემდეგ სტანდარულ სიტუაციაზე. მონაცემთა პარამეტრების მნიშვნელობები როგორც წესი, უცნობი სიდიდეებია. მონაცემებზე დაყრდნობით საჭიროა:

1. განისაზღვროს მნიშვნელობათა ინტერვალი, რომელიც სავარაუდოდ შეიცავს უცნობ პარამეტრს (ან ფუნქციას, რომელიც დამოკიდებულია ერთ ან მეტ პარამეტრზე) და
2. რაოდენობრივად „რამდენად ალბათურად“ ფარავს ეს ინტერვალი ზუსტ მნიშვნელობას.

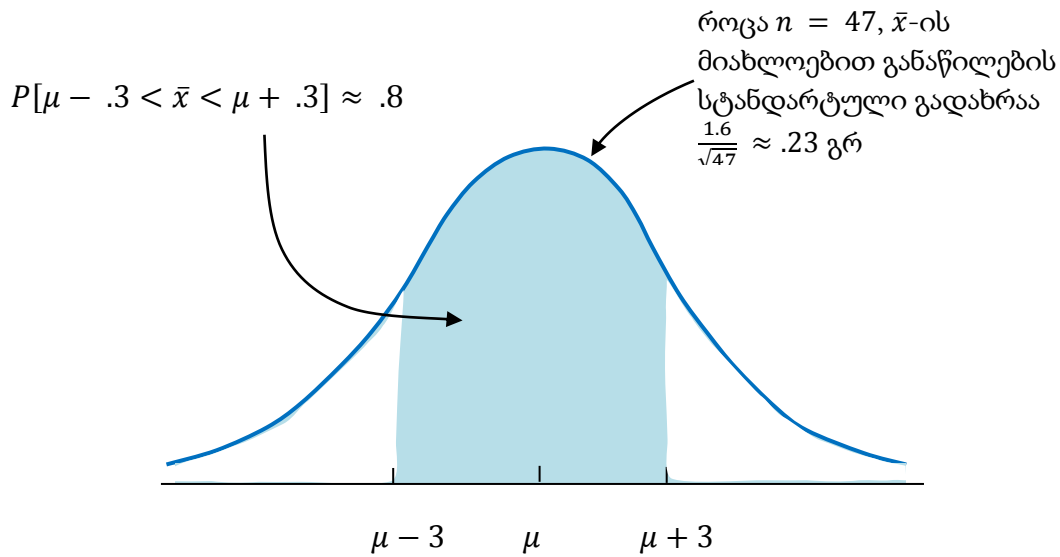
მაგალითად, მოწყობილობას, რომელიც ბავშვის საკვებს ათავსებს ქილებში, გააჩნია შევსების საშუალო დონე μ , რომელიც უცნობი სიდიდეა. საჭიროა განისაზღვროს ინტერვალი, რომელიც სავარაუდოდ შეიცავს μ -ს და დადგინდეს საიმედოობა, რამდენად მნიშვნელოვანია ეს ინტერვალი. ასეთი ტიპის სტატისტიკური დასკვნებს, რომლებიც განკუთვნილია იმისთვის, რომ გადაწყვიტოს მსგავსი პრობლემები, ეწოდება *ნდობის ინტერვალის შეფასება*.

n -სიდიდის ნდობის ინტერვალი μ -სთვის, როდესაც σ ცნობილია.

გავიხსენოთ ამოცანა ბავშვის საკვებით ქილების შევსების შესახებ, სადაც სუფთა წონის სტანდარტული გადახრა $\sigma = 1.6$ გრამია. შერჩევის საშუალოს შესაფასებლად გავითვალისწინოთ, რომ, რადგანაც n დიდია, ამიტომ *iid* შემთხვევითი სიდიდეები დაახლოებით ნორმალურია. ექსპერიმენტით დადგინდა, რომ, მაგალითად, როცა $n=47$ და

$\bar{x}$ = ქილების შევსების პროცესში შერჩევის ზუსტი წონების საშუალო (გრ),

მაშინ არსებობს 80% -იანი შანსი იმისა, რომ $\bar{x}$ მოთავსებულია $[\mu - .3; \mu + .3]$ შუალედში. ეს ფაქტი გამოსახულია შემდეგ ნახაზზე.



გრაფიკი 6.1. $\bar{x}$ -ის მიახლოებითი ალბათობის განაწილება, როცა $n = 47$

შევნიშნოთ, რომ $\bar{x}$ აღნიშვნას ვიყენებთ შერჩევის საშუალოს აღსანიშნავად. იგი საკმაოდ გავრცელებული აღნიშვნაა სტატისტიკაში.

გრაფიკზე პირველივე დაკვირვებისას გვიჩნდება აზრი იმის შესახებ, თუ რა არის იმის შესაძლებლობა, რომ $\bar{x}$ ჩავარდება

$$\mu - .3 < \bar{x} < \mu + .3 \quad (6.1)$$

შუალედში, რომლის სიგრძეა $2 \cdot 0.3 = 0.6$ და ცენტრი არის μ . შესაძლებელია განვიხილოთ მისი ეკვივალენტური სიტუაცია, ანუ არსებობს თუ არა 0.6 სიგრძის ინტერვალი ცენტრით $\bar{x}$, რომელშიც ჩავარდება μ სიდიდე. ალგებრულად წინა უტოლობა ტოლფასია შემდეგი უტოლობის

$$\bar{x} - .3 < \mu < \bar{x} + .3, \quad (6.2)$$

რომელიც ჩვენს ყურადღებას წარმართავს მეორე მიმართულებით. ცხადია, რომ (6.2) გამოსახულებას გააჩნია შანსი, დაახლოებით 80%, რომ $n = 47$ ზომის შერჩევის წონები ხვდებიან

$$(\bar{x} - .3; \bar{x} + .3) \quad (6.3)$$

ინტერვალში, რომელიც შეიძლება ჩაითვალოს μ -ს ნდობის ინტერვალად 80% -იანი საიმედოობით.

მაგალითი 6.1. შევსების წონის საშუალოს ნდობის ინტერვალი.

კვლავ განვიხილოთ შემთხვევა, რომ შერჩევის ზომაა $n=47$ და $\bar{x} = 138.2$ გრ. მაშინ (6.3) გამოსახულება გვთავაზობს ინტერვალს, რომლის კიდურა წერტილებია

$$138.2 \text{ გრ} \pm .3 \text{ გრ}$$

(ეს ნიშნავს, რომ ინტერვალი 137.9 გრ -დან 138.5 გრ -მდე), გამოიყენება როგორც შევსების წონის საშუალოს 80%-იანი ნდობის ინტერვალი.

ძნელი არ არის (6.3) გამოსახულების ლოგიკური განზოგადება. როგორც ვიცით, *iid* მოდელი შეესაბამება დიდი შერჩევის ელემენტებს. ცენტრალური ზღვარითი თეორემიდან გამომდინარეობს, რომ შერჩევის საშუალო $\bar{x}$ დაახლოებით არის ნორმალური, რომლის საშუალოა μ და სტანდარტული გადახრაა $\sigma/\sqrt{n}$. მაშინ $p > .5$ რიცხვისთვის z არის სტანდარტული ნორმალური განაწილების p ქვანტილი. მაშინ ალბათობა იმისა, რომ

$$\mu - z \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z \frac{\sigma}{\sqrt{n}} \quad (6.4)$$

დაახლოებით არის $1 - 2(1 - p)$. ეს უტოლობა შეიძლება ასეც ჩაიწეროს:

$$\bar{x} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \frac{\sigma}{\sqrt{n}} \quad (6.5)$$

და საბოლოოდ მივიღებთ ინტერვალს, რომლის კიდურა წერტილებია:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \quad (6.6)$$

ე.ი. ინტერვალი აღნიშნული წერტილებით არის ნდობის ინტერვალი μ საშუალოსთვის (სანდოობის დონით ანუ ალბათობით $1 - 2(1 - p)$.)

ამ ტოლობაში z შერჩეულია ისე, რომ სტანდარტული ნორმალური ალბათობა $-z$ -სა და z -ს შორის შესაბამება სანდოობის სასურველ დონეს. ცხრილი B.1 (სტანდარტული ნორმალური დაგროვებითი ალბათობები) შესაძლებელია გამოვიყენოთ შესაბამისობების შესამოწმებლად, რომლებიც მოცემულია, შემდეგ ცხრილში:

სასურველი სანდობა	z
80%	1.28
90%	1.645
95%	1.96
98%	2.33
99%	2.58

ცხრილი 6.2. n - დიდი ინტერვალები μ -სთვის

ეს ცხრილი გაძლევს z -ის მნიშვნელობებს, რომელებიც საჭიროა მონიშნულ ფორმულაში ზოგიერთი ნდობის დონის დასადგენად.

6.2. ზოგადად გამოყენებადი ნდობის ინტერვალი μ -სთვის

გამოსახულება (6.6) წარმოგვიდგენს მათემატიკურად სწორ ფორმულას, რომელიც მოხერხებულად გამოყენებადია, როცა σ ცნობილია. თუმცა ეს არ არის ჩვეულებრივი მოვლენა, რომ ვაფასებთ μ -ს და ამ დროს σ ცნობილია. ასეთი სიტუაციები გამოჩნდებიან ხოლმე ქარხანულ პროცესთან დაკავშირებულ ამოცანებში, როგორც იყო, მაგალითად, ქილების შევსების ამოცანები. წარსულში მიღებულმა მნიშვნელოვანმა გამოცდილებებმა ზოგჯერ შეიძლება მოგვცენ საკმარის გონივრული მნიშვნელობა σ -სთვის. სანამ ფიზიკური პროცესები მიმდინარეობს, არის მომენტები, როცა შესაძლებელი ხდება ამოცანაში μ -ს მიმდინარე მნიშვნელობების გამოყენება.

საბედნიეროდ, იმ მსჯელობების მოდიფიკაცია, რომელმაც გამოიწვია ფორმულა (6.6), წარმოგვიდგენს ნდობის ინტერვალის ფორმულას μ -სთვის და დამოკიდებულია მხოლოდ შერჩევის მახასიათებლებზე. არგუმენტი, რომელსაც მივყავართ (6.6) ფორმულამდე დამოკიდებულია ფაქტზე, რომ დიდი n -სთვის $\bar{x}$ მიახლოებით არის ნორმალური განაწილება μ საშუალოთი და სტანდარტული გადახრით $\sigma/\sqrt{n}$, ანუ

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (6.7)$$

არის სტანდარტული ნორმალური. აქაც გამოჩნდა σ , რომელსაც ადგილი ჰქონდა (6.6) ფორმულაში. მაგრამ ამ შემთხვევაში ცენტრალური ზღვართი თეორემის მსუბუქი განზოგადება იძლევა იმის გარანტიას, რომ დიდი n -სთვის

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad (6.8)$$

არის აგრეთვე სტანდარტული ნორმალური, და (6.8) ფორმულა აღარ შეიცავს σ -ს.

ვიწყებთ იმით, რომ თუ მოდელი შეესაბამება *iid*-ს და n საკმარისად დიდია, მაშინ (6.8) ცვლადი მიახლოებით არის სტანდარტული ნორმალური და დანარჩენი მსჯელობა თითქმის იგივეა, რაც მანამდე იყო. დადებითი z -ებისთვის, უტოლობა

$$-z < \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} < z$$

ეკვივალენტურია შემდეგი უტოლობის:

$$\mu - z \frac{s}{\sqrt{n}} < \bar{x} < \mu + z \frac{s}{\sqrt{n}} \quad (6.4)$$

ეს კი თავის მხრივ ეკვივალენტურია უტოლობის:

$$\bar{x} - z \frac{s}{\sqrt{n}} < \mu < \bar{x} + z \frac{s}{\sqrt{n}} .$$

ე.ი. ინტერვალი, რომლის ცენტრია შემთხვევითი $\bar{x}$ და შემთხვევითი სიგრძე— $2zs/\sqrt{n}$, შემთხვევითი კიდურა წერილობითა:

$$\bar{x} \pm z \frac{s}{\sqrt{n}} \quad (6.5)$$

ეს წერტილები შეიძლება უშუალოდ გამოვიყენოთ μ -ს ნდობის ინტერვალის ასაგებად. სასურველი სანდოობისთვის შესაძლებელია z ისე შეირჩეს, რომ სტანდარტული ნორმალური ალბათობა $-z$ -სა და z -ს შორის (ესენი ინტერვალის საზღვრებია) შეესაბამებოდეს სანდოობის დონეს.

6.3. რამდენიმე დამატებითი კომენტარი ნდობის ინტერვალთან

დაკავშირებით

ფორმულები (6.6) და (6.9) ნდობის ინტერვალების დასადგენად გამოიყენება შემდეგი ტიპის ამოცანების შემთხვევაში: „ μ მოთავსებულია a და b მნიშვნელობებს შორის“. მაგრამ ხშირად ამოცანა ყალიბდება ასედაც: „ μ არის არანაკლებ c “, ან „ μ არ აღემატება d -ს“, რომელთაც ზოგჯერ შეიძლება ჰქონდეთ მეტი პრაქტიკული მნიშვნელობა. მაგალითად, ავტოინჟინერს შეიძლება ჰქონდეს სურვილი გადაწყვიტოს ამოცანა: „მოცმული ძრავისთვის NO emission -ის (არა გაფრქვევა) საშუალო არის არაუმეტეს 5 ppm (ნაწილაკი მილიონზე). ან სამოქალაქო ინჟინერს აინტერესებს საკითხი, ვთქვათ ასეთი: „კუმშვაზე გამძლეობის საშუალო შესაბამისი ტიპის ნიმუშებისთვის არის არანაკლებ 4 188 psi (ფუნტი კვადრატულ დუიმზე)“. ეს პრაქტიკული საინჟინრო ამოცანები პიდაპირ მიგვითითებენ ცალმხრივ ნდობის ინტერვალზე.

ცალმხრივი ინტერვალების მიღება

არანაირ სირთულეს არ წარმოადგენს ფორმულებიდან ცალმხრივი ინტერვალების მიღება. თუ გვაქვს ქმედითი ორმხრივი ფორმულა, მაშინ უნდა მოვიქცეთ შემდეგნაირად:

1. შევცვალოთ ქვედა საზღვარი $-\infty$ -ით, ან ზედა $+\infty$ -ით.
2. შეიცვალოს მითითებული ნდობის დონე სათანადოდ ზევით (ზოგადად ეს ნიშნავს, რომ „არასანდოობის დონე“ გავყოთ ორზე.)

ეს მითითებული პროცედურა გამოიყენება არა მხოლოდ (6.6) და (6.9) ფორმულებისთვის, არამედ ყველა დანარჩენი ორმხრივი ნდობის ინტერვალისთვისაც.

მეორე მნიშვნელოვანი საკითხი ნდობის ინტერვალების გამოყენებისას, არის *ნდობის ინტერვალის* ტექნიკური მნიშვნელობის სწორად გაგების პრობლემა. სამწუხაროდ, არსებობს მრავალი შესაძლო გაურკვეველობა. ამიტომ მნიშვნელოვანია, გარკვეულად ჩამოყალიბდეს ნდობის ინტერვალის რას ნიშნავს და რას არ ნიშნავს.

მაგალითად, შერჩევის აღება და მისი ჩასმა (6.6) და (6.9) ფორმულებში ნიშნავს, რომ სანდოობის დონე თვალსჩინოა. ავირჩიოთ (ორმხრივი) 90%-იანი ნდობის დონე და დავუშვათ, შესაბამისი z -ის მნიშვნელობაა, ვთქვათ, $z=1.645$, რომელიც გამოიყენება ფორმულა (6.9)-ში. მაშინ ვამბობთ, არის 90%-იანი შანსი იმისა, რომ μ მოხვდება ინტერვალში

$$\bar{x} - 1.645 \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.645 \cdot \frac{s}{\sqrt{n}} ,$$

ან შეიძლება ეს ასედაც ჩაიწეროს:

$$P \left[\bar{x} - 1.645 \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.645 \cdot \frac{s}{\sqrt{n}} \right] \approx .90$$

შერჩევის ზომა μ -ს შესაფასებლად

შევნიშნოთ, რომ ნდობის ინტერვალის განხილვისას ფორმულები (6.6) და (6.9) ტოვებენ შთაბეჭდილებას, რომ ისინი იძლევიან არც თუ ისე ზუსტ რაოდენობრივ პასუხს კითხვაზე: „რამდენად დიდი უნდა იყოს n ” (6.9) ფორმულაში? მაგალითად, ვთქვათ, გინდათ

(1) - სასურველი სანდოობის დონე, ან

(2) - შერჩევის სტანდარტული გადახრის ყველაზე ცუდი შემთხვევის რა მოლოდინი არსებობს, ან

(3) - სასურველი სიზუსტე μ -ს შეფასებაში.

უფრო მარტივი ამოცანაა, ვიპოვოთ შესაბამისი შერჩევის ზომა. დავუშვათ, რომ სასურველი სანდოობის დონე მიგვითითებს, გამოვიყენოთ z -ის მნიშვნელობა (6.9) ფორმულაში, ხოლო s არის შერჩევის სტანდარტული გადახრის სავარაუდოდ ყველაზე ცუდი მნიშვნელობა. დავუშვათ, გვინდა გვქონდეს $\bar{x} \pm \Delta$ ფორმის სანდოობის ზღვრები.

განვსაზღვროთ

$$\Delta = z \cdot \frac{s}{\sqrt{n}}$$

და ვიპოვოთ აქედან n ,

$$n = \left(\frac{zs}{\Delta} \right)^2 .$$

ზოგჯერ შესაძლებელია მოხდეს, რომ შერჩევის ზომის გამოთვლების შედეგად მივიღოთ განსხვავებული პასუხები და არა მყარი ზუსტი რიცხვი. თუ გამოთვლები ისეთია, რომ რომ ჯერ მივიღეთ მცირე სიდიდე და მერე დიდი, ანუ მნიშვნელობები

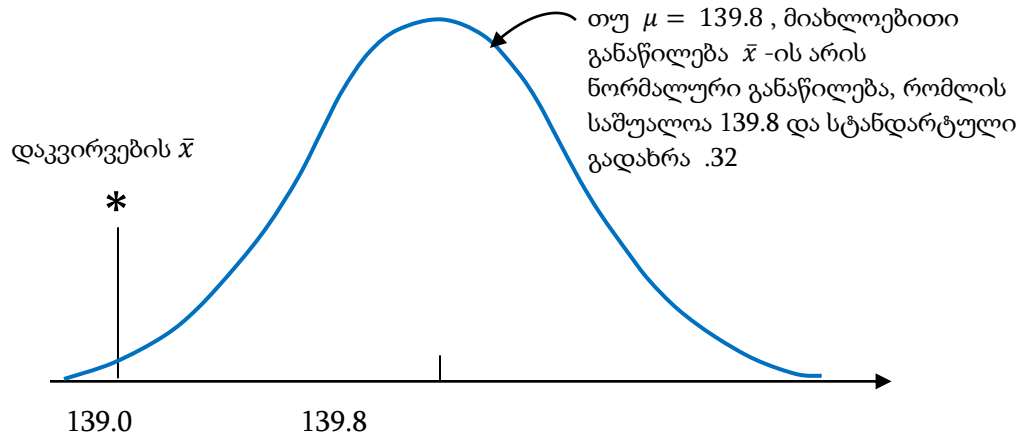
თანდათან იზრდება, მაშინ ეს არის საკმარისი გარანტია იმისა, რომ შესაძლებელია (6.9) ფორმულის გამოყენება.

6.4. დიდი შერჩევის მნიშვნელობის ტესტი საშუალოსთვის

საინტერესოა, რის საშუალებას გვაძლევს ალბათობა ნდობის ინტერვალის შეფასების დროს. მნიშვნელობის შეფასების ტესტი იყენებს მონაცემებს იმისთვის, რომ რიცხობრივად შეაფასოს შერჩევის პარამეტრების მნიშვნელობათა დამაჯერებლობა. მნიშვნელობის ტესტს ჩვენ აღვწერთ μ საშუალოსთვის, როდესაც სტანდარტული გადახრა σ ცნობილია.

კვლავ გავიხსენოთ ქილების შევსების პროცესი და დავუშვათ, რომ სუფთა წონისთვის ცნობილია $\sigma = 1.6$ გრ. დავუშვათ, რომ ეტიკეტზე აღნიშნული წონაა 135გრ. საიჟინრო პროცესს გააჩნია მიზანი, რომ სუფთა წონის საშუალო იყოს $135 + 3\sigma = 139.8$ გრამამდე. ბოლოს, დავუშვათ, რომ შევსების შემოწმების რუტინული პროცესი მიზნად ისახავს წარმოების პროცესში აღმოაჩინოს, ცვლილება ხომ არ არის საშუალოს დასაშვებ მნიშვნელობებში. შერჩევა აიღება $n = 25$ ქილა და შევსების $\bar{x} = 139.0$ გრ. რა შეუძლია გვითხრას ამ სიდიდემ მიმდინარე პროცესის დამაჯერებლობაზე, თუ რალური სამიზნე არის 139.8 გრ?

აქ შეიძლება მოვუხმოთ ცენტრალურ ზღვარით თეორემას. თუ მართლაც საშუალო არის 139.8 გრ და $\bar{x}$ -ს გააჩნია დაახლოებით ნორმალური განაწილება, რომლის სტანდარტული გადახრაა $\sigma/\sqrt{n} = 1.6/\sqrt{25} = .32$ გრ. შემდეგ გრაფიკზე გამოსახულია ეს სიტუაცია, როდესაც დასაკვირვებელი სიდიდეა $\bar{x} = 139.0$ გრ.

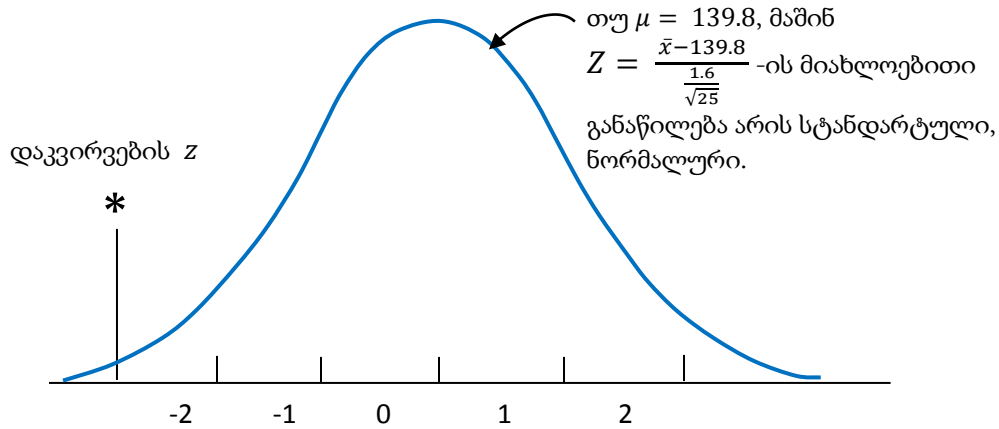


გრაფიკი 6.3. მიახლოებითი ალბათობის განაწილება $\bar{x}$ -სთვის, როდესაც $\mu = 139.8$ და დასაკვირვებელი სიდიდე $\bar{x} = 139.0$

შემდეგი გრაფიკი გვიჩვენებს სტანდარტულ ნორმალურ განაწილებას, რომელიც შეესაბამება 6.3 გრაფიკს. იგი დამყარებულია იმ ფაქტზე, რომ, თუ მიმდინარე პროცესის საშუალო არის მიზნის მნიშვნელობა 139.8 გრ, მაშინ $\bar{x}$ დაახლოებით ნორმალურია μ საშუალოთი და $\sigma/\sqrt{n} = .32$ გრ სტანდარტული გადახრით. აქედან გამომდინარეობს, რომ

$$Z = \frac{\bar{x} - 139.8}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 139.8}{.32}$$

დაახლოებით არის სტანდარტული ნორმალური განაწილება. დაკვირვებად $\bar{x} = 139.0$ მნიშვნელობას 6.3 გრაფიკიდან შეესაბამება $z = -2.5$, რომელიც წარმოდგენილია შემდეგ სურათზე:



გრაფიკი 6.4. გრაფიკი 6.3 -ის შესაბამისი სტანდარტული ნორმალური სურათი

ცხადია, რომ ან გრაფიკი 6.3 -დან ან გრაფიკი 6.4 -დან, თუ პროცესის საშუალო არის მიზნის მნიშვნელობა 139.8 გრ (და აქედან გამომდინარე გრაფიკებიც სწორია), მაშინ საკმაოდ ექსტრემალური / იშვიათად დაკვირვებადია $\bar{x}$, ან იგივეა რაც z . ცხადია, რომ ექსტრემალური / იშვიათი მოვლენები ზოგჯერ შეიძლება მოხდეს, მაგრამ ეს ბუნებაში იმდენად იშვიათად ხდება, რომ დაკვირვების $\bar{x}$ (ან იგივეა რაც z) იძლევიან იმის საშუალებას, რომ ჩაითვალოს, პროცესი მიზნის მნიშვნელობაზე არადამაჯერებელია.

გრაფიკები გვიჩვენებენ იმასაც, რომ „ექსტრემალური მნიშვნელობები არა მხოლოდ μ -ს ქვემოთაა, არამედ ზემოთაც. ანუ არა მხოლოდ დაკვირვების მნიშვნელობები $x \leq 139.0$ გრ (ან $z \leq -2.5$) არის ექსტრემალური, არამედ ექსტრემალურია აგრეთვე $x \geq 140.6$ გრ (ან $z \geq 2.5$). (პირველი ხდება მაშინ, როცა μ კლებულობს და მეორე - როცა μ იზრდება). მიზნის მნიშვნელობათა არადამაჯერებლობა რაოდენობრივად შემდეგნაირად შეიძლება გამოისახოს Φ ფუნქციის საშუალებითაც:

$$\Phi(-2.5) + (1 - \Phi(2.5)) = .01$$

განმარტება 6.1. სტატისტიკური მნიშვნელობის ტესტირება არის მინაცემების საშუალებით რაიმე ცდის პარამეტრის (ან ფუნქციის) დამაჯერებლობის რაოდენობრივი შეფასება.

ლოგიკურად, მნიშვნელობის ტესტი იწყება ცდის ან ჰიპოტეზური მნიშვნელობის განსაზღვრით.

განმარტება 6.2. ნულოვანი ჰიპოთეზა არის დებულება შემდეგი ფორმის

$$\text{პარამეტრი} = \#$$

ან

$$\text{პარამეტრების ფუნქცია} = \#$$

(ზოგიერთი $\#$ რიცხვისთვის). ნულოვანი ჰიპოთეზა, როგორც წესი, განასახიერებს პარამეტრს და ზოგადად აღნიშნავენ H_0 -ით. ნულოვანი ჰიპოთეზის აღნიშვნას ცენტრალური ადგილი უკავია ჰიპოთეზათა ტესტების ჩაწერაში. ფრაზა „ნულოვანი ჰიპოთეზა“ არის დებულება, რომ არ არის *განსხვავება* ან ტოლია. მაგალითად, სტანდარტული ჩანაწერი

$$H_0 : \mu = 139.8$$

ნიშნავს, რომ არ არის განსხვავება μ -სა და მიზნის 139.8 გრ მნიშვნელობას შორის.

განმარტება 6.2. ალტერნატიული ჰიპოთეზა არის ნულოვანი ჰიპოთეზის საპირისპირო დებულება. იგი განსაზღვრავს იმ ფორმებს, რომლებიც ნულოვანი ჰიპოთეზის ამოღების შემდეგ განიცდიან შეშფოთებას. ალტერნატიული ჰიპოთეზა ზოგადად აღნიშნება H_a -თი. იგი ემყარება იმავე მონაცემებს, რასაც ნულოვანი ჰიპოთეზა, იმ განსხვავებით, რომ ტოლობის ნიშანი უნდა შეიცვალოს ნიშნებით $\neq$, $>$ ან $<$.

ამ ორი განმარტებიდან მხოლოდ μ საშუალოსთვის ნულოვანი ჰიპოთეზის სამი შესაძლო ალტერნატიული ჰიპოთეზების წყვილებია:

$$H_0 : \mu = \#$$

$$H_0 : \mu = \#$$

$$H_0 : \mu = \#$$

$$H_a : \mu > \#$$

$$H_a : \mu < \#$$

$$H_a : \mu \neq \#$$

თუ ქილების შევსების ამოცანაში წონაში დაკლების შესაძლებლობაა $\mu < 139.8$ გრ, ხოლო გადავსების შესაძლებლობაა $\mu > 139.8$ გრ, მაშინ მათი შესაბამისი ალტერნატიული ჰიპოთეზა იქნება:

$$H_a: \mu \neq 139.8$$

სტატისტიკური ტესტი არის კერძო შენთხვევა იმ რიცხვითი მონაცემების, რომლებიც გამოიყენება მნიშვნელობთა ტესტში. იგი ზოგადად შეიცავს იმ რიცხვებს, რომლებიც გამოჩნდებიან ხოლმე ნულოვან ჰიპოთეზაში. ჩვენს შემთხვევაში ფორმულა

$$Z = \frac{\bar{x} - 139.8}{\frac{\sigma}{\sqrt{n}}}$$

არის სპეციალიზებული ფორმა ზოგადი (დიდი - n , ცნობილი σ) სტატისტიკური ტესტი μ -სთვის და მას აქვს ასეთი სახე:

დიდი შერჩევის სტატისტიკური ტესტი μ -სთვის, როდესაც σ ცნობილია

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ჩვენი სცენარისთვის μ -ს ჰიპოთეზური მნიშვნელობა არის 139.8 , $n = 25$ და $\sigma = 1.6$. შეიძლება ვიფიქროთ, რომ ამ ტიპის სტატისტიკური ტესტი ყველაზე უფრო მოსახერხებელია სტანდარტული ფორმით (იმ ფორმით, როგორც არის უკანასკნელ ფორმულაში), ვიდრე როგორც თავად $\bar{x}$ იყო. ეს ფორმულა იმაზეც მიგვანიშნებს, რომ განაწილება ყოველთვის იქნება ერთი და იგივე —კერძოდ, სტანდარტული ნორმალური.

თუ გავაგრძელებთ ამ მაგალითს და ნულოვანი ჰიპოთეზის მაგიერ განვიხილავთ ალტერნატიულ

$$H_a: \mu \neq 139.8$$

ჰიპოთეზას, დასაკვირვებელი $\bar{x}$ ძალიან მეტია ან ძალიან ნაკლებია 139.8-ზე, მიისწრაფვის შედეგისკენ. ანუ დაკირვების ასეთი მნიშვნელობა უარყოფს ნულ ჰიპოთეზის არადამაჯერებლობას.

მივუთითოთ, როგორ შეიძლება, გამოვიყენოთ მონაცემები ნულოვანი ჰიპოთეზის დამაჯერებლობის დასადგენად. ამისთვის საჭიროა, მოვაგროვოთ მონაცემები,

ჩავრთოთ ისინი სტატისტიკური ტესტის ფორმულაში და მითითებული განაწილებისა და გამოთვლილი სიდიდეების საშუალებით მივიღებთ H_0 ჰიპოთეზის რაოდენობრივ შეფასებას.

განმარტება 6.3. მნიშვნელობებზე დაკვირვების დონე ან p -მნიშვნელობა არის ალბათობა იმისა, რომ მითითებული განაწილება აფასებს სტატისტიკური ტესტის იმ შესაძლო მნიშვნელობათა სიმრავლეს, რომლებიც არიან რეალურად დაკვირვებადი უკიდურესი მნიშვნელობები (იმ ტერმინებში, რომ ეჭვი შეგვაქვს H_0 ჰიპოთეზის ჭეშმარიტებაში).

მცირე p -მნიშვნელობები არიან მტკიცებულებები H_0 -ის წინააღმდეგ

მნიშვნელობებზე დაკვირვებათა მცირე დონე არის ძლიერი მტკიცებულება ნულოვანი ჰიპოთეზის მოქმედების წინააღმდეგ. შევსების პროცესის ამოცანაში სტატისტიკური ტესტის დაკვირვებადი სიდიდის

$$z = -2.5$$

შეფასების p -მნიშვნელობა ანუ დაკვირვების დონე არის

$$\Phi(-2.5) + (1 - \Phi(2.5)) = .01 ,$$

რომელიც იძლევა საკმაოდ ძლიერ მტკიცებულებას იმის წინააღმდეგ, რომ პროცესის საშუალო შესაძლოა იყოს მიზნის მნიშვნელობა.

ხუთსაფეხურიანი ფორმატი მნიშვნელობათა ტესტის შეჯამებისთვის

პირველი საფეხური: ჩამოვაცალიბოთ ნულოვანი ჰიპოთეზა.

მეორე საფეხური: ჩამოვაცალიბოთ ალტერნატიული ჰიპოთეზა.

მესამე საფეხური: ჩამოვაცალიბოთ ტესტის კრიტერიუმი. ამისთვის, უნდა დავწეროთ სტატისტიკური ტესტის ფორმულა (მასში ჩართული უნდა იყოს მხოლოდ ჰიპოთეზური მნიშვნელობები ნულოვანი ჰიპოთეზიდან და არანაირი ინფორმაცია შერჩევიდან) და მივუთითოთ განაწილება. შემდეგ ზოგად ტერმინებში ჩამოვაცალიბოთ რა არის სტატისტიკური ტესტის დაკვირვებადი მნიშვნელობა და რა წარმოადგენს ძლიერ მტკიცებულებას ნულოვანი ჰიპოთეზის წინააღმდეგ.

მეოთხე საფეხური: ვაჩვენოთ ნიმუშზე დაფუძნებული გამოთვლები.

მეხუთე საფეხური: წარმოვადგინოთ ანგარიში მნიშვნელობის დაკვირვებად დონეზე და განვაცხადოთ (რამდენადაც ეს შესაძლებელია) მისი შედეგები რეალური საინჟინრო ამოცანების კონტექსტში.

მაგალითი 6.2. მნიშვნელობის ტესტი შევსების პროცესის საშუალოსთან დაკავშირებით

ხუთ-საფეხურიანი მნიშვნელობათა ტესტის ფორმატი შეიძლება გამოვიყენოთ ადრე განხილული ქილების შევსების პროცესის შემთხვევისთვის.

1. $H_0 : \mu = 139.8$
2. $H_a : \mu \neq 139.8$
3. სტატისტიკური ტესტი არის

$$Z = \frac{\bar{x} - 139.8}{\frac{\sigma}{\sqrt{n}}}$$

მითითებული განაწილება არის სტანდარტული ნორმალური და დიდი დაკვირვებადი მნიშვნელობა $|z|$ წარმოადგენს მტკიცებულებას H_0 ჰიპოთეზის წინააღმდეგ.

4. შერჩევა გაგძლევს

$$Z = \frac{139.0 - 139.8}{\frac{1.6}{\sqrt{100}}} = -2.5$$

5. მნიშვნელობის დასაკვირვებელი დონე არის

$$\begin{aligned} &P[\text{სტანდარტული ნორმალური ცვლადი} \leq -2.5] + \\ &+ P[\text{სტანდარტული ნორმალური ცვლადი} \geq 2.5] = \\ &= P[|\text{სტანდარტული ნორმალური ცვლადი}| \geq 2.5] = 0.1 \end{aligned}$$

ეს არის გონივრული მკაცრი მტკიცებულება იმისა, რომ შევსების დონის საშუალო არ არის მიზნის მნიშვნელობა.

ზოგიერთი ფაქტი ალბათობის თეორიიდან საშუალებას იძლევა, რომ დიდი შერჩევის μ საშუალოს ნდობის ინტერვალი თავისუფალია σ -გან. შესაძლებელია აგრეთვე, გავაკეთოთ დიდი- n მნიშვნელობის ტესტი μ -სთვის σ -ს გარეშე. თუ n საკმარისად დიდი, მაშინ

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}},$$

არის მიახლოებით სტანდარტული ნორმალური. ეს ნიშნავს, რომ დიდი- n -თვის ტესტი

$$H_0 : \mu = \#$$

არის მარტივი მეთოდი, მაგრამ ფართოდ გამოყენებადი სტატისტიკისთვის

$$Z = \frac{\bar{x} - \#}{\frac{s}{\sqrt{n}}}.$$

მას ეწოდება *დიდი შერჩევის სტატისტიკური ტესტი μ -სთვის*.

სავარჯიშოები:

1. ახსენით და გააკეთეთ ინტერპრეტაცია დებულების:
„ინტერვალი 6.3 -დან 79.9 -მდე არის 95% -იანი ნდობის ინტერვალი μ საშუალოსთვის.
2. სტუდენტები ზომავენ მანქანის რაღაც ნაწილის დიამეტრს. იყენებენ ორ განსხვავებულ შტანგენფარგალს (ციფრულს და ვერნერის სკალის). მათი გაზომვების მონაცემები ჩაწერილია შემდეგი სახით. მოცემულია მცირე ფორმატის სიხშირეთა ცხრილი, ციფრული შტანგენფარგლით მიღებული 50 გაზომვის შედეგი.

დიამეტრი (მმ)	სიხშირე
4.52	1
4.66	4
4.67	7
4.68	7
4.69	14
4.70	9
4.71	4
4.72	4

- (ა) გამოთაღეთ შერჩევის საშუალო და სტანდარტული გადახრა ამ მონაცემებისთვის.
- (ბ) გამოიყენეთ (ა)-ში მიღებული შედეგები და იპოვეთ 98%-იანი ორმხრივი ნდობის ინტერვალი დიამეტრის საშუალოსთვის.
- (გ) გაიმეორეთ (ბ) ნაწილი 99%-იანი ნდობის ინტერვალისთვის. როგორ შევადაროთ ეს ინტერვალი (ბ)-ში მიღებულს?
- (დ) გამოიყენეთ (ა)-ში მიღებული შედეგები და იპოვეთ 98%-იანი სანდოობის ქვედა ზღვარი დიამეტრის საშუალოსთვის. (იპოვეთ რიცხვი # , ისეთი რომ $(# ; \infty)$ იყოს 98%-იანი ნდობის ინტერვალი.) როგორ ედრება ეს სიდიდე (ბ)-ში მიღებული ინტერვალის ქვედა კიდურა წერტილს?
- (ე) გაიმეორეთ (დ) ნაწილი 99%-იანი ნდობის ინტერვალისთვის. შეადარეთ აქ მიღებული მნიშვნელობა (დ)-ში მიღებულ პასუხს.
- (ვ) მოკლედ სტატისტიკურად აღწერეთ (ბ)-ში მიღებული ნდობის ინტერვალი.

დანართი

A.1. ელემენტარული ალბათობის ძირითადი ცნებები და აქსიომები

ალბათობის თეორიაში არსებობს ძირითადი, დაუყვანადი, პოტენციური შედეგები, რომელთაც ზოგჯერ შანსაც უწოდებენ. ამ შედეგების გავრცელებული ტერმინია ხდომილება, რომელიც მიეკუთვნება ე.წ. ელემენტარულ ხდომილებათა სივრცეს.

განმარტება A.1. შანსების სიტუაციიდან მხოლოდ ერთ პოტენციურ შედეგს, რომელიც ერთადერთი გზით შეიძლება განხორციელდეს **ელემენტარულ ხდომილებას** უწოდებენ, ხოლო ყველა ელემენტარულ ხდომილებათა ერთობლიობა ქმნის **ელემენტარულ ხდომილებათა სივრცეს**.

მათემატიკურად ელემენტარული ხდომილებები წარმოადგენენ უნივერსალური სიმრავლის წერტილებს და ეს სიმრავლე არის *ელემენტარულ ხდომილებათა სივრცე*. მას ხშირად აღნიშნავენ Ω ასოთი („ომეგა დიდი“). ამ სიმრავლის ის ქვესიმრავლეები, რომლებიც შეიცავენ ერთზე მეტ ელემენტს წარმოადგენენ საინტერესო ობიექტებს.

განმარტება A.2. Ω სივრცის ნებისმიერ ქვესიმრავლეს (ანუ ელემენტარულ ხდომილებათა ერთობლიობას) ეწოდება **ხდომილება**. მას ზოგადად აღნიშნავენ დიდი ლათინური ასოებით და აღნიშნავენ ფრაზას, რომელიც აღწერს ამ ხდომილებას.

რადგან ხდომილებები განმარტებულია სიმრავლურ ენაზე, ამიტომ სიმრავლეთა თეორიის ყველა მარტივი ცნება თუ ფორმულა ვრცელდება ხდომილებებზეც. მაგალითად, როგორცაა სიმრავლეთა *გაერთიანება*, *თანაკვეთა* ან *დამატება*. ალბათობათა თეორიაში იხმარება მათი შეასტყვისი ტერმინები *ან*, *და*, *არა*.

განმარტება A.3. A და B ხდომილებებისთვის, რომლებიც არიან Ω ელემენტარულ ხდომილებათა სივრცის ქვესიმრავლეები, განიმარტებიან ახალი ხდომილებები:

1. არა A (ანუ A -ს უარყოფა), არის ისეთი ხდომილება, რომელიც შედგება ყველა იმ ელემენტარული ხდომილებებისგან, რომლებიც არ ეკუთვნიან A-ს.
2. A ან B (ანუ ხდომილებათა ჯამი $A + B$), არის ისეთი ხდომილება, რომელიც შედგება ყველა იმ ელემენტარული ხდომილებებისგან, რომლებიც

მეკუთვნებიან A -ს ან B -ს, ან ორივეს ერთდროულად. ასეთ დროს ვამბობთ, „მიეკუთვნებიან ერთერთს მაინც“.

3. A და B (ანუ ხდომილებათა ნამრავლი $A \cdot B$), არის ისეთი ხდომილება, რომელიც შედგება ყველა იმ ელემენტარული ხდომილებებისგან, რომლებიც მეკუთვნებიან როგორც A -ს, ასევე B -ს, ანუ მიეკუთვნებიან ორივეს ერთდროულად.

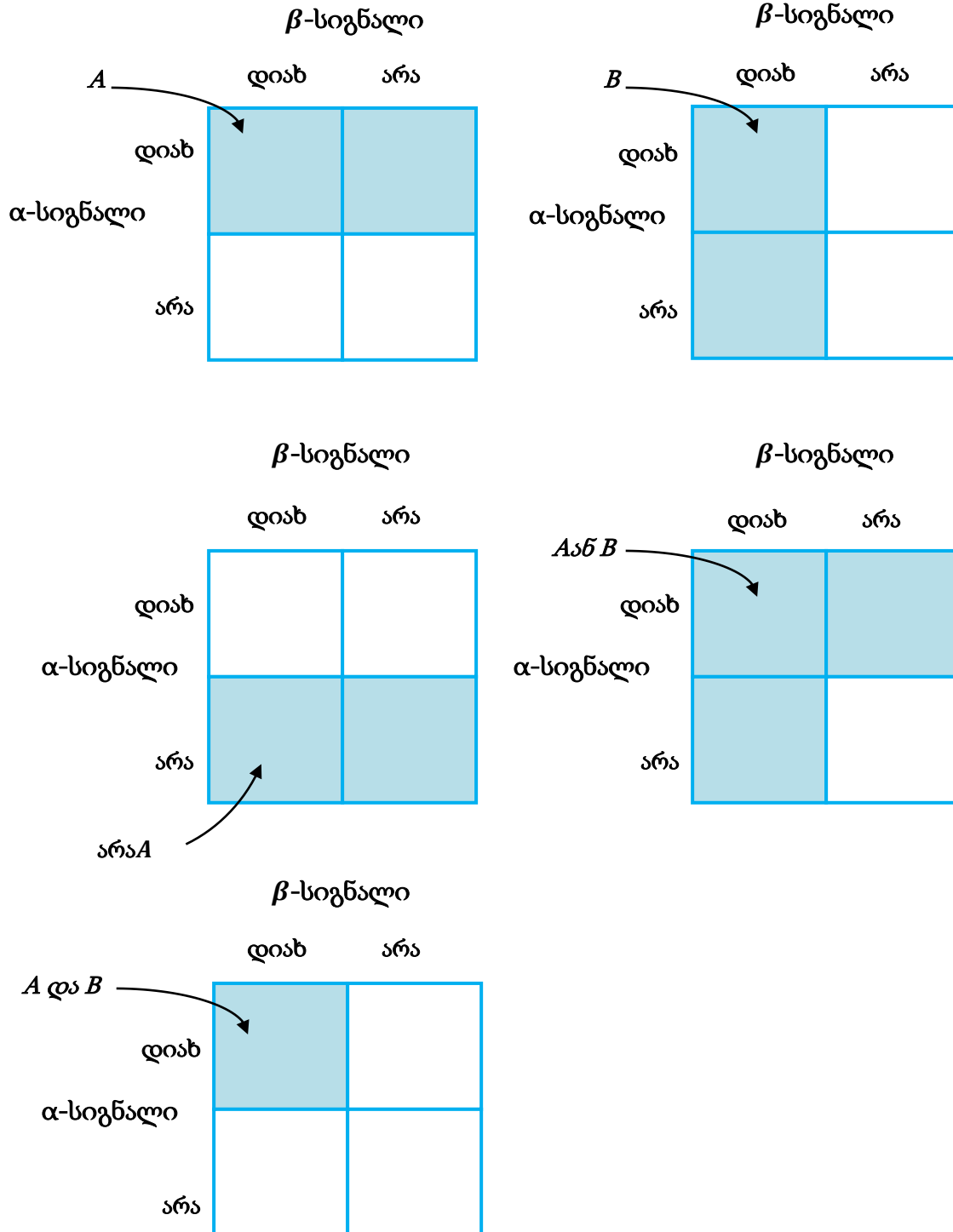
სიმრავლეთა ელემენტარული თეორია უშვებს ისეთი სიმრავლის არსებობას, რომელსაც არცერთი ელემენტი არ გააჩნია. მას *ცარიელ სიმრავლეს* უწოდებენ. ასეთივე საჭიროებაა ალბათობის თეორიაშიც.

განმარტება A.4. ცარიელი ხდომილება ანუ შეუძლებელი ხდომილება ისეთი ხდომილებაა, რომელიც არც ერთ შედეგს არ შეიცავს, ანუ ისეთი ხდომილებაა, რომლის განხორციელებაც შეუძლებელია. ისევე როგორც სიმრავლეთა თეორიაში, ალბათობაშიც მიღებულია, რომ ცარიელი ხდომილება აღინიშნოს $\emptyset$ სიმბოლოთი.

ალბათობის თეორიაში $\emptyset$ სიმბოლო ყველაზე უფრო მოსახერხებელია ხდომილებათა იმ დამოკიდებულების აღწერისთვის, როდესაც მათ არ გააჩნიათ საერთო ელემენტარული ხდომილებები. სპეციალური ტერმინოლოგიით საბოლოოდ გვქენება, რომ **A და $B = \emptyset$** .

განმარტება A.5. თუ A და B ხდომილებებს საერთო ელემენტები არ გააჩნიათ (ანუ **A და $B = \emptyset$**), მაშინ მათ ეწოდებათ **განცალკევებული**, ანუ **ურთიერთ გამომრიცხავი**, (ქართული ტერმინოლოგიით - **არათავსებადი**) ხდომილებები.

მაგალითი A.1. ვთქვათ, რაღაც მოწყობილობაში α -სიგნალის შემოსვლის ხდომილება აღნიშნულია A-თი, ხოლო β -სიგნალისა – B-თი. ქვემოთ მოყვანილ ნახაზებზე გრაფიკულადაა გამოსახული განმარტებების ინტერპრეტაციები.



ნახაზი A.1. სიტუაციების გრაფიკული ინტერპრეტაციები: A, B , არა A , A ან B , A და B .

მანიპულაციები ხდომილებებზე, როგორცაა დამატება, გაერთიანება, თანაკვეთა და ა. შ. წარმოადგენს აუცილებელ ფონს ალბათობის შესასწავლად, მაგრამ იგი ნაკლებად ასახავს ალბათობის ძირითად მიზანს. ალბათობას საფუძვლად უდევს აქსიომატიკა, რომელიც ასე შეიძლება ჩამოყალიბდეს.

განმარტება A.6. ალბათობის სისტემა ეწოდება ყოველი A ხდომილებებისთვის ისეთი $P[A]$ რიცხვის შესაბამებას, რომ სრულდებოდეს შემდეგი პირობები:

1. ყოველი A -სთვის $0 \leq P[A] \leq 1$,
2. $P[\Omega] = 1$ და $P[\emptyset] = 0$,
3. ურთიერთ გამომრიცხავი (არათავსებადი) $A_1, A_2, A_3, \dots$ ხდომილებებისთვის

$$P[A_1 \text{ ან } A_2 \text{ ან } A_3 \text{ ან } \dots] = P[A_1] + P[A_2] + P[A_3] + \dots$$

(1), (2) და (3) დამოკიდებულებები წარმოადგენენ ალბათობათა თეორიის აქსიომებს.

აქსიომა (1) გვეუბნება, რომ, მაგალითად, ფარდობითი სიხშირეების შემთხვევაში ალბათობებს მხოლოდ 0-დან 1-მდე შუალედში აქვთ აზრი. აქსიომა (2) გვეუბნება, რომ 1-ის ტოლი ალბათობის ინტერპრეტაცია ნიშნავს ხდომილების აუცილებლად მოხდენას (ანუ ეს იგივეა, რომ Ω -დან ერთი ხდომილება უეჭველად მოხდება), ხოლო ალბათობა 0 კი ნიშნავს, რომ ხდომილება არასდროს მოხდება. აქსიომა (3) ამბობს, რომ თუ ხდომილება შედგება პატარ - პატარა თანაუკვეთი ხდომილებებისგან, მაშინ ხდომილების ალბათობა ტოლია პატარა ნაწილების ხდომილებათა ჯამის. ამ აქტის ერთერთ გამოყენებას ჩვენ შევხვდით წინა თავებში, როდესაც განვიხილავდით მაგალითს, სადაც დავასკვნით, რომ პუასონის X შემთხვევითი ცვლადისთვის

$$\begin{aligned} P[2 \leq X \leq 5] &= P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5] = \\ &= f(2) + f(3) + f(4) + f(5), \end{aligned}$$

როდესაც რეალურად გამოვიყენეთ მესამე აქსიომა ხდომილებებისთვის:

$$A_1 = \{X = 2\}, \quad A_2 = \{X = 3\}, \quad A_3 = \{X = 4\}, \quad A_4 = \{X = 5\}.$$

ეს არის ძალიან მარტივი შემთხვევა. ზოგადად, ყოველთვის შეგვიძლია გამოვიყენოთ განმარტება A.6, მხოლოდ უნდა შევამოწმოთ, რომ ხდომილებები აკმაყოფილებენ ალბათობის აქსიომების პირობებს. უფრო ხშირად გვიწევს მივანიჭოთ ალბათობები (რომელთა ჯამი ტოლია 1-ის) კონკრეტულ ხდომილებებს და შემდეგ მარტივად გამოვიყენოთ მესამე აქსიომა საჭირო ალბათობების გამოსათვლელად. (ეს სტრატეგია გარანტიას გვაძლევს, რომ მომდევნო ალბათობები იქნება ლოგიკურად თანმიმდევრული).

მაგალითი A.2. ბზარები ლითონის დეტალზე. განვიხილოთ განსაკუთრებულად მარტივი სიტუაცია. ერთჯერადი შემოწმებისას დეტალებზე ბზარების აღმოჩენის ელემენტარულ ხდომილებათა სივრცე იქნება ასეთი:

$$\Omega = \{ \text{ბზარი აღმოჩნდა, ბზარი არ აღმოჩნდა} \},$$

აქ შესაძლებელია მხოლოდ ოთხი ხდომილება:

$$\Omega, \quad \{ \text{ბზარი აღმოჩნდა} \}, \quad \{ \text{ბზარი არ აღმოჩნდა} \}, \quad \emptyset.$$

მათი ალბათობები, რომლებიც შეესატყვისებიან განმარტება A.6 -ს, შეიძლება იყოს

$$P[\Omega] = 1, \quad P[\text{ბზარი აღმოჩნდა}] = .3$$

$$P[\text{ბზარი არ აღმოჩნდა}] = .7, \quad P[\emptyset] = 0.$$

ცხადია, რომ ეს მათემატიკურ თანხვედრაშია ალბათობის განმარტებასთან, მაგრამ ცალკე საკითხია შევადგინოთ ისეთი რეალური და ამასთან სასარგებლო მოდელი, რომელიც ნამდვილად გასცემს პასუხს საბაზისო ცდისეულ მტკიცებულებებს.

A. 2. უმარტივესი თეორემები ალბათობათა თეორიიდან.

როგორც ვიცით, ალბათობათა თეორიაში შესაძლო ხდომილებათა ალბათობები არ არიან ზედმიწევნით ცხადად დაწერილები. საკმაოდ ხშირად არსებობს სიტუაციები, როდესაც ერთობლივად მოცემული ხდომილებების ანალიზისა და აქსიომების დახმარებით ვითვლით სხვა ხდომილებების შესაბამის ალბათობებს და მათი საშუალებით ვადგენთ იმ ხდომილებების ალბათობებს, რომელთა ცოდნაც

წარმოადგენს ჩვენს ინტერესს. ერთერთი ასეთი მარტივი თეორემა, რომელიც ამყარებს დამოკიდებულებას $P[A]$ -სა და $P[\text{არა } A]$ -ს შორის არის შემდეგი:

თეორემა A.1. ნებისმიერი A ხდომილებისთვის

$$P[\text{არა } A] = 1 - P[A]$$

ეს ფაქტიც იყო გამოყენებული წინა თავებში ზუსტი დასკვნის გარეშე. მაგალითად, დამოუკიდებელ ცდათა შემთხვევაში წარმატება/წარუმატებლობის შედეგით, ალბათობა იმისა, რომ წარმატებას ადგილი ექნებოდა ერთხელ მაინც (ანუ, ბინომური შემთხვევითი X ცვლადისთვის $P[X \geq 1]$), არის 1-ს მინუს ნულოვანი წარმატების ალბათობა. ე.ი. $1 - P[X = 0] = 1 - f(0)$ არის ზუსტად რომ თეორემა A.1 -ის შედეგი.

მეორე მარტივი თეორემა შეეხება აქსიომა (3) -ზე დაყრდნობით განმარტება A.6-ის ცვლილებას იმ შემთხვევისთვის, როდესაც ხდომილებები არ არიან აუცილებლად განცალკევებულებები (დამოუკიდებლები). ამ თეორემას ზოგჯერ უწოდებენ **ჯამის ალბათობის თეორემას**.

თეორემა A.2. ნებისმიერი ორი A და B ხდომილებებისთვის

$$P[A \text{ ან } B] = P[A] + P[B] - P[A \text{ და } B]$$

ვინაიდან $P[A \text{ და } B]$ სათვალავში ჩაითვალა ორჯერ, ამიტომ გამოკლებამ გამოთვლების უზუსტობა გაასწორა და მივიღეთ ის, ფორმულა რომელიც გვაქვს $P[A \text{ ან } B]$ -ს გამოსათვლელად.

შევნიშნოთ, რომ თუ გვაქვს ურთიერთ გამომრიცხავი (არათავსებადი) ხდომილებები, მაშინ ზემოთ მოცემულ ფორმულაში ბოლო წევრი

$$P[A \text{ და } B] = P[\emptyset] = 0.$$

ამიტომ თეორემის ფორმულა გამარტივდება და ორი ხდომილების შემთხვევისთვის ვღებულობთ განმარტება A.6-ის მე-3 აქსიომას. ანუ თუ ხდომილებები არათავსებადია მაშინ გვექნება უბრალოდ შეკრება $P[A] + P[B]$. ეს ფორმულა პრაქტიკაში ხშირად გამოიყენება მაშინ, როდესაც ფორმულაში მონაწილე ოთხი

წევრიდან ვიცით რომელიმე სამი,; მაშინ შეგვიძლია მარტივი არითმეტიკული გამოთვლებით ვიპოვოთ მეოთხე წევრი.

მესამე მარტივი თეორემა შეეხება შემთხვევას, როდესაც ელემენტარულ ხდომილებათა სივრციდან ყოველი შედეგის დადგომა თანაბრად მოსალოდნელია, ან სხვანაირად რომ ვთქვათ, ტოლალბათურია.

თეორემა A.3. თუ Ω არის სასრულ ელემენტარულ ხდომილებათა სივრცე და ყოველი ხდომილება ტოლალბათურია, მაშინ ნებისმიერი A ხდომილების ალბათობა

$$P[A] = \frac{A \text{ სიმრავლის ელემენტების რაოდენობა}}{\Omega \text{ სივრცის ელემენტების რაოდენობა}}$$

ეს თეორემა გვიჩვენებს, რომ თუ ყოველი ხდომილება ტოლალბათურია, მაშინ ალბათობების მოძებნა ან შეფასებები მარტივი გამოთვლებითაა შესაძლებელი.

ლემმა A.1. (ტოლალბათური შემთხვევითი შერჩევა) დავუშვათ, რომ ოთხი ერთნაირი წიგნიდან ერთი არის ხარვეზიანი. განვსაზღვროთ ალბათობა იმისა, რომ შემთხვევით აღებული ორი წიგნიდან ორივე კარგი იქნება (ვირჩევთ ჯერ ერთ წიგნს, შემდეგ მეორეს).

კარგი წიგნები აღვნიშნოთ $G1$, $G2$, და $G3$ სიმბოლოებით, დეფექტური კი - D -თი, მაშინ ამ აღნიშვნებში Ω სივრცეს სიმრავლურ აღნიშვნებში ექნება ასეთი სახე:

$$\Omega = \{ (G1, G2), (G1, G3), (G1, D), (G2, G1), (G2, G3), (G2, D), \\ (G3, G1), (G3, G2), (G3, D), (D, G1), (D, G2), (D, G3) \}$$

Ω სივრცის გრაფიკული ინტერპრეტაცია მოცემულია შემდეგ ნახაზზე.

მეორედშერჩეული

	G1	G2	G3	D
G1	X			
G2		X		
G3			X	
D				X

პირველადშერჩეული

ნახაზი A.2. 12 შესაძლო შემთხვევის გრაფიკული ინტერპრეტაციები, როდესაც ვირჩევთ ოთხიდან ორს.

როგორც ვხედავთ სივრცე შედგება 12 ელემენტისგან, რომლებიც არიან ტოლალბათურნი და მხოლოდ 6 მათგანი არ შეიცავს D-ს არც პირველ და არც მეორე ადგილას. მესამე თეორემა გვთავაზობს ასეთ გამოსავალს,

$$P[\text{ორი კარგი წიგნი}] = \frac{6}{12} = .50$$

A. 3. პირობითი ალბათობა და ხდომილებათა დამოუკიდებლობა

გამოვიყენოთ დამოუკიდებლობა-დამოკიდებულების ადრე გამოყენებული აღნიშვნები. ამასთან ერთად, შევნიშნოთ ისიც, რომ ერთი ხდომილების მოხდენის ალბათობა დამოკიდებულია მეორე ხდომილების მოხდენა ან არ მოხდენაზე. ეს იდეა სიტყვიერად ასე ჩამოყალიბდება.

განმარტება A.7. ვთქვათ, მოცემული გვაქვს ორი A და B ხდომილება. ამავე დროს ვიგულისხმობთ, რომ B ხდომილების ალბათობა არ უდრის 0-ს. მაშინ **A ხდომილების**

პირობითი ალბათობა B პირობით (ანუ იმ პირობით, რომ B უკვე მოხდა) განიმარტება ასე:

$$P[A | B] = \frac{P[A \text{ და } B]}{P[B]}$$

სცენარი A.1. (გაგრძელება) თეორემაში მოყვანილი ფორმულის საფუძველზე გამოვთვალოთ შემდეგი პირობითი ალბათობა:

$P[\text{მეორე შერჩეული წიგნი დეფექტურია} \mid \text{პირველი შერჩეული წიგნი კარგია}]$.

მარტივი გამოთვლებით 12 შედეგიან ელემენტარულ ხდომილებათა სივრცისთვის გვექნება, რომ

$$P[\text{პირველი შერჩეული წიგნი კარგია}] = \frac{9}{12} = .75$$

$$P[\text{პირველი შერჩეული წიგნი კარგია და მეორე დეფექტურია}] = \frac{3}{12} = .25$$

მაშინ, განმარტება A.7-ის გამოყენებით ვღებულობთ, რომ

$P[\text{მეორე შერჩეული წიგნი დეფექტურია} \mid \text{პირველი შერჩეული წიგნი კარგია}]$

$$= \frac{\frac{3}{12}}{\frac{9}{12}} = \frac{1}{3}$$

ზოგჯერ ვხვდებით ისეთ გარემოებას, როდესაც თვალნათელია როგორ გამოვთვალოთ პირობითი ალბათობა. ამ მაგალითშიც ასეთი სიტუაციაა.

$$P[\text{მეორე შერჩეული წიგნი დეფექტურია} \mid \text{პირველი შერჩეული წიგნი კარგია}] = \frac{1}{3}$$

იმიტომ, რომ თუ პირველი შერჩევა კარგია, მაშინ მეორის შესარჩევად უკვე რჩება სამი შემთხვევა, რომელთაგან ერთია დეფექტური.

განმარტება A.7-ში მოყვანილი ფორმულიდან განვსაზღვრავთ $P[A \text{ და } B]$ -ს, მივიღებთ ნამრავლის ალბათობის ფორმულას.

თეორემა A.4. ვიგულისხმობთ, რომ $P[B] > 0$, ანუ $P[A|B]$ განსაზღვრულია, მაშინ

$$P[A \text{ და } B] = P[B] \cdot P[A|B]$$

როგორც წესი, $P[A|B]$ და $P[A]$ მნიშვნელობები ზოგადად განსხვავდებიან. ამ განსხვავებას იწვევს B ხდომილების მოხდენის ან არ მოხდენის ფაქტი. ზოგჯერ შეიძლება მოხდეს რომ $P[A|B]$ და $P[A]$ მნიშვნელობები იყვნენ ერთმანეთის ტოლი. ამ დროს საჭირო ხდება ახალი ტერმინის, *დამოკიდებულობის* ცნების შემოტანა.

განმარტება A.8. A და B ხდომილებებს ეწოდებათ *დამოუკიდებლები* თუ

$$P[A|B] = P[A] .$$

სხვა შემთხვევაში (ანუ, როცა $P[A|B] \neq P[A]$), ეწოდებათ *დამოკიდებულები*.

A. 4. კომბინატორიკის ელემენტები

გამრავლების პრინციპი. ვიგულისხმობთ, რომ რაღაც კომპლექსური ქმედება შედგება r რაოდენობის ქმედებებისგან. ამასთან პირველ კომონენტს შუძლია შეიცვალოს n_1 სხვადასხვა გზით, ანუ მიიღოს n_1 სხვადასხვა მნიშვნელობა. შესაბამისად მეორეს შეუძლია მიიღოს n_2 სხვადასხვა მნიშვნელობა, მესამეს — n_3 და ა.შ. მაშინ საერთო რაოდენობა, რომელიც შეიძლება რომ მიიღოს კომპლექსურმა ქმედებამ ტოლია:

$$n = n_1 \cdot n_2 \cdots n_r$$

გრაფიკულად რომ წარმოვიდგინოთ ხის დიაგრამაზე, გვაქვს n_1 პირველი დონის კვანძი, n_2 ცალი მეორე დონის კვანძი და ა.შ. მთლიანობაში იქნება $n_1 \cdot n_2 \cdots n_r$ ცალი r დონიანი კვანძი.

მაგალითად, დაუშვათ, რომ სამი სხვადასხვა A, B, C ქმედება შესაძლებელია განხორციელდეს რამდენიმე გზით. A შეიძლება $n_1 = 3$ გზით, B — $n_2 = 4$ და C — $n_3 = 2$. ამ დროს $r = 3$. მაშინ, გვექნება

$$n = n_1 \cdot n_2 \cdot n_3 = 3 \cdot 4 \cdot 2 = 24$$

სხვადასხვა კომბინაციას.

განმარტება A.9. თუ n განსხვავებული ობიექტების ერთობლიობიდან შერჩეული გვაქვს r ცალი ობიექტი და მათ ვალაგებთ მიმდევრობით 1-დან r -მდე (ანუ ისინი იკავებენ r განსხვავებულ პოზიციას), მაშინ ყოველ ასეთ პოტენციურ წყობას ეწოდება გადანაცვლება. მას აღნიშნავენ $P_{n,r}$ სიმბოლოთი და იკითხება „ n ობიექტის გადანაცვლებათა რაოდენობა r -ად“. მისი გამოსათვლელი ფორმულაა

$$P_{n,r} = n(n-1)(n-2) \cdots (n-r+1),$$

ანუ

$$P_{n,r} = \frac{n!}{(n-r)!}$$

შენიშვნა: ქართულ ენაზე არსებულ ლიტერატურაში ამ სიდიდეს უწოდებენ წყობას, ხოლო იმ კერძო შემთხვევას, როცა $r = n$ ჰქვია გადანაცვლება და უდრის $n!$ -ს,

$$P_n = P_{n,n} = n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1,$$

მეორე ზოგადი გამოთვლითი ამოცანა დაკავშირებულია გადანაცვლების იდეასთან და პრაქტიკულად შესაფერისია იმ პრობლემასთან, როდესაც ხდება შემთხვევითი შერჩევების აღწერა. ეს არის ამოცანა, როდესაც ხდება n განსხვავებული საგნების

(ობიექტების) ერთობლიობიდან r არადალაგებული საგნების შერჩევა. მისი ტერმინოლოგია და აღნიშვნები არის შემდეგი:

განმარტება A.10. თუ n განსხვავებული ობიექტების ერთობლიობიდან შერჩეული გვაქვს r ცალი არადალაგებული ობიექტი, მაშინ ყოველ შესაძლო ერთობლიობას ეწოდება **კომბინაცია**. ამ კომბინაციათა ყველა შესაძლო რაოდენობა აღინიშნება $\binom{n}{r}$ -ით და ეწოდება „კომბინაციათა რაოდენობა n -დან r -ად“.

შევნიშნოთ, რომ ამ ტერმინოლოგიამ შეიძლება გამოიწვიოს მსუბუქი კონფლიქტი „კომბინაციის“ სხვა კონტექსტებში ხმარების დროს, რაც ძალიან იშვიათია და მაშინაც იოლად ხდება დაკონკრეტება.

(ქართულ ტერმინოლოგიაში მას *ჯუფთებას* უწოდებენ).

ადვილი დასანახია, რომ გადანაცვლების გამოთვლის პროცესი შედგება ორი საფეხურისგან. პირველი, n განსხვავებული საგნიდან ვირჩევთ r საგანს ანუ ვითვლით კომბინაციას და მეორე, შემდეგ ამორჩეულ საგნებში ვითვლით ადგილების მიხედვით დალაგებათა რაოდენობას. ყოველივე ეს შეიძლება ასე ჩაიწეროს:

$$P_{n,r} = \binom{n}{r} \cdot P_{r,r},$$

რაც ნიშნავს შემდეგს:

$$\frac{n!}{(n-r)!} = \binom{n}{r} \cdot \frac{r!}{0!}, \quad (\text{განმარტებით } 0! = 1)$$

საიდანაც მივიღებთ, *კომბინაციათა რაოდენობის გამოსათვლელ ფორმულას*:

$$\binom{n}{r} = \frac{n!}{n!(n-r)!}.$$

ამ ფორმულაში მიღებული შეფარდება, როგორც გვახსოვს, გამოყენებული იყო $p^x(1-p)^{n-x}$ სახით ბინომური ალბათობის ფუნქციაში, როდესაც ვითვლიდით x წარმატებათა რაოდენობას n ცდათა მიმდევრობაში.

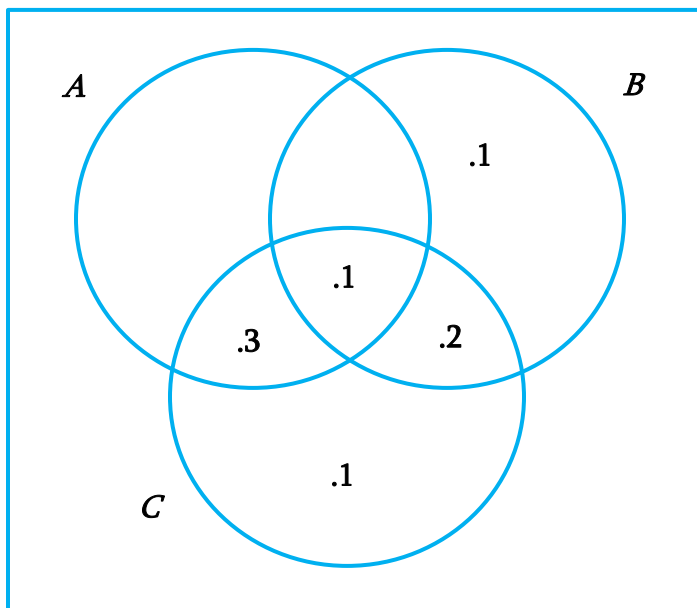
სავარჯიშოები:

1. ერთ ყუთში შერეულად ერთად ყრია ქანჩები. ქანჩების 30% არის $\frac{1}{2}$ დუიმიანი (in.), ხოლო 70% არის $\frac{9}{16}$ (in.). მეორე ყუთში აგრეთვე შერეულად ერთად ყრია ჭანჭიკები. მათი 40% არის $\frac{1}{2}$ (in.) და 60% არის $\frac{9}{16}$ (in.). ვიგულისხმობთ, რომ თითოეული ყუთიდან დამოუკიდებლად და შემთხვევით შევარჩიეთ თითო ქანჩი და ჭანჭიკი.

(ა) რა არის ალბათობა იმისა, რომ ეს ქანჩი და ჭანჭიკი ერთი და იმავე ზომის იქნება (ანუ ერთმანეთს მოერგებიან) ?

(ბ) პირობითი ალბათობა იმისა, რომ ქანჩი $\frac{9}{16}$ (in.) მოცემულია და ამოღებული ჭანჭიკი მას მოერგება?

2. შემდეგი ნახაზი წარმოადგენს ვენის დიაგრამას, რომელზედაც მონიშნულია მათი შესაბამისი ალბათობები. დამატებით იმისგან, რაც მონიშნულია გვაქვს, ვიგულისხმობთ, რომ $P[B] = .4$ და $P[C | A] = .8$.



(ა) დაამთავრეთ დიაგრამის შევსება. გამოთვალეთ შემდეგი სამი ალბათობა:

$P[A \text{ და } B \text{ და არა } C]$, $P[A \text{ და არა } B \text{ და არა } C]$ და

$P[\text{არა } (A \cap B \cap C)] = P[\text{არა } A \text{ და არა } B \text{ და არა } C]$.

- (ბ) გამოიყენეთ ალბათობები, რომლებიც დიაგრამაზეა მოცემული (აგრეთვე თქვენი პასუხები, რომლებიც მიიღეთ (ა)-ში) და გამოთვალეთ $P[A \text{ და } B]$.
- (გ) გამოიყენეთ დიაგრამის ალბათობები და გამოთვალეთ $P[B | C]$.
- (დ) აქ მოყვანილ ინფორმაციაზე დაყრდნობით უპასუხეთ, არიან თუ არა B და C ხდომილებები? პასუხი ახსენით.
3. გვაქვს მანქანის ერთი და იმავე ნაწილების 100 ელემენტური სიმრავლე. აქედან 10 არის არასტანდარტული დიამეტრის მცირე ზომის გამო და 20 არის არასტანდარტული დიამეტრის დიდი ზომის გამო, ხოლო 70 სტანდარტულია.
- (ა) რამდენი სხვადასხვა შესაძლო $n = 10$ ზომის შერჩევა არსებობს მთელი ამ სიმრავლიდან?
- (ბ) რამდენი სხვადასხვა შესაძლო $n = 10$ ზომის შერჩევა არსებობს ისეთი, რომელიც შეიცავს ზუსტად ერთ დიამეტრის მცირე ზომიან ნაწილს, ზუსტად ორ დიამეტრის დიდ ზომიან ნაწილს და 7 სტანდარტულ დიამეტრიან ნაწილს.
- (გ) წინა (ა) და (ბ) კითხვებზე დაყრდნობით უპასუხეთ, რა არის ალბათობა იმისა, რომ შემთხვევით აღებული $n = 10$ ზომის შერჩევა შეიცავს ზუსტად ერთ დიამეტრის მცირე ზომიან ნაწილს, ზუსტად ორ დიამეტრის დიდ ზომიან ნაწილს და 7 სტანდარტულ დიამეტრიან ნაწილს?
4. სტუდენტმა გადაწყვიტა გამოიყენოს კალკულატორის შემთხვევითი ციფრული ფუნქცია და მისი დახმარებით შეარჩიოს სამნიშნა PIN კოდი. (ვიგულისხმობთ, რომ ყველა რიცხვი 000 -დან 999 -მდე ტოლ შესაძლებელია).
- (ა) რა არის ალბათობა იმისა, რომ შერჩეული იქნება რიცხვი, რომელიც შედგება მხოლოდ კენტი ციფრებისაგან?
- (ბ) რა არის ალბათობა იმისა, რომ სამივე შერჩეული ციფრი განსხვავებულია?
- (გ) რა არის ალბათობა იმისა, რომ შერჩეული რიცხვები განსხვავებულია ერთმანეთისგან და ციფრები დალაგებულია ზრდადობის ან კლებადობის რიგით?
5. მოცემული გვაქვს სიმრავლე, რომელიც შედგება ათი ერთნაირი ბარომეტრისგან, რომელთაგან სამი არაზუსტია. ტექნიკოსები ამოწმებენ მათ ვარგისიანობას. ისინი შემთხვევით თითო-თითოდ იღებენ ბარომეტრებს,

ამოწმებენ მათ და შემოწმებულს უკან აღარ აბრუნებენ.

(ა) რა არის ალბათობა იმისა, რომ შერჩეული პირველი ოთხი ბარომეტრიდან ზუსტად ერთი იქნება არაზუსტი?

(ბ) რა არის ალბათობა იმისა, რომ ტექნიკოსები აღმოაჩენენ მეორე არაზუსტ ხელსაწყოს მეხუთე გასიჯვაზე?

B. ცხრილები

B.1. სტანდარტული ნორმალური დაგროვებითი ალბათობები

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$$

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455

-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0594	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

B.2. სტანდარტული ნორმალური კვანტილები

	.00	.01	.12	.03	.04	.05	.06	.07	.08	.09
.0		-2.33	-2.05	-1.88	-1.75	-1.65	-1.55	-1.48	-1.41	-1.34
.1	-1.28	-1.23	-1.18	-1.13	-1.08	-1.04	-.99	-.95	-.92	-.88
.2	-.84	-.81	-.77	-.74	-.71	-.67	-.64	-.61	-.58	-.55
.3	-.52	-.50	-.47	-.44	-.41	-.39	-.36	-.33	-.31	-.28
.4	-.25	-.23	-.20	-.18	-.15	-.13	-.10	-.08	-.05	-.03
.5	0.00	.03	.05	.08	.10	.13	.15	.18	.20	.23
.6	.25	.28	.31	.33	.36	.39	.41	.44	.47	.50
.7	.52	.55	.58	.61	.64	.67	.71	.74	.77	.81
.8	.84	.88	.92	.95	.99	1.04	1.08	1.13	1.18	1.23

.9	1.28	1.34	1.41	1.48	1.55	1.65	1.75	1.88	2.05	2.33
----	------	------	------	------	------	------	------	------	------	------

სარჩევი

თავი 1 შესავალი

1.3 საინჟინრო სტატისტიკა: რა და რატომ

1.4 საბაზისო ტერმინოლოგია

სავარჯიშოები

თავი 2 ელემენტარული აღწერითი სტატისტიკა

2.1. რაოდენობრივი მონაცემების ელემენტარული გრაფიკული და ცხრილური დამუშავება

2.2. სიხშირეთა ცხრილები და ჰისტოგრამები

2.3. კვანტილები და მათთან დაკავშირებული გრაფიკები

2.4. სტანდარტული რიცხვითი მახასიათებლები

2.5 დამოკიდებულებები ცვლადებს შორის

სავარჯიშოები

თავი 3 დისკრეტული შემთხვევითი სიდიდე

3.1. შემთხვევითი სიდიდეები და მათი განაწილებები

3.2. დისკრეტული ალბათობის ფუნქცია და დაგროვებითი ალბათობის ფუნქცია

3.3. დისკრეტული ალბათობის განაწილების შეჯამება

3.4. ბინომური და გეომეტრიული განაწილებები

3.5. პუასონის განაწილება.

სავარჯიშოები

თავი 4 უწყვეტი შემთხვევითი სიდიდე

4.1. ალბათობის სიმკვრივის ფუნქციები და დაგროვებითი ალბათობის ფუნქციები

4.2. უწყვეტი განაწილების საშუალო და დისპერსია

4.3. ალბათობის ნორმალური განაწილება

4.4. მაჩვენებლიანი განაწილება

4.5. ვეიბულის განაწილება

სავარჯიშოები

თავი 5 ერთობლივი განაწილებები და დამოუკიდებლობა

5.1. დისკრეტულ შემთხვევით სიდიდეთა ერთობლობა

5.2. პირობითი ალბათობა და დამოუკიდებლობა დისკრეტული შემთხვევითი სიდიდეებისთვის

5.3. მრავალი ცვლადის ფუნქციები

5.4. საშუალო და დისპერსია შემთხვევით სიდიდეთა წრფივი კომბინაციებისთვის

5.5. ცენტრალური ზღვარითი ეფექტი

სავარჯიშოები

თავი 6 ფორმალური სტატისტიკური დასკვნები

- 6.1. დიდი შერჩევის ნდობის ინტერვალი საშუალოსთვის
- 6.2. ზოგადად გამოყენებადი ნდობის ინტერვალი μ -სთვის
- 6.3. რამდენიმე დამატებითი კომენტარი ნდობის ინტერვალთან დაკავშირებით
- 6.4. დიდი შერჩევის მნიშვნელობის ტესტი საშუალოსთვის

სავარჯიშოები

დანართი

- B. 1. ელემენტარული ალბათობის ძირითადი ცნებები და აქსიომები
- B. 2. უმარტივესი თეორემები ალბათობათა თეორიიდან.
- B. 3. პირობითი ალბათობა და ხდომილებათა დამოუკიდებლობა
- C. 4. კომბინატორიკის ელემენტები

სავარჯიშოები

- D. ცხრილები

