მონაცემთა ანალიზი და სტატისტიკა
(ლექციათა კურსი – 2013წ)

(პროფ. გია სირბილაძე)

ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი
ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი
კომპიუტერულ მეცნიერებათა დეპარტამენტი

ავტორის წინასიტყვაობა

მონაცემთა ანალიზი და სტატისტიკა შეისწავლის როგორც სტატისტიკურ მონაცემთა სისტემატიზაციის, ანალიზისა და გამოყენების მეთოდებს თეორიული თუ პრაქტიკული დასკვნების მიღების მიზნით, ასევე სტატისტიკური პროგრამული უზრუნველყოფის გამოყენების შესაძლებლობებსა და პერსპექტივებს. მას ხშირად **სტატისტიკურ გადაწყვეტილებათა თეორიასაც** უწოდებენ. როგორც **დისკრეტულ ალბათობათა ანალიზი**, რომელსაც თქვენ *საგნის სწავლების პირველ ლექციებზე გაეცანით*, ასევე *სტატისტიკის საფუძვლები* შეისწავლის მასობრივი მოვლენების შემთხვევით ხდომილებათა თუ პარამეტრთა რაოდენობრივი ხასიათის კანონზომიერებებს. *ალბათობათა ანალიზში*, ვიცით რა რაიმე მოვლენის ბუნება და მისი საწყისი ალბათური განაწილება, ვიკვლევთ ამა თუ იმ მახასიათებლის განაწილებას და მის პარამეტრებს, რომელიც დაკავშირებულია მასობრივ მოვლენასთან. *სტატისტიკაში ყველაფერი პირიქითაა:* საწყისად ითვლება **ექსპერიმენტული მონაცემები**. როგორც წესი ეს *შემთხვევით პოპულაციებზე* დაკვირვებით მიღებული მონაცემებია, ხოლო მოითხოვება გავაკეთოდ დასკვნები მასობრივი მოვლენის ევოლუციაზე თუ ქცევაზე (მისი განაწილების თერმინებში) ან შევაფასოთ პოპულაციათა პარამეტრები და ამით *შევიქმნათ სარწმუნო წარმოდგენები მასობრივი მოვლენის შესახებ.* დიდი რაოდენობა მონაცემების დამუშავება, ანალიზი და მისი მიზნობრივი გამოყენება **თანამედროვე ინფორმატიკის** პრიორიტეტული მიმართულებაა. ცხოვრების დღევანდელ რთულ გარემოში ადამიანი, მისი ინტელექტი თუ ცოდნა უძლურია თავი მოუყაროს ერთად ყველა საჭირო ინფორმაციას და მასზე დაყრდნობით **მიიღოს ოპტიმალური გადაწყვეტილება**. მიზეზი მრავალია: ესაა მონაცემთა **დიდი რაოდენობა მისი განუზღვრელი და წინააღმდეგობრივი, ხშირად კი არაზუსტი** წარმოდგენებით. ყოველივე ეს ეხება მრავალ გამოყენებით დარგს: ეკონომიკას და ბიზნესს, ექსპერიმენტულ ფიზიკას, ქიმიას და ბიოლოგიას, სოციოლოგიას და ფსიქოლოგიას, ლინგვისტიკას და სხვა. **ინფორმატიკოსს**, რომელიც შეიარაღებულია გამოყენებითი სტატისტიკის მეთოდების ცოდნითა და მისი პროგრამული უზრუნველყოფით, შეუძლია დაეხმაროს **ყველას** მათი სტატისტიკური მონაცემთა **დამუშავების, ანლიზისა და სინთეზის** ამოცანებში. ამისათვის საჭიროა შევისწავლოთ სტატისტიკურ მონაცემთა **დამუშავების მეთოდები**, მასობრივი მოვლენის აღმწერი სიდიდეების **შეფასების პარამეტრული მეთოდები,** ასევე **ინტერვალური შეფასების მეთოდები**, და ბოლოს **სტატისტიკური გამოცდათა და სტატისტიკური დასკვნების** თეორიის ელემენტები. მნიშვნელოვანია **რეგრესიული ანალიზის** ელემენტების ცოდნაც.

პარალელურად ყოველივე ამისა უნდა გავეცნოთ **EXCEL** – ის იმ პროცედურა-ფუნქციებს, რომლებიც უზრუნველყოფენ ზემოთ ჩამოთვლილი ამოცანების კონკრეტულ რეალიზაციას.

ლექციათა კურსი განკუთვნილია **Computing**-ის (ზოგადად ინფორმატიკა) ყველა სპეციალობების (მათ შორის "კომპიუტერული მეცნიერება") სწავლების სტუდენტებისათვის. ლექციათა კურსი შეესაბამება მსოფლიოში აღიარებული კომპიუტერული საზოგადოებების- **ACM –სა და IEEE**-ის **Computing**-ის მეთოდოლოგიას.

ლექციათა კურსი შედგება თხუთმეტი ლექციისგან და დანართისგან. ყოველ ლექციაში სალექციო მასალასთან ერთად წარმოდგენილია სავარჯიშო ამოცანები,

EXCEL – ის სტატისტიკური ფუნქციები და კითხვარი თვითშემოწმებისთვის. დანართი შედგება სტატისტიკური ცხრილებისგან.

ლექცია I უწყვეტი შემთხვევითი სიდიდეები

§ 1. განაწილების ფუნქცია და განაწილების სიმკვრივე

ვთქვათ, X არის შემთხვევითი სიდიდე. x ნამდვილი ცვლადის

$$F_X(x) = P\{X \le x\} \tag{1.1}$$

ფუნქციას ეწოდება X შემთხვევითი სიდიდის განაწილების ფუნქცია.

დისკრეტული X შემთხვევითი სიდიდე იღებს $x_1<x_2<\cdots<x_r$ მნიშვნელობებს, შესაბამისად $p_1,p_2,\ldots, p_r$ ალბათობებით, მაშინ

$$F_X(x_k) = P\{X \le x_k\} = p_1 + p_2 + \cdots + p_k = \sum_{i=1}^{k} p_i ,$$

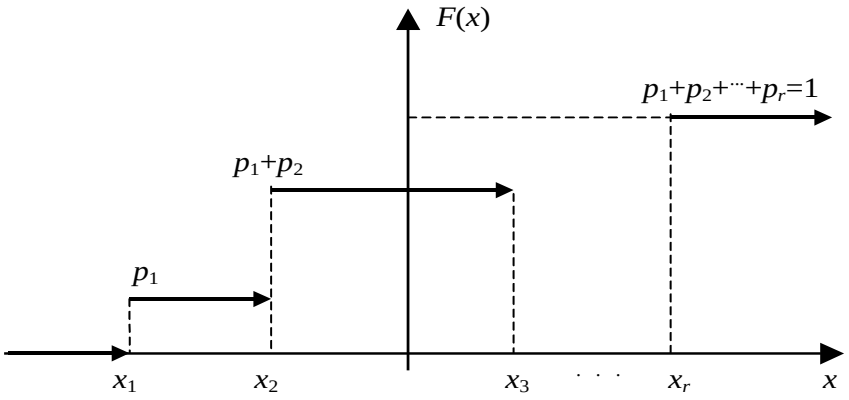
ხოლო ყოველი x-ისათვის $[x_k, x_{k+1})$ ინტერვალიდან

$$F_X(x) = P\{X \le x\} = F(x_k) .$$

ამრიგად, $F_X(x)$ წარმოადგენს არაკლებად ფუნქციას და ის ასე ჩაიწერება:

$$F_X(x) = \begin{cases} 0, & \text{როცა} \quad x < x_1, \\ p_1, & \text{როცა} \quad x_1 \le x < x_2, \\ p_1 + p_2, & \text{როცა} \quad x_2 \le x < x_3, \\ \cdots & \cdots \quad \cdots \\ \sum_{i=1}^{k} p_i, & \text{როცა} \quad x_k \le x < x_{k+1}, \\ \cdots & \cdots \quad \cdots \\ 1, & \text{როცა} \quad x_r \le x \end{cases} \tag{1.2}$$

ამ ფუნქციის გრაფიკს აქვს შემდეგი სახე:



ნახ. 3.2.

ნებისმიერი დისკრეტული შემთხვევითი სიდიდის განაწილების ფუნქცია არაკლებადი, უბან-უბან მუდმივი, მარჯვნიდან უწყვეტი ფუნქციაა. ამასთან $F(x)$-ის წყვეტის წერტილებია x_k, $k=1,2,\ldots$, და ნახტომის სიდიდე x_k წერტილში $p_k=P\{X=x_k\}$ ალბათობის ტოლია.

(1.2)-იდან ცხადია, რომ ადგილი აქვს ტოლობას:

$$p_k = F_X(x_k) - F_X(x_{k-1}), \tag{1.3}$$

იგულისხმება, რომ $F_X(x_0)=0$.

ნებისმიერი X შემთხვევითი სიდიდის განაწილების ფუნქციას აქვს შემდეგი თვისებები

1. მონოტონურობის თვისება: თუ $x_1<x_2$, მაშინ $F_X(x_1)\le F_X(x_2)$;
2. $F_X(-\infty)=\lim\limits_{x\to-\infty} F_X(x)=0$ და $F_X(\infty)=\lim\limits_{x\to\infty} F_X(x)=1$;
3. განაწილების ფუნქცია მარჯვნიდან უწყვეტია.

თუ მოცემულია X დისკრეტული შემთხვევითი სიდიდის განაწილების ფუნქცია $F_X(x)$, მაშინ ადვილია მისი *განაწილების კანონის აღდგენა*: $F_X(x)$ განაწილების ფუნქციის წყვეტის წერტილები წარმოადგენენ X შემთხვევითი სიდიდის შესაძლო მნიშვნელობებს, ხოლო შესაბამისი ალბათობები გამოითვლება (1.3) ფორმულით.

ვიტყვით, რომ X არის უწყვეტი შემთხვევითი სიდიდე, თუ არსებობს მთელ ღერძზე განსაზღვრული ისეთი არაუარყოფითი $f(x)$ ფუნქცია, რომ X შემთხვევითი სიდიდის განაწილების $F_X(x) = P\{X\le x\}$ ფუნქციისათვის

$$F_X(x)=\int_{-\infty}^{x} f(y)dy. \qquad (1.4)$$

$f(x)$ ფუნქციას X შემთხვევითი სიდიდის განაწილების სიმკვრივე ეწოდება.

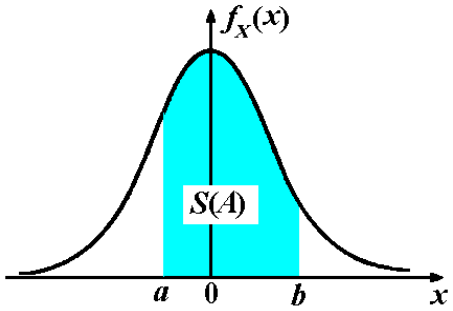
ჩვენ ვგულისხმობთ, რომ $f(x)$ ფუნქცია უწვეტია მთელ ღერძზე გარდა, შესაძლებელია, წერტილების სასრული რაოდენობისა. (1.4)-დან გამომდინარე განაწილების სიმკვრივის თვისებებია

1. $f(x)\ge 0$, $-\infty<x<\infty$,
2. $\int\limits_{=\infty}^{\infty} f(x)dx=1$,
3. $F_X'(x)=f(x)$, $f(x)$ ფუნქციის უწყვეტობის წერტილებში.
4. თუ a და b სასრული რიცხვებია

$$P\{a\le X\le b\}=P\{a\le X<b\}=P\{a<X\le b\}=P\{a<X<b\}=\int_a^b f(x)dx\equiv F(b)-F(a). \qquad (1.5)$$

უკანასკნელი ტოლობა ხაზს უსვამს იმ გარემოებას, რომ ნებისმიერი x-ისათვის $P\{X=x\}=0$.

ამრიგად, ალბათობა იმისა, რომ X შემთხვევითი სიდიდე მოთავსდება რაიმე ინტერვალში სიმკვრივის ფუნქციის გრაფიკსა და Ox ღერძს შორის მოთავსებული იმ ფიგურის $S(A)$ ფართობის ტოლია, რომელიც $x=a$ და $x=b$ წერტილებში აღმართულ მართობებს შორისაა მოქცეული (იხ.ნახ. 1.1).



$$P\{a \le X \le b\} = S(a,b) = S(A)$$

ნახ.1.1

შევადაროთ (1.2) და (1.3) ფორმულები, რომელთაც ადგილი აქვს დისკრეტული შემთხვევითი სიდიდისათვის, შესაბამისად (1.4) და (1.5) ფორმულებს. ადვილი შესამჩნევია ამ ფორმულების მსგავსება, განსხვავება კი იმაში მდგომარეობს, რომ შეკრების ოპერაცია შეიცვალა სიმკვრივიდან ინტეგრალის აღების ოპერაციით, რაც, როგორც ვიცით, შეკრების ოპერაციის უწვეტი ანალოგია. აქვე აღვნიშნოთ, რომ უსასრულოდ მცირე Δx-ისათვის ადგილი აქვს შემდეგ მიახლოებით ტოლობას

$$P\{x \le X \le x+\Delta x\} \approx f_X(x)\Delta x \tag{1.6}$$

მაგალითი 1.1. მოცემულია ფუნქცია

$$f(x) = \begin{cases} 0, & \text{როცა} \quad x < 0, \\ ax^2, & \text{როცა} \quad 0 \le x < 3, \\ 0, & \text{როცა} \quad x \ge 3. \end{cases}$$

ა) განსაზღვრეთ a პარამეტრის ის მნიშვნელობა, რომლისთვისაც $f(x)$ განაწილების სიმკვრივეა;

ბ) იპოვეთ შესაბამისი $F_X(x)$ განაწილების ფუნქცია;

გ) იპოვეთ ალბათობა იმისა, რომ $F_X(x)$ განაწილების ფუნქციის მქონე შემთხვევითი სიდიდე მოხვდება (2, 4] შუალედში.

ამოხსნა. ვინაიდან $f(x) \ge 0$, როცა $a \ge 0$, a პარამეტრის მნიშვნელობა გამოითვლება სიმკვრივის იმ თვისებიდან, რომ

$$\int_{-\infty}^{\infty} f(u)du = 1.$$

ჩავატაროთ გამოთვლები. გვაქვს

$$1 = \int_{-\infty}^{\infty} f(u)du = \int_{-\infty}^{0} f(u)du + \int_{0}^{3} f(u)du + \int_{3}^{\infty} f(u)du = \int_{0}^{3} ax^2 dx = a\frac{x^3}{3}\Big|_0^3 = 9a.$$

ამრიგად, მივიღეთ, რომ $a=1/9$.

განაწილების ფუნქციის გამოსათვლელად ვისარგებლოთ ფორმულით $F_X(x) = \int_{-\infty}^{x} f_X(u)du$,

$$F_X(x) = \begin{cases} 0, & \text{როცა} \quad x < 0, \\ \dfrac{x^3}{27}, & \text{როცა} \quad 0 \le x < 3, \\ 1, & \text{როცა} \quad x \ge 3. \end{cases}$$

მესამე შეკითხვაზე პასუხის მისაღებად ვისარგებლოთ (1.5) ფორმულით. მივიღებთ

$$P\{2<X<4\} = F_X(4) - F_X(2) = 19/27. \qquad \blacklozenge$$

§ 2. უწყვეტი შემთხვევითი სიდიდის მათემატიკური ლოდინი, დისპერსია

უწყვეტი ტიპის X შემთხვევითი სიდიდის მათემატიკური ლოდინი ეწოდება

$$EX = \int_{-\infty}^{+\infty} xf(x)dx \tag{1.7}$$

რიცხვს, სადაც $f_X(x)$ არის X შემთხვევითი სიდიდის განაწილების სიმკვრივე, თუ

ცნობილია, რომ $\int_{-\infty}^{+\infty} | u | f(u)\, du < \infty$.

უწყვეტი განაწილების შემთხვევაში მათემატიკური ლოდინის შემდეგი თვისებები ინტეგრალის თვისებებიდან გამომდინარეობს.

1. $Ec=c$, ანუ მუდმივის მათემატიკური ლოდინი ამ მუდმივის ტოლია.
2. $E(cX)=cEX$, ანუ მუდმივი გადის მათემატიკური ლოდინის ნიშნის გარეთ;
3. $E(c_1X_1+c_2X_2)=c_1EX_1+c_2EX_2$, ანუ წრფივი კომბინაციის მათემატიკური ლოდინი ლოდინების წრფივი კომბინაციის ტოლია;
4. თუ X და Y შემთხვევითი სიდიდეები დამოუკიდებელია, მაშინ $EXY=EX\cdot EY$.

შემთხვევითი სიდიდის მეორე მნიშვნელოვანი მახასიათებელი შემთხვევითი სიდიდის დისპერსიაა

$$DX=E(X-EX)^2=EX^2-(EX)^2.$$

სადაც EX^2 - X შემთხვევითი სიდიდის მეორე რიგის მომენტია. ამასთან უწყვეტი შემხვევითი სიდიდისათვის

$$DX = \int_{-\infty}^{\infty} (x - EX)^2 f(x)dx,$$

ხოლო

$$EX^2 = \int_{-\infty}^{\infty} x^2 f(x)dx.$$

დისპერსია აკმაყოფილებს შემდეგ თვისებებს:

1) თუ c მუდმივია, მაშინ $Dc=0$.
2) $D(aX+b)=a^2DX$, ე.ი. შემთხვევით სიდიდეზე მუდმივის დამატება მის დისპერსიას არ ცვლის, ხოლო მუდმივი მამრავლი კვადრატული ხარისხით გადის დისპერსიის ნიშნის გარეთ.
3) თუ X და Y შემთხვევითი სიდიდეები დამოუკიდებელია, მაშინ $D(X+Y) =DX+DY$.

§ 3. მოდა, მედიანა, განაწილების კვანტილი

უწყვეტი ტიპის განაწილების მოდა ეწოდება რიცხვითი ღერძის იმ წერტილის აბსცისას, რომელშიც განაწილების სიმკვრივე აღწევს ლოკალურ მაქსიმუმს; მოდა Mo სიმბოლოთი აღინიშნება

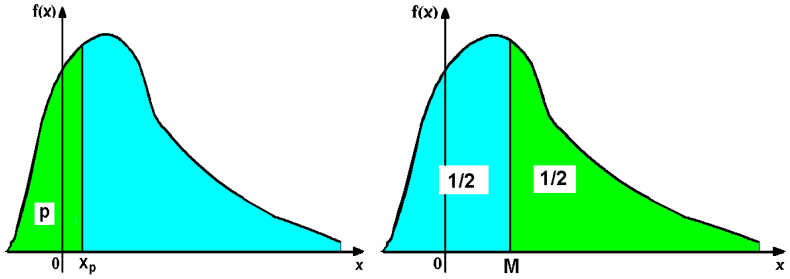
ცხადია, შესაძლებელია არსებობდეს რამდენიმე მოდა. მაშასადამე, განაწილების მოდის მოსაძებნად საჭიროა მისი სიმკვრივის მაქსიმალური მნიშვნელობის შესაბამისი აბსცისის მოძებნა. განაწილებებს, რომელთა სიმკვრივესაც გააჩნია ერთი ასეთი წერტილი, უნიმოდალური ეწოდება. ცალკეული გამონაკლისების გარდა (როგორიცაა მაგალითად, $[a,b]$ ინტერვალში თანაბარი განაწილება), მომავალში როგორც წესი, ჩვენ საქმე გვექნება უნიმოდალურ განაწილებებთან.

უწყვეტი განაწილების მქონე X შემთხვევითი სიდიდისათვის, Me მედიანა ისეთი წერტილია, რომლის „მარცხნივ“ და „მარჯვნივ“ X შემთხვევითი სიდიდის მოხვედრის ალბათობები ტოლია.

$$P\{X\le Me\}=P\{X\ge Me\}=1/2 \qquad (1.8)$$

ვთქვათ, უწყვეტი X შემთხვევითი სიდიდის განაწილების ფუნქციაა $F(x)$. განაწილების p-კვანტილი არის არგუმენტის ის x_p მნიშვნელობა, რომლისთვისაც $F(x_p) = p$. შევნიშნოთ, რომ მედიანა შეიძლება განისაზღვროს, როგორც 1/2 რიგის კვანტილი ანუ $Me=x_{1/2}$.

განაწილების სიმკვრივის ტერმინებში p-კვანტილისა და მედიანის გრაფიკული წარმოდგენა შემდეგია (იხ. ნახ. 1.2)



ნახ. 1. 2

ქვემოთ, 1.1 ცხრილში მითითებულია შემთხვევითი სიდიდეების რიცხვითი მახასიათებლების გამოთვლის ფორმულები დისკრეტული და უწყვეტი განაწილებებისათვის ცხრილი 1.1.

რიცხვითი მახასიათებლები	**დისკრეტული ტიპი**	**უწყვეტი ტიპი**
მათემატიკური ლოდინი	$EX=\sum_{i=1}^{r} x_i f_i$	$EX=\int_{-\infty}^{+\infty} uf(u)du$
k-ური რიგის მომენტი	$EX^k=\sum_{i=1}^{r} x_i^k f_i$	$EX^k=\int_{-\infty}^{+\infty} u^k f(u)du$
დისპერსია	$DX=E(X-EX)^2=EX^2-(EX)^2$	
საშუალო კვადრატული გადახრა	$\sigma_X = +\sqrt{DX}$	
p-კვანტილი	$x_p=\{x : F(x) = p\}$	
მედიანა	$Me = x_{1/2} = \min\{x : F(x) \geq 1/2\}$	

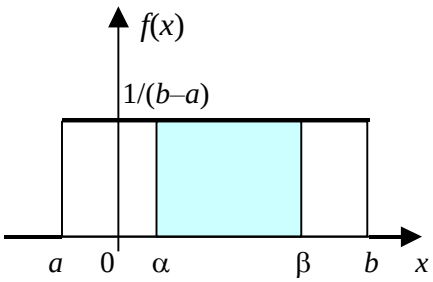
§ 4. ძირითადი უწყვეტი განაწილებანი

4.1. თანაბარი განაწილება $[a,b]$ შუალედში

ვიტყვით, რომ X შემთხვევით სიდიდეს აქვს თანაბარი განაწილება $[a,b]$ შუალედში $(-\infty<a<b<+\infty)$, ანუ X თანაბრად განაწილებულია $[a,b]$ შუალედში, თუ მისი განაწილების სიმკვრივეა

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{თუ} \quad a \leq x \leq b, \\ 0, & \text{წინააღმდეგ შემთხვევაში}. \end{cases} \tag{1.9}$$

მის გრაფიკს აქვს შემდეგი სახე:

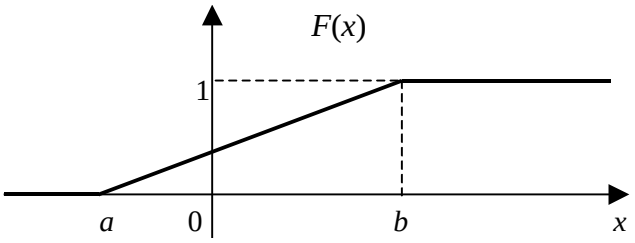


ნახ. 1.3

X შემთხვევითი სიდიდის $F(x) = \int_{-\infty}^{x} f(s)ds$ განაწილების ფუნქციას აქვს სახე

$$F(x) = \begin{cases} 0, & \text{როცა} \quad x < a, \\ (x-a)/(b-a), & \text{როცა} \quad a \le x \le b, \\ 1, & \text{როცა} \quad x > b. \end{cases} \tag{1.10}$$

ნახ. 1.4-ზე ამ ფუნქციის გრაფიკია მოცემული



ნახ. 1.4

თუ $a \le \alpha < \beta \le b$, მაშინ (1.7)-დან მივიღებთ, რომ

$$P\{\alpha < X \le \beta\} = F(\beta) - F(\alpha) = (\beta - \alpha)/(b-a), \tag{1.11}$$

ანუ ამ შუალედში მოხვედრის ალბათობა ტოლია 1.3 ნახაზზე გამოსახული დაშტრიხული მართკუთხედის ფართობისა. X შემთხვევითი სიდიდის რიცხვითი მახასიათებლებია:

$$EX = (a+b)/2, \qquad DX = EX^2 - (EX)^2 = (b-a)^2/12,$$

$$\sigma_X = (b-a)/(2\sqrt{3}).$$

(1.10)-ის ძალით გვექნება, რომ

$$F((a+b)/2) = ((a+b)/2 - a)/(b-a) = 1/2$$

და მაშასადამე, $Me = (a+b)/2$ წარმოადგენს განაწილების მედიანას. რაც შეეხება თანაბარი განაწილების მოდას, ცხადია, რომ $[a, b]$ შუალედის ნებისმიერი წერტილი შეიძლება ჩავთვალოთ განაწილების მოდად.

(a,b) ინტერვალზე თანაბრად განაწილებული X შემთხვევითი სიდიდე

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{როცა} \quad a < x \le b, \\ 0, & \text{როცა } x < a \text{ ან } x > b. \end{cases} \qquad F(x) = \begin{cases} 0, & \text{როცა} \quad x < a, \\ (x-a)/(b-a), & \text{როცა} \quad a \le x \le b, \\ 1, & \text{როცა} \quad x > b. \end{cases}$$

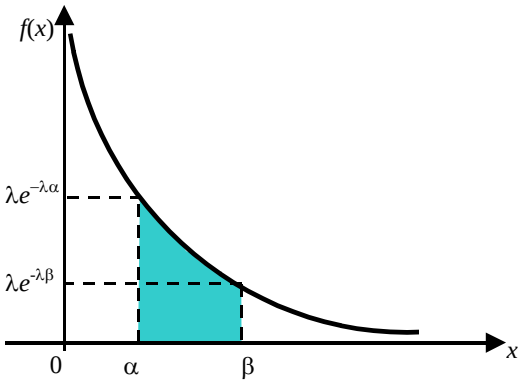
$$EX = (a+b)/2, \qquad DX = (b-a)^2/12, \qquad Me = (a+b)/2$$

4.2 ექსპონენციალური განაწილება

λ პარამეტრით ($\lambda>0$) ექსპონენციალურად განაწილებული ეწოდება ისეთ X შემთხვევით სიდიდეს, რომლის განაწილების სიმკვრივეა:

$$f(x)=\begin{cases} 0, & x<0, \\ \lambda e^{-\lambda x}, & x\geq 0. \end{cases}$$

სიმკვრივის გრაფიკს აქვს შემდეგი სახე:



ნახ. 1.5

დაშტრიხული არის ფართობი წარმოადგენს ექსპონენციალური განაწილების მქონე X შემთხვევითი სიდიდის $[\alpha,\beta]$ ინტერვალში მოხვედრის ალბათობას. შესაბამის განაწილების ფუნქციას აქვს შემდეგი სახე:

$$F(x)=\begin{cases} 0, & \text{როცა} \quad x<0, \\ 1-e^{-\lambda x}, & \text{როცა} \quad x\geq 0. \end{cases}$$

ექსპონენციალურად განაწილებული X შემთხვევითი სიდიდის მათემატიკური ლოდინია $EX=\int_{-\infty}^{+\infty} uf(u)du=\int_{0}^{+\infty}\lambda u e^{-\lambda u}du=\frac{1}{\lambda}$, ხოლო დისპერსია $DX=1/\lambda^2$ და მაშასადამე, $\sigma_X=1/\lambda$. ამრიგად, ექსპონენციალური განაწილების პარამეტრი წარმოადგენს მისი მათემატიკური ლოდინის შებრუნებულ სიდიდეს. რადგან განაწილების სიმკვრივე კლებადია, ის თავის მაქსიმალურ მნიშვნელობას აღწევს $x=0$ წერტილში და ამიტომ განაწილების მოდა $Mo=0$; რაც შეეხება მედიანას, ადვილი საჩვენებელია, რომ $M=(1/\lambda)\cdot\ln 2$.

ექსპონენციალურად განაწილებული X შემთხვევითი სიდიდე

$$f(x)=\begin{cases} 0, & x<0, \\ \lambda e^{-\lambda x}, & x\geq 0. \end{cases}$$

$$F(x)=\begin{cases} 0, & \text{როცა} \quad x<0, \\ 1-e^{-\lambda x}, & \text{როცა} \quad x\geq 0. \end{cases}$$

$$EX=\frac{1}{\lambda}, \quad DX=1/\lambda^2, \quad Me=\frac{1}{\lambda}\ln 2,$$

§ 4.3 ნორმალური განაწილება

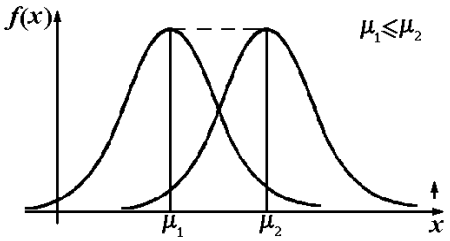
X შემთხვევით სიდიდეს, რომლის განაწილების სიმკვრივეს აქვს შემდეგი სახე:

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (1.12)$$

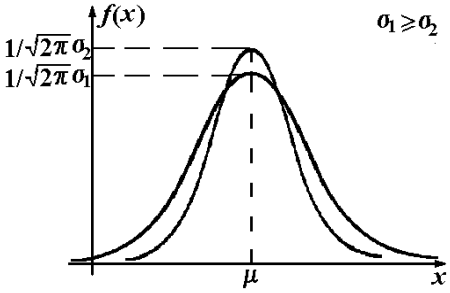
ეწოდება ნორმალურად განაწილებული შემთხვევითი სიდიდე.

განაწილების μ და σ პარამეტრების შინაარსი იმაში მდგომარეობს, რომ ისინი შესაბამისად განაწილების მათემატიკურ ლოდინსა და საშუალო კვადრატულ გადახრას წარმოადგენენ: $\mu=EX$, ხოლო $\sigma=+\sqrt{DX}$. μ პარამეტრით განისაზღვრება განაწილების სიმკვრივის მდებარეობა რიცხვთა ღერძზე, კერძოდ კი – მისი მაქსიმუმის წერტილი, ანუ მოდა: რაც უფრო დიდია მათემატიკური ლოდინი, მით უფრო მარჯვნივაა წანაცვლებული განაწილების სიმკვრივის გრაფიკი. ამასთანავე, μ პარამეტრის ცვალებადობა არავითარ გავლენას არ ახდენს განაწილების სიმკვრივის გრაფიკის ფორმაზე. σ გვიჩვენებს განაწილების სიმკვრივის გაშლილობას: რაც უფრო დიდია σ, მით უფრო გაშლილია განაწილების სიმკვრივის გრაფიკი (იხ. ნახ. 1.6 და 1.7). (1.12)-ით განსაზღვრული სიმკვრივის ფუნქცია x არგუმენტის სიმეტრიული ფუნქციაა, $x=\mu$ წრფე წარმოადგენს მისი სიმეტრიის ღერძს ანუ $f_{\mu,\sigma^2}(\mu - x) = f_{\mu,\sigma^2}(\mu + x)$; $f_{\mu,\sigma^2}(x)$ უნიმოდალური სიმკვრივეა, ის მაქსიმალურ მნიშვნელობას ღებულობს, როცა $x = \mu$ და $f_{\mu,\sigma^2}(x) = 1/\sqrt{2\pi\sigma^2}$; ამ განაწილების მოდა, მედიანა და საშუალო ტოლია; $M_0 = Me = \mu$.

ის ფაქტი, რომ X შემთხვევითი სიდიდე ნორმალურადაა განაწილებული პარამეტრებით μ და σ, მოკლედ ასე ჩაიწერება $X \sim N(\mu,\sigma)$.



ნახ. 1.6



ნახ. 1.7

აღვნიშნოთ ნორმალური განაწილების შემდეგი თვისებები:

N1. თუ $X\sim N(\mu,\sigma)$, მაშინ ამ შემთხვევითი სიდიდის ნებისმიერი $Y=aX+b$ წრფივი კომბინაცია, სადაც a და b ნამდვილი რიცხვებია, კვლავ ნორმალურადაა განაწილებული, $Y\sim N(a\mu+b,a\sigma)$. კერძოდ,

ა) თუ $X\sim N(\mu,\sigma)$, მაშინ $Z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$, (1.13)

ბ) თუ $Z\sim N(0,1)$, მაშინ $X=\sigma Z+\mu\sim N(\mu,\sigma)$ (1.14)

N2. თუ $X_1\sim N(\mu_1,\sigma_1)$, $X_2\sim N(\mu_2,\sigma_2)$ და X_1 და X_2 დამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ $X_1\pm X_2$ შემთხვევითი სიდიდე კვლავ ნორმალურადაა განაწილებული პარამეტრებით $\mu_1\pm\mu_2$ და $\sigma=\sqrt{\sigma_1^2+\sigma_2^2}$.

$N(0,1)$ განაწილების მქონე შემთხვევით სიდიდეს ეწოდება სტანდარტული ნორმალური განაწილების მქონე შემთხვევითი სიდიდე, ხოლო $N(0,1)$ განაწილებას – სტანდარტული ნორმალური განაწილება. სტანდარტული ნორმალური განაწილების სიმკვრივეს აღნიშნავენ $\varphi(x)$-ით, ხოლო განაწილების ფუნქციას – $\Phi(x)$-სიმბოლოთი.

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

და

$$\Phi(x)=\int_{-\infty}^{x}\varphi(u)du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du.$$

არსებობს ცხრილები, რომლებშიც ტაბულირებულია $\Phi(x)$ ფუნქციის მნიშვნელობები, რაც საშუალებას გვაძლევს გამოვთვალოთ ნორმალურ შემთხვევით სიდიდესთან დაკავშირებული ნებისმიერი ხდომილობის ალბათობა.

$\varphi(x)$ ფუნქციის ლუწობიდან ($\varphi(x)=\varphi(-x)$) გამომდინარეობს, რომ

$$\Phi(-x)=1-\Phi(x),\ \Phi(0)=0{,}5. \quad (1.15)$$

ამ თვისებების გამო სტანდარტული ნორმალური განაწილების ფუნქციის ცხრილები, როგორც წესი, მოცემულია მხოლოდ x არგუმენტის დადებითი მნიშვნელობებისათვის.

მოვიყვანოთ კიდევ ერთი მნიშვნელოვანი ფაქტი: თუ$X\sim N(0,1)$, მაშინ

$$P\{|X|\le x\}=2\Phi(x)-1, \quad (1.16)$$

რაც გამომდინარეობს $P\{|X|\le x\}=\Phi(x)-\Phi(-x)$ და (1.15) თანაფარდობებიდან.

მაგალითი 1.2. გამოვთვალოთ $X\sim N(0,1)$-სათვის შემდეგი ალბათობები:

ა) $P\{X<1.2\}$, ბ) $P\{X>0.5\}$, გ) $P\{X<-0.3\}$, დ) $P\{-0.3<X\le 0.7\}$, ე) $P\{|X|\le 2.3\}$.

ამოხსნა. ა) განსაზღვრების თანახმად, $P\{X<1.2\}=\Phi(1.2)$ და უშუალოდ ცხრილის x სვეტში ვეძებთ მნიშვნელობას 1.2 და იგივე სტრიქონის $\Phi(x)$ სვეტიდან ამოვწერთ $\Phi(1.2)$-ის მნიშვნელობას, რომელიც 0.88-ის ტოლია. ამრიგად $P\{X<1.2\}=0.88$.

ბ) $\{X>0.5\}$ და $\{X\le 0.5\}$ ურთიერთსაწინააღმდეგო ხდომილობებია, ამიტომ

$$P\{X>0.5\}=1-P\{X\le 0.5\}$$

და კვლავ ცხრილის საშუალებით ვპოულობთ ამ ხდომილობის ალბათობას:

$$P\{X>0.5\}=1-P\{X\le 0.5\}=1-0.69=0.31.$$

გ) $P\{X<-0.3\}$-ის გამოთვლისას ვისარგებლოთ (1.12) ფორმულით

$$P\{X<-0.3\}=\Phi(-0.3)=1-\Phi(0.3)=1-0.62=0.38.$$

დ) შემთხვევაში ალბათობის გამოსათვლელად ვისარგებლოთ (1.5) ფორმულით განაწილების ფუნქციისათვის. გვექნება

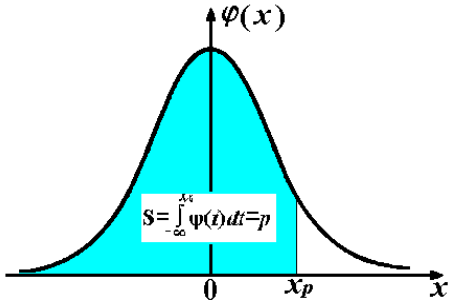
$$P\ \{-0.3<X\le 0.7\} = \Phi(0.7)-\Phi(-0.5) = 0.75-0.31 = 0.44.$$

ე) (1.12)-ის ძალით $P\{|X|\le 1.3\}= 2\Phi(1.3)-1 = 2\cdot 0.9-1 = 0.8.$ ♦

როგორც ვიცით, $N(0,1)$ განაწილების p $(0<p<1)$ რიგის x_p კვანტილი აკმაყოფილებს ტოლობას:

$$\Phi(x_p) = p.$$

თვალსაჩინოებისათვის იხ. ნახ. 1.8.



ნახ. 1.8

ვინაიდან $\Phi(0)=1/2$, საიდანაც $x_{1/2}=0$, ამიტომ $x_p<0$, როცა $p<1/2$ და $x_p>0$, როცა $p>1/2$. რადგან $\Phi(-x)=1-\Phi(x)$, ამიტომ

$$x_p = -\,x_{1-p} \qquad (1.17)$$

და ნორმალური განაწილების კვანტილების ცხრილი მოცემულია p-ს მხოლოდ იმ მნიშვნელობებისათვის, რომლებიც 1/2-ზე მეტია.

მაგალითი 1.3. ვთქვათ, $X\sim N(0,1)$. ვიპოვოთ k-ს ის მნიშვნელობა, რომლისათვისაც ა) $P\{X\le k\}=0.8$; ბ) $P\{X\le k\}=0.3$; გ) $P\{X>k\}=0.1$; დ) $P\{-k<X<k\}=0.9$; ე) $P\{|X|>k\}=0.15$.

ამოხსნა. ა) პირობის თანახმად, $\Phi(k)=0.8$, ანუ საძიებელია $N(0,1)$ განაწილების 0.8-კვანტილი. Φ-ების სვეტში ვეძებთ რიცხვით მნიშვნელობას 0.8-ს და მის გასწვრივ x-ების სვეტში შესაბამის მნიშვნელობას, რომელიც მიახლოებით 0.85-ის ტოლია, ე.ი. $k\approx 0.85$.

ბ) პირობის თანახმად, $\Phi(k)=0.3$, ანუ საძიებელია $N(0,1)$ განაწილების 0.3-კვანტილი. მაგრამ Φ-ის სვეტში არსად არ გვხვდება მნიშვნელობა 0.3. რას ნიშნავს ეს? რადგან $0.3<0.5$ და $0.5=\Phi(0)$, ამიტომ $\Phi(k)<\Phi(0)$ და ე.ი., $k<0$. ამიტომ (4.15)-ის ძალით გვექნება: $\Phi(k)=1-\Phi(-k)=0.3$,

ანუ $\Phi(-k)=0.7$. ახლა უკვე ა)-ს მსგავსად, Φ-ის სვეტში ვეძებთ რიცხვით მნიშვნელობას 0.7-ს და მის გასწვრივ x-ების სვეტში შესაბამის მნიშვნელობას, რომელიც მიახლოებით 0.55-ის ტოლია, ე.ი. $-k\approx 0.55$ და $k\approx -0.55$.

გ) $P\{X>k\}=0.1$ ტოლობიდან გამომდინარეობს, რომ $P\{X\le k\}=0.9$ და წინა შემთხვევების ანალოგიურად ვპოულობთ, რომ $k\approx 1.3$.

დ) ცხადია, რომ რადგან $P\{-k<X\le k\}=2\Phi(k)-1=0.9$, გვექნება, რომ $\Phi(k)=0.95$, საიდანაც $k\approx 1.65$.

ე) ცხადია, რომ $P\{|X|>k\}=0.15$ ტოლობიდან გამომდინარეობს $P\{|X|\le k\}=0.85$ ტოლობა, საიდანაც თავის მხრივ მივიღებთ, რომ $2\Phi(k)-1=0.85$ ანუ $\Phi(k)=0.925$ და $k\approx 1.45$. ♦

გადავიდეთ ზოგადი შემთხვევის განხილვაზე. ვთქვათ $X\sim N(\mu,\sigma^2)$, სადაც μ და σ პარამეტრების რიცხვითი მნიშვნელობები ცნობილია. ამ შემთხვევაში $\{X\le x\}$ ხდომილობის ალბათობა მარტივად გამოითვლება. მართლაც

$$P\{X\le x\} = P\left(\frac{X-\mu}{\sigma}\le\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right). \qquad (1.18)$$

ანალოგიურად

$$P\{a<X\le b\} = P\{X\le b\}-P\{X\le a\}=\Phi((b-\mu)/\sigma)-\Phi((a-\mu)/\sigma). \qquad (1.19)$$

მაგალითი 1.4. ცნობილია, რომ საკონდიტრო ქარხნის მიერ გამოშვებული ფილა შოკოლადის წონა ნორმალურადაა განაწილებული 400 გრ-ის ტოლი საშუალოთი და 20 გრ-ის ტოლი საშუალო კვადრატული გადახრით. რას უდრის ალბათობა იმისა, რომ მაღაზიაში ნაყიდი ფილა შოკოლადის წონა იქნება

ა) არ აღემატება 410 გრ-ს,

ბ) აღემატება 425 გრ-ს,

გ) არ აღემატება 380 გრ-ს,

დ) აღემატება 395 გრ-ს,

ე) მეტია 390 და ნაკლებია 412 გრ-ზე.

ამოხსნა ა) ამრიგად $X\sim N(400,20)$ საძიებელი $P\{X<410\}$ ალბათობის რიცხვითი მნიშვნელობა (4.18)-ის თანახმად

$$P\{X\le 410\} = P\left(\frac{X-400}{20}\le\frac{410-400}{20}\right)=\Phi\left(\frac{1}{2}\right).$$

უშუალოდ ცხრილის x სვეტში ვეძებთ მნიშვნელობას 0,5 და იგივე სტრიქონის $\Phi(x)$ სვეტიდან ამოვწერთ $\Phi(0,5)$-ის მნიშვნელობას, რომელიც 0.69-ის ტოლია. საბოლოოდ, $P\{X<410\}=0.69$. (69% შემთხვევაში ფილა შოკოლადის წონა ნაკლებია 410 გრ-ზე)

ბ) $P\{X>425\} = 1-P\{X<425\}=1-\Phi\left(\frac{425-400}{20}\right)=1-\Phi(1.25)=1-0,89=0,11.$

ამრიგად, ალბათობა იმისა, რომ მაღაზიაში ნაყიდი ფილა შოკოლადის წონა აღემატება 425 გრ-ს არის 0,11-ის ტოლი (ანუ 11% შემთხვევაში ფილა შოკოლადის წონა მეტია 410 გრ-ზე).

გ) $P\{X<380\}=P\left(\frac{X-400}{20}\le\frac{380-400}{20}\right)=\Phi(-1)=!-\Phi(1)=1-0,84=0,16$

ამრიგად, ალბათობა იმისა, რომ მაღაზიაში ნაყიდი ფილა შოკოლადის წონა ნაკლებია 380გრ-ზე არის 0,16-ის ტოლი (ანუ 16%) .

დ) ანალოგიურად ალბათობა იმისა, რომ მაღაზიაში ნაყიდი ფილა შოკოლადის წონა აღემატება 395გრ.-ს არის 0,59-ის ტოლი (ან 59%) .

ე) საძიებელი ალბათობის რიცხვითი მნიშვნელობა გამოითვლება ცრილური მნიშვნელობების მარტივი კომბინაციით.

$$P(390<X<412)=\Phi(0,6)-\Phi(-0,5)=0,41\ (41\%).$$ ♦

ახლა გამოვთვალოთ შემდეგი ალბათობები:

$$P\{|X-\mu|\le k\sigma\}=P\{\mu-k\sigma<X\le\mu+k\sigma\},\qquad k=1,2,3,4.$$

(4.18) და (4.19) ფორმულებიდან ცხადია, რომ სამართლიანია ტოლობა

$$P\{|X-\mu|\le k\sigma\}=P\{|Y|\le k\}=2\Phi(k)-1,$$

სადაც $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$, საიდანაც ცხრილის დახმარებით გვექნება:

k	1	2	3	4
$P\{\mu-k\sigma<X\le\mu+k\sigma\}$	0.6827	0.9545	0.9973	0.9999

ის ფაქტი, რომ $P\{|X-\mu|\le 3\sigma\}=0,9973$, ცნობილია სამი სიგმას წესის სახელწოდებით.

სტანდარტული გადახრის ინტუიციური ინტერპრეტაცია შემდეგში მდგომარეობს: თუ მონაცემები მიღებულია ნორმალური განაწილების მქონე პოპულაციიდან, მაშინ

ინტერვალი $[\bar{x}-s,\ \bar{x}+s]$ შეიცავს მონაცემთა ~ 68%-ს,

ინტერვალი $[\bar{x}-2s,\ \bar{x}+2s]$ შეიცავს მონაცემთა ~ 95%-ს,

ინტერვალი $[\bar{x}-3s,\ \bar{x}+3s]$ შეიცავს მონაცემთა ~ 99%-ს,

ინტერვალი $[\bar{x}-4s,\ \bar{x}+4s]$ შეიცავს პრაქტიკულად ყველა მონაცემს.

ახლა ვიპოვოთ $N(\mu,\sigma)$ განაწილების p-კვანტილი:

$$P\{X\le y_p\}=p.$$

სადაც, როგორც უკვე აღვნიშნეთ, $X=\mu+\sigma\cdot Y$, სადაც $Y\sim N(0,1)$. ამიტომ

$$P\{X\le y_p\}=P\{(X-\mu)/\sigma)\le(y_p-\mu)/\sigma\}=P\{X\le(y_p-\mu)/\sigma\}=\Phi((y_p-\mu)/\sigma)=p,$$

საიდანაც $(y_p-\mu)/\sigma=x_p$, ანუ

$$y_p=\mu+\sigma\cdot x_p. \qquad (1.20)$$

მაგალითი 1.5. X შემთხვევითი სიდიდე $N(5,4)$ კანონითაა განაწილებული. ვიპოვოთ ისეთი a, b და c რიცხვები, რომელთათვისაც:

ა) $P\{X\le a\}=0.7$, ბ) $P\{X>b\}=0.4$, გ) $P\{|X-5|\le c\}=0.8$.

ამოხსნა. ა) განაწილების ფუნქციისა და კვანტილის განსაზღვრების თანახმად, (1.20)-იდან მივიღებთ, რომ $a=y_{0.7}=5+4\cdot x_{0.7}$. ცხრილი A1-ის მიხედვით (იხ. დანართი) $x_{0,7}=0.55$, ე.ი. $a=5+4\cdot 0.55=7.2$.

ბ) რადგან $P\{X>b\}=1-P\{X\le b\}$, $P\{X\le b\}=0.6$ და $N(0,1)$ განაწილების კვანტილების ცხრილიდან $x_{0,6}=0.25$, ამიტომ $b=4\cdot 0.25+5=6$.

გ) $P\{|X-5|\le c\} = P\{|X-5|/4\le c/4\} = \Phi(c/4) - \Phi(-c/4) = 2\Phi(c/4) - 1 = 0.8$, საიდანაც $\Phi(c/4) = 0.9$ ანუ $c = 4\cdot X_{0.9} = 4\cdot 1.3 = 5.2$.

ნორმალურად განაწილებული X შემთხვევითი სიდიდე

$X\sim N(\mu,\sigma^2)$	$X\sim N(0,1)$
$f_{m,\sigma}(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
$F_{\mu,\sigma}(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$
$EX=\mu$, $DX=\sigma^2$	$EX=0$, $DX=1$
$\mu_{2k}=1\cdot 3\cdots(2k-1)\cdot(\sigma^2)^k$, $k=1,2,...$	$\mu_{2k}=1\cdot 3\cdot\cdots\cdot(2k-1)$, $k=1,2,...$
$\mu_{2k+1}=0$, $k=0,1,...$	$\mu_{2k+1}=0$, $k=0,1,...$

§ 4.4. χ^2-განაწილება

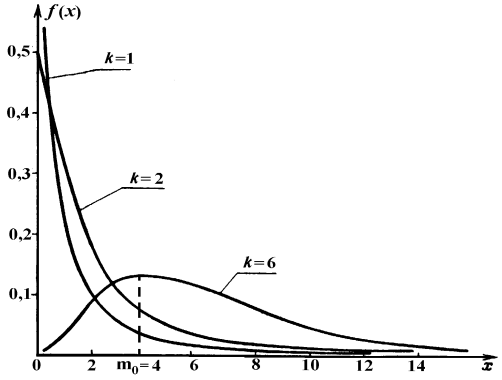
ვთქვათ, $X_1,X_2,...,X_k$ დამოუკიდებელი შემთხვევითი სიდიდეებია, სტანდარტული ნორმალური განაწილებით ($X_i\sim N(0,1)$, $i=1,2,...,k$). მაშინ ამ შემთხვევითი სიდიდეების

$$X^2(k)=X_1^2+X_2^2+\cdots+X_k^2 \qquad (1.21)$$

კვადრატების ჯამის განაწილებას უწოდებენ $\chi^2(k)$ განაწილებას (ანუ χ^2 (''ხი კვადრატ'')-განაწილებას k თავისუფლების ხარისხით). შესაკრებთა რაოდენობა k (1.21)-ში ამ განაწილების პარამეტრს წარმოადგენს. ვინაიდან $X^2(k)$ არაუარყოფითია, $\chi^2(k)$ განაწილების სიმკვრივე ნულია უარყოფით ნახევარღერძზე. როცა $k\le 2$, $\chi^2(k)$-განაწილების სიმკვრივე კლებადი ფუნქციაა, ხოლო როცა $k>2$ მას გააჩნია მაქსიმუმი $x=k-2$ წერტილში. რადგან $EX_i^2=1$, (1.21)-დან მივიღებთ, რომ

$$EX^2(k)=EX_1^2+EX_2^2+\cdots+EX_k^2=\sum_{i=1}^{k}EX_i^2=k.$$

ასევე მიიღება, რომ $DX^2(k)=2k$. ნახ. 1.9-ზე მოყვანილია $\chi^2(k)$ განაწილების სიმკვრივის გრაფიკი k-ს სხვადასხვა მნიშვნელობებისათვის.



ნახ. 1.9

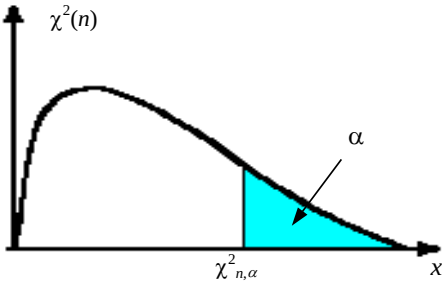
მოგვყავს $\chi^2(k)$-განაწილების ზოგიერთი თვისება:

თუ $X\sim\chi^2(k)$, $Y\sim\chi^2(m)$ და X და Y დამოუკიდებელი შემთხვევითი სიდიდეებია, მაშინ $X+Y\sim\chi^2(k+m)$, ანუ χ^2-განაწილების მქონე დამოუკიდებელი შემთხვევითი სიდიდეების შეკრებით მიიღება კვლავ χ^2-განაწილების მქონე შემთხვევითი სიდიდე, რომლის თავისუფლების ხარისხი შესაკრებთა თავისუფლების ხარისხების ჯამის ტოლია.

ვთქვათ, $X\sim\chi^2(k)$ და

$$P\{X>x_{n,\alpha}\}=P\{X\in(x_{n,\alpha},+\infty)\}=\alpha.$$

$(x_{n,\alpha},+\infty)$ არეს χ^2განაწილების ცალმხრივ, მარჯვენა (ზედა) α ზომის არეს, ხოლო $x_{n,\alpha}$–ს ზედა α კრიტიკულ წერტილს უწოდებენ. დანართში (ცხრილი A3) მოცემულია χ^2-განაწილების ზედა α-კრიტიკული წერტილების რიცხვითი მნიშვნელობები სხვადასხვა k-სა (k=1,2,..,30) და α-სათვის (α=0.01; 0.025; 0.05; 0.95; 0.975; 0.99);



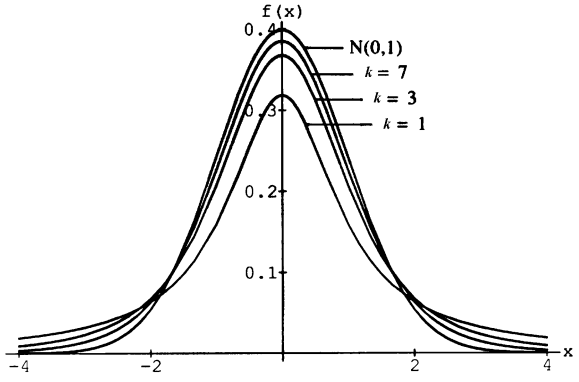
ნახ 1.10

§ 4.5. სტიუდენტის $t(k)$-განაწილება

სტიუდენტის $t(k)$-განაწილება ან როგორც ამბობენ, სტიუდენტის t-განაწილება k თავისუფლების ხარისხით აქვს შემთხვევით სიდიდეს

$$T(k)=\frac{X}{\sqrt{X^2(k)/k}}, \qquad (1.22)$$

სადაც $X \sim N(0,1)$, $X^2(k)$-ს აქვს χ^2-განაწილება k თავისუფლების ხარისხით და X და $X^2(k)$ შემთხვევითი სიდიდეები დამოუკიდებელია. მოვიყვანოთ ამ განაწილების სიმკვრივის გრაფიკი:



ნახ. 1.11

$t(k)$-განაწილების რიცხვითი მახასიათებლებია:

$$ET(k)=0,\ DT(k) = \begin{cases} \dfrac{k}{k-2}, & \text{როცა} \quad k>2, \\ \infty, & \text{როცა} \quad 0<k\le 2. \end{cases}$$

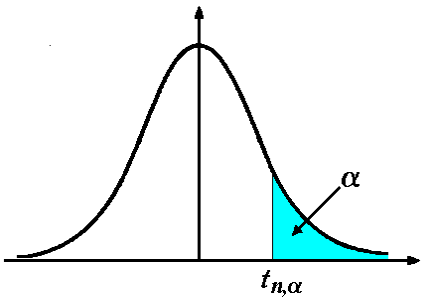
ნორმალურის მსგავსად, t-განაწილების სიმკვრივეც სიმეტრიული და უნიმოდალურია. $x=0$ წერტილი წარმოადგენს ამ განაწილების მედიანასაც, მოდასაც და საშუალოსაც.

ვთქვათ, $X \sim t(k)$ და

$$P\{X>t_{n,\alpha}\}=P\{X\in(t_{n,\alpha}, +\infty)\}=\alpha.$$

$(t_{n,\alpha}, +\infty)$ არეს $t(k)$-განაწილების ცალმხრივ, მარჯვენა (ზედა) α ზომის არეს, ხოლო $t_{n,\alpha}$-ს ზედა α კრიტიკულ წერტილს უწოდებენ. დანართში (ცხრილი A4) მოცემულია t-განაწილების ზედა α-კრიტიკული წერტილების, $t_{k,\alpha}$-ს რიცხვითი მნიშვნელობები სხვადასხვა k-სა (k=1,2,..,30) და α-სათვის (α=0.01; 0.025; 0.05; 0.95; 0.975; 0.99). შევნიშნოთ, რომ $t(k)$-განაწილებას (როგორც 0-ის მიმართ სიმეტრიულ განაწილებას), აქვს ისეთივე თვისებები, როგორიც აქვს სტანდარტული ნორმალურ განაწილებას. კერძოდ, $t_{k,1-\alpha}=-t_{k,\alpha}$. როცა თავისუფლების ხარისხი $k\to\infty$ (პრაქტიკულად, როცა k>30), t-განაწილება სწრაფად უახლოვდება სტანდარტულ ნორმალურ განაწილებას:

$$P\{T(k)\le x\} \approx \Phi(x).$$



ნახ. 1.12. t განაწილების ზედა α-კრიტიკული წერტილები, $t_{n,\alpha}$.

§ 5. ასიმეტრია და ექსცესი

ჩვენ განვიხილეთ განაწილებათა რიცხვითი მახასიათებლები, მათემატიკური ლოდინი, დისპერსია, მთელი დადებითი k რიგის სასრული საწყისი და ცენტრალური მომენტები.

ცენტრალური მომენტების მეშვეობით განისაზღვრება ისეთი მნიშვნელოვანი მახასიათებლები როგორიცაა ასიმეტრიისა და ექსცესის კოეფიციენტები.

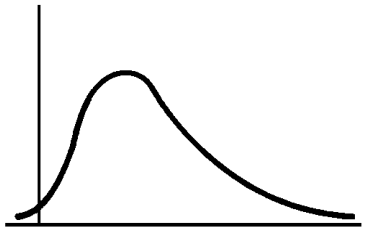
ასიმეტრიის ზომას საფუძვლად ედება საშუალო კუბური გადახრა და ასიმეტრიის კოეფიციენტს ანგარიშობენ

$$\alpha = \frac{E(X - EX)^3}{\sigma^3}$$

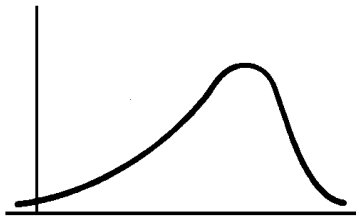
ფორმულით. საშუალო კუბური გადახრა საშუალებას გვაძლევს უფრო სრულად გავითვალისწინოთ შემთხვევითი სიდიდის დიდი გადახრები. განაწილების ასიმეტრიის შემთხვევაში განაწილების მრუდის ერთი მხარე იძლევა უფრო დიდ კუბურ გადახრას მეორე მხარესთან შედარებით და რადგან კუბური გადახრის დროს გადახრის ნიშანი ნარჩუნდება, კუბურ გადახრებს შორის განსხვავება აჩვენებს დადებით ან უარყოფით ასიმეტრიას.

თუ X შემთხვევითი სიდიდის განაწილება სიმეტრიულია თავისი $\mu=EX$ საშუალო მნიშვნელობის მიმართ, მაშინ X შემთხვევითი სიდიდის ყველა კენტი რიგის ცენტრალური მომენტი ნულის ტოლია. აქედან გამომდინარე ასიმეტრიის კოეფიციენტიც ნულის ტოლია.

თუ $\alpha>0$, მაშინ განაწილება მარჯვნივ ასიმეტრიულია (ნახ. 1.13), ხოლო $\alpha<0$ ნიშნავს, რომ განაწილება მარცხნივ ასიმეტრიულია (ნახ. 1.14).



ნახ. 1.13



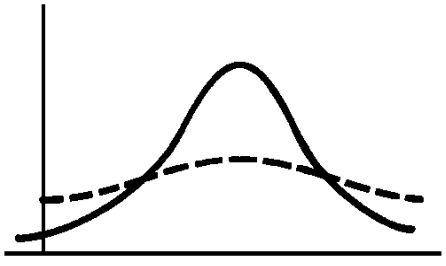
ნახ. 1.14

შევნიშნოთ, რომ ნორმალური და კერძოდ სტანდარტული ნორმალური განაწილების და t-განაწილების ასიმეტრიის კოეფიციენტი ნულის ტოლია, ხოლო χ^2 განაწილების ასიმეტრიის კოეფიციენტი დადებითია.

ექსცესის კოეფიციენტი e ახასიათებს განაწილების კონცენტრაციის ხარისხს საშუალო მნიშვნელობის ირგვლივ. იგი გამოითვლება

$$e = \frac{E(X - EX)^4}{\sigma^4} - 3$$

ფორმულით. რაც უფრო დიდია e, მით მეტადაა კონცენტრირებული განაწილება μ-ს ირგვლივ, ანუ სიმკვრივეს μ-წერტილში აქვს მაღალი პიკი, და პირიქით



ნახ. 1.15

თუ რაიმე უწვეტი განაწილების მქონე X შემთხვევითი სიდიდის ექსცესის კოეფიციენტი დადებითია, ეს მიუთითებს იმაზე, რომ სტანდარტული შემთხვევითი სიდიდის $\frac{X-EX}{\sqrt{DX}}$ განაწილების სიმკვრივე ნულის მიდამოში უფრო გაწელილია ზევით, ვიდრე სტანდარტული ნორმალური განაწილების სიმკვრივე.

t-განაწილებისათვის e ექსცესის კოეფიციენტი უარყოფითია.

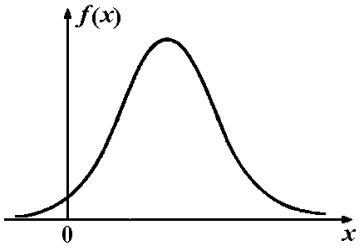
კითხვები თვითშემოწმებისათვის

- როგორ შემთხვევით სიდიდეს ეწოდება უწყვეტი?
- როგორ ფუნქციას ვუწოდებთ უწყვეტი შემთხვევითი სიდიდის განაწილების სიმკვრივეს?
- რა დამოკიდებულობაა უწვეტი შემთხვევითი სიდიდის განაწილების სიმკვრივესა და განაწილების ფუნქციას შორის?
- შეიძლება თუ არა, რომ არგუმენტის რაიმე მნიშვნელობისათვის ა) განაწილების ფუნქცია მეტი იყოს ერთზე? ბ) განაწილების სიმკვრივე მეტი იყოს ერთზე? გ) განაწილების ფუნქცია იყოს უარყოფითი? დ) განაწილების სიმკვრივე იყოს უარყოფითი?
- როგორ გამოითვლება X შემთხვევითი სიდიდის განაწილების ფუნქციით ან განაწილების სიმკვრივით $(a<X<b)$ ხდომილობის ალბათობა?
- რას ეწოდება უწყვეტი შემთხვევითი სიდიდის მათემატიკური ლოდინი; დისპერსია; საშუალო კვადრატული გადახრა?
- როგორ განიმარტება უწყვეტი შემთხვევითი სიდიდის მედიანა და მოდა?
- რა შემთხვევაში ეწოდება უწყვეტ შემთხვევით სიდიდეს თანაბრად განაწილებული?
- რას უდრის (a,b) –ზე თანაბრად განაწილებული შემთხვევითი სიდიდის მათემატიკური ლოდინი და დისპერსია?
- რა შემთხვევაში ეწოდება უწყვეტ შემთხვევით სიდიდეს ნორმალურად განაწილებული?
- როგორი სახე აქვს ნორმალურად გენაწილებული შემთხვევითი სიდიდის სიმკვრივეს? რა შინაარსი აქვს სიმკვრივის გამოსახულებაში შემავალ პარამეტრებს? რას უდრის ნორმალურად განაწილებული შემთხვევითი სიდიდდის მოდა, მედიანა?
- როგორ შემთხვევით სიდიდეთა როგორი ფუნქციაა χ^2 კანონით განაწილებული შემთხვევითი სიდიდე? რისი ტოლია განაწილების თავისუფლების ხარისხი?
- როგორ შემთხვევით სიდიდეთა როგორი ფუნქციაა სტიუდენტის კანონით განაწილებული შემთხვევითი სიდიდე? რისი ტოლია განაწილების თავისუფლების ხარისხი?
- როგორ შემთხვევით სიდიდეთა როგორი ფუნქციაა ფიშერის კანონით განაწილებული შემთხვევითი სიდიდე?

სავარჯიშოები

1.1. X შემთხვევითი სიდიდის განაწილების ფუნქციაა $F_X(x)$. როგორია $Y=aX+b$ შემთხვევითი სიდიდის $F_Y(x)$ განაწილების ფუნქცია, სადაც a და b მუდმივებია.

1.2. მოცემულია X შემთხვევითი სიდიდის განაწილების სიმკვრივის გრაფიკი:



იგივე გრაფიკზე ააგეთ $Y=X+b$ შემთხვევითი სიდიდის განაწილების სიმკვრივის გრფიკი, სადაც $b>0$ მუდმივია

1.3. უწყვეტი X შემთხვევითი სიდიდის განაწილების ფუნქციაა

$$F(x)=\begin{cases} 0, & x \le 0, \\ \sin x, & 0 < x \le \pi/2, \\ 1, & x > \pi/2. \end{cases}$$

რისი ტოლია: ა) სიმკვრივე, ბ) $P\{-2 \le x \le \pi/4\}$;

1.4. უწყვეტი X შემთხვევითი სიდიდის განაწილების ფუნქციაა

$$F(x)=\begin{cases} 0, & x \le 0, \\ \sin 2x, & 0 < x \le \pi/4, \\ 1, & x > \pi/4. \end{cases}$$

რისი ტოლია ა) $f(x)$ სიმკვრივე, ბ) $P\{\pi/6 \le X \le \pi/2\}$.

1.5. უწყვეტი X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases} 0, & x \le 1, \\ x-\dfrac{1}{2}, & 1 < x \le 2, \\ 0, & x > 2. \end{cases}$$

რისი ტოლია ა) $F(x)$ განაწილების ფუნქცია, ბ) $P\{1/2<X\le 3/2\}$, გ) EX, დ) $p=1/3$ დონის კვანტილი, ე) DX.

1.6. X შემთხვევითისიდიდის განაწილების ფუნქციაა

$$F(x)=\begin{cases} 0, & \text{როცა} \quad x \le 2, \\ 0{,}5x-1, & \text{როცა} \quad 2 < x \le 4, \\ 1, & \text{როცა} \quad x > 4 \end{cases}$$

იპოვეთ ალბათობა იმისა, რომ X შემთხვევითი სიდიდე მიიღებს ა) 0,2-ზე ნაკლებ მნიშვნელობას, ბ) 3-ზე ნაკლებ მნიშვნელობას, გ) 3-ზე არანაკლებ მნიშვნელობას, დ) 5-ზე არანაკლებ მნიშვნელობას.

1.7. X შემთხვევითისიდიდის განაწილების ფუნქციაა

$$F(x)=\begin{cases} 0, & \text{როცა} \quad x \le 0, \\ x^2, & \text{როცა} \quad 0 < x \le 1, \\ 1, & \text{როცა} \quad x > 1 \end{cases}$$

იპოვეთ ალბათობა იმისა, რომ X შემთხვევითი სიდიდე ოთხი დამოუკიდებელი დაკვირვების შედეგად სამჯერ მიიღებს მნიშვნელობას $(0.25;\ 0.75)$ ინტერვალიდან.

1.8. X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases} 2x, & \text{როცა} \quad x \in (0,1), \\ 0, & \text{როცა} \quad x \notin (0,1) \end{cases}$$

რისი ტოლია ა) $F(x)$ განაწილების ფუნქცია, ბ) EX, DX, გ) $p=4/9$ დონის კვანტილი

1.9. X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases}(1/2)x, & \text{როცა} \quad x\in(0,2),\\ 0, & \text{როცა} \quad x\notin(0,2)\end{cases}$$

რისი ტოლია X შემთხვევითი სიდიდის მათემატიკური ლოდინი?

1.10. X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases}-(3/4)x^2+(9/2)x-6, & \text{როცა } x\in(2,4),\\ 0, & \text{როცა } x\notin(2,4).\end{cases}$$

იპოვეთ ა) მოდა, ბ) მათემატიკური ლოდინი, გ) მედიანა, დ)განაწილების ფუნქცია, ე)გამოთვალეთ P{2<X<3}.

1.11. X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases}-(3/4)x^2+6x-45/4, & \text{როცა } x\in(3,5),\\ 0, & \text{როცა } x\notin(3,5).\end{cases}$$

იპოვეთ ა) მოდა=4, ბ) მათემატიკური ლოდინი, გ) მედიანა, დ) განაწილების ფუნქცია.

1.12. უწყვეტი X შემთხვევითი სიდიდის განაწილების სიმკვრივეა

$$f(x)=\begin{cases}C(x^2+2x), & x\in(0,1),\\ 0, & x\notin(0,1).\end{cases}$$

რისი ტოლია

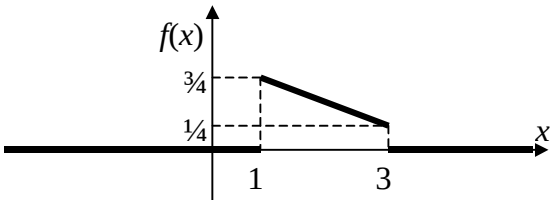
ა) C კოეფიციენტი, ბ) EX, გ) $F(x)$ განაწილების ფუნქცია, დ) $P\{-3\le X\le 1/2\}$

1.13.

1.მოცემულია უწყვეტი X შემთხვევითი სიდიდის განაწილების სიმკვრივის გრაფიკი.

ა ანალიზური სახით ჩაწერეთ განაწილების სიმკვრივე და განაწილების ფუნქცია.

ბ გამოთვალეთ $P\{2\le X\le 4\}$

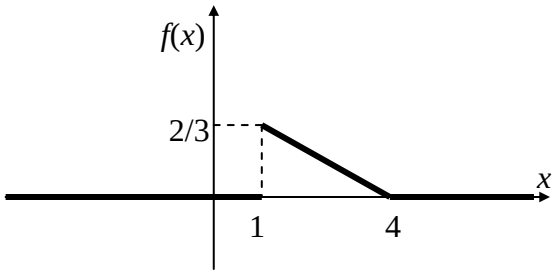


1.14.

1.მოცემულია უწყვეტი X შემთხვევითი სიდიდის განაწილების სიმკვრივის გრაფიკი.

ა ანალიზური სახით ჩაწერეთ განაწილების სიმკვრივე

ბ გამოთვალეთ$P\{2\le X\le 3\}$



ლექცია II,III ალბათობის თეორიის ზოგიერთი ზღვარითი თეორემა

შემთხვევით სიდიდეთა დიდი რაოდენობის ჯამების ყოფაქცევა სპეციალურ შესწავლას საჭიროებს. მაგალითად, განვიხილოთ დამოუკიდებელი, ერთნაირად განაწილებული $X_1,X_2,...,X_n$ შემთხვევითი სიდიდეების არითმეტიკული საშუალო

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

დავუშვათ, რომ $EX_i=\mu$, $DX_i=\sigma^2$, $i=1,2,..,n$. მათემატიკური ლოდინისა და დისპერსიის თვისებების გამო

$$E\,\overline{X}_n = \mu, \quad D\,\overline{X}_n = \frac{\sigma^2}{n}, \tag{2.1}$$

ანუ $\overline{X}_n$ შემთხვევით სიდიდეს იგივე საშუალო აქვს, რაც ყოველ შესაკრებს, მაგრამ ხასიათდება n-ჯერ უფრო მცირე გაფანტულობით თავისი საშუალოს მიმართ, ვიდრე თითოეული შესაკრები. ქვემოთ ჩვენ მოვიყვანთ დებულებებს არითმეტიკული საშუალოს ყოფაქცევის შესახებ შესაკრებთა რაოდენობის ზრდისას, რომლებიც ეყრდნობა ჯამის დისპერსიის შემცირების აღწერილ ეფექტს და შევეხებით შემთხვევით შესაკრებთა ჯამების განაწილების სხვა თვისებებს.

§ 1. ჩებიშევის უტოლობა

X შემთხვევითი სიდიდის საკუთარი მათემატიკური ლოდინისაგან გადახრის შესაფასებლად იყენებენ ე.წ. *ჩებიშევის უტოლობას*, რომელიც მიიღება შემდეგი ელემენტარული უტოლობიდან.

ვთქვათ, X რაიმე დადებით მნიშვნელობებიანი შემთხვევითი სიდიდეა, რომლის მათემატიკური ლოდინი სასრულია: $EX < \infty$. მაშინ ნებისმიერი დადებითი ε რიცხვისათვის

$$P\{X \ge \varepsilon\} \le EX / \varepsilon. \tag{2.2}$$

მართლაც, თუ შემოვიფარგლებით X დისკრეტული შემთხვევითი სიდიდით, რომელიც ღებულობს $x_1, x_2,..., x_r$ დადებით მნიშვნელობებს $p_1, p_2, ..., p_r$ ალბათობებით, გვექნება

$$EX = \sum_{i=1}^{r} x_i p_i = \sum_{i:x_i \ge \varepsilon} x_i p_i + \sum_{i:x_i < \varepsilon} x_i p_i \ge \sum_{i:x_i \ge \varepsilon} x_i p_i \ge \varepsilon \sum_{i:x_i \ge \varepsilon} p_i = \varepsilon\, P\{X \ge \varepsilon\}.$$

ამით უტოლობა (2.2) ნაჩვენებია. იგი ასევე მარტივად მტკიცდება ზოგად შემთხვევაში.

ჩებიშევის უტოლობა. ვთქვათ, X რაიმე შემთხვევითი სიდიდეა სასრული DX დისპერსიით, მაშინ სამართლიანია შემდეგი უტოლობა:

$$P\{|X-EX| \ge \varepsilon\} \le DX / \varepsilon^2. \tag{2.3}$$

დამტკიცება. ცხადია, რომ $P\{|X-EX| \ge \varepsilon\} = P\{(X-EX)^2 \ge \varepsilon^2\}$. (6.2) უტოლობის ძალით $P\{|X-EX| \ge \varepsilon\} = P\{(X-EX)^2 \ge \varepsilon^2\} \le E(X-EX)^2 / \varepsilon^2 = DX / \varepsilon^2$.

ჩებიშევის უტოლობა შეიძლება ჩაიწეროს ეკვივალენტური ფორმით:

$$P\{|X-EX| \geq \varepsilon\sqrt{DX}\ \} \leq 1/\varepsilon^2. \qquad (2.4)$$

ეს უტოლობა ნათლად წარმოაჩენს დისპერსიას, როგორც გაბნევის მახასიათებელს: (2.4)-ის ძალით X შემთხვევითი სიდიდის საკუთარი EX მათემატიკური ლოდინისაგან $\varepsilon\sqrt{DX}$ -ზე არანაკლები გადახრის ალბათობა არ აღემატება $1/\varepsilon^2$ სიდიდეს.

შენიშვნა. (2.3) უტოლობა შეიძლება ჩაიწეროს შემდეგი ექვივალენ-ტური სახით:

$$P(|X-EX| \leq \varepsilon) > 1-\frac{DX}{\varepsilon^2}.$$

ამ სახით ჩაწერილი ჩებიშევის უტოლობა განსაზღვრავს $|X-EX|<\varepsilon$ ხდომილობის ალბათობას, რომელიც ქვევიდანაა შემოსაზღვრული $1-DX/\varepsilon^2$ სიდიდით.

მაგალითი 2.1. შეაფასეთ ზემოდან $P\{|X-EX|\geq 2\}$, თუ ცნობილია, რომ $DX=2$.

ამოხსნა. ვისარგებლოთ (2.3) ფორმულით გვექნება

$$P(|X-EX| \geq 2) \leq DX / 2^2 = 2/2^2 = 1/2.$$

საზოგადოდ ჩებიშევის უტოლობა იძლევა მათემატიკური ლოდინისაგან შემთხვევითი სიდიდის გადახრის მხოლოდ უხეშ შეფასებას, რასაც ადასტურებს შემდეგი

მაგალითი 2.2. ვთქვათ, $EX=\mu$, $DX=\sigma^2$ და $\varepsilon=3\sigma$. მაშინ ჩებიშევის უტოლობით გვექნება:

$$P\{|X-\mu|\geq 3\sigma\}<\sigma^2/(9\sigma^2) = 1/9 \approx 0.1111.$$

თუ დავუშვებთ, რომ X განაწილებულია ნორმალურად μ და σ^2 პარამეტრებით, მაშინ, როგორც შესაბამისი ცხრილებიდან ჩანს, იმავე ხდომილობის ალბათობა დიდი სიზუსტით ტოლია 0.0027 რიცხვისა, რაც გაცილებით ნაკლებია ვიდრე 1/9.

§ 2. დიდ რიცხვთა კანონი

ვთქვათ $X_1,X_2,...,X_n,...$ დამოუკიდებელი ერთნაირად განაწილებული შემთხვევით სიდიდეთა მიმდევრობაა μ მათემატიკური ლოდინითა და σ^2 დისპერსიით.

განვიხილოთ

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

შემთხვევითი სიდიდეების მიმდევრობა. (2.1) ფორმულისა და ჩებიშევის უტოლობის ძალით, ნებისმიერი დადებითი ε რიცხვისათვის გვექნება:

$$P\{|\overline{X}_n - E\ \overline{X}_n|>\varepsilon\} = P\{|\overline{X}_n-\mu|>\varepsilon\} \leq \frac{D\overline{X}_n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}, \qquad (2.5)$$

საიდანაც იმის გათვალისწინებით, რომ $\frac{\sigma^2}{n\varepsilon^2} \to 0$, როცა $n \to \infty$, მივიღებთ, რომ

$$P\{|\overline{X}_n-\mu| > \varepsilon\} \to 0, \quad \text{როცა } n \to \infty. \qquad (2.6)$$

უკანასკნელი თანაფარდობა სამართლიანია იმ შემთხვევაშიც, როდესაც შემთხვევით სიდიდეებს არ გააჩნიათ სასრული დისპერსია. საბოლოოდ შეიძლება ჩამოვაყალიბოთ თეორემა, რომელიც დიდ რიცხვთა კანონის სახელითაა ცნობილი

დიდ რიცხვთა კანონი. თუ $X_1,X_2,...,X_n,...$ ერთნაირად განაწილებული დამოუკიდებელი შემთხვევითი სიდიდეებია საერთო μ მათემატიკური ლოდინით, მაშინ ნებისმიერი დადებითი

ε-ისათვის

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right\} = 0\,. \qquad (2.7)$$

შემთხვევით სიდიდეთა $Y_1,Y_2,..,Y_n,...$ მიმდევრობას ეწოდება ალბათობით კრებადი Y შემთხვევითი სიდიდისაკენ, რაც $Y_n \xrightarrow{P} Y$ სიმბოლოთი აღინიშნება, თუ ნებისმიერი დადებითი ε-ისათვის

$$\lim_{n\to\infty} P\{|Y_n - Y| > \varepsilon\} = 0. \qquad (2.8)$$

აქედან გამომდინარე დიდ რიცხვთა კანონი მოკლედ ასე ჩაიწერება:

$$\overline{X}_n \xrightarrow{P} \mu.$$

ახლა განვიხილოთ ერთი მნიშვნელოვანი კერძო შემთხვევა, როდესაც X_i შემთხვევითი სიდიდეები ბერნულის შემთხვევითი სიდიდეებია წარმატების p ალბათობით

X_i	1	0
	p	$1-p$

ამ შემთხვევაში $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$ წარმოადგენს წარმატების ფარდობით სიხშირეს n დამოუკიდებელ ცდაში. ვინაიდან $EX_i=p$, ამიტომ დიდ რიცხვთა კანონს ბერნულის სქემისათვის ექნება შემდეგი სახე:

ბერნულის თეორემა .

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} = 0\,, \qquad \left(\frac{S_n}{n} \xrightarrow{P} p\right)$$

ლიტერატურაში $\frac{S_n}{n}$ ფარდობით სიხშირეს უწოდებენ შერჩევით პროპორციას და მას $\hat{P}_n$ სიმბოლოთი აღნიშნავენ, $\hat{P}_n = \frac{S_n}{n}$.

ჩავთვალოთ, რომ p არის იმ უსასრულო პოპულაციის პროპორციად, რომელიც ბერნულის შემთხვევით სიდიდეს შეესაბამება. მაშინ ბერნულის თეორემა შემდეგნაირად ჩაიწერება:

$$\hat{P}_n \xrightarrow{P} p.$$

§ 3. ცენტრალური ზღვარითი თეორემა (კლასიკური)

ვთქვათ $X_1,X_2,...,X_n,...$ დამოუკიდებელ ნორმალურად განაწილებული შემთხვევით სიდიდეთა მიმდევრობაა μ საშუალოთი და σ^2 დისპერსიით. როგორც ცნობილია $S_n=X_1+\cdots+X_n$

შემთხვევითი სიდიდე კვლავ ნორმალურადაა განაწილებული $n\mu$ საშუალოთი და $n\sigma^2$ დისპერსიით, ხოლო ნორმირებული ჯამების განაწილება ნორმალურია ნულოვანი საშუალოთი და ერთეულოვანი დისპერსიით, ე.ი.

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1). \qquad (2.9)$$

პირიქით, თუ დამოუკიდებელ ერთნაირად განაწილებულ შემთხვევით შესაკრებთა ჯამისათვის ძალაშია (2.9), მაშინ ამ შემთხვევით სიდიდეთა საერთო განაწილება აუცილებლად ნორმალური აღმოჩნდება. ამრიგად, როდესაც X_i შემთხვევით სიდიდეთა განაწილება ნორმალური არაა, S_n შემთხვევითი სიდიდის განაწილება არ შეიძლება იყოს ნორმალური. მიუხედავად ამისა ნორმირებული ჯამებისათვის გარანტირებულია მიახლოებითი ნორმალურობა, როდესაც შესაკრებთა რაოდენობა უსასრულოდ იზრდება. ამ ფაქტის ზუსტი ფორმულირება ცნობილია ცენტრალური ზღვარითი თეორემის სახელწოდებით.

ცენტრალური ზღვარითი თეორემა. თუ $X_1, X_2, \ldots, X_n$ დამოუკიდებელი ერთნაირად განაწილებული შემთხვევითი სიდიდეებია საერთო μ მათემატიკური ლოდინითა და სასრული σ^2 დისპერსიით, ხოლო $S_n = X_1 + X_2 + \cdots + X_n$, მაშინ

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) - \Phi(x) \to 0, \quad \text{როცა } n \to \infty, \qquad (2.10)$$

სადაც $\Phi(x)$ არის სტანდარტული ნორმალური განაწილების ფუნქცია.

მაგრამ ცენტრალური ზღვარითი თეორემის (2.10) ფორმულირება არაა საკმარისი იმ ფართო გავრცელების ასახსნელად, რომელიც ნორმალურ განაწილებას გააჩნია სულ სხვადასხვა მოვლენის მოდელის როლში. არსებობს ბევრად უფრო ზოგადი პირობები, როცა ადგილი აქვს დამოუკიდებელ (ან თუნდაც დამოკიდებულ) შემთხვევით სიდიდეთა ჯამების განაწილების სიახლოვეს ნორმალურ განაწილებასთან პირობის ყველაზე ზოგადი ფორმულირება სიტყვიერად შეიძლება ასე გამოითქვას: შესაკრებთა რაოდენობის ზრდისას თითოეულ მათგანს გარკვეული აზრით თანაბრად მცირე წვლილი უნდა შეჰქონდეს მთელს ჯამში. ამიტომ

პრაქტიკული სიტუაციების განხილვისას, თუ რაიმე სიდიდე ბევრი სხვა მცირე სიდიდის ჯამურ ეფექტს წარმოადგენს, მაშინ მივიჩნევთ, რომ აღნიშნული სიდიდის განაწილების კანონი ნორმალურია.

(2.10) თანაფარდობიდან მარტივად გამომდინარეობს შემდეგი ფაქტი

$$P\left(\frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \le x\right) \to \Phi(x), \quad \text{როცა } n \to \infty.$$

ამრიგად, დავასკვნით, რომ:

შესაკრებთა საკმაოდ დიდი რაოდენობისათვის ამ სიდიდეთა არითმეტიკული საშუალო $\bar{X}_n$ მიახლოებით ნორმალურადაა განაწილებული μ საშუალოთი და σ^2/n დისპერსიით.

ლიტერატურაში, რომელიც ეხება სტატისტიკის გამოყენებით ასპექტებს, ცენტრალური ზღვარითი თეორემა სწორედ ამ ტერმინებშია ჩამოყალიბებული.

ბერნულის სქემისათვის წარმატების მუდმივი p ალბათობით, $0<p<1$, $q=1-p$ ცენტრალური ზღვარითი თეორემა იძენს შემდეგ ფორმას: ყოველი ნამდვილი x-ისათვის

$$P\left\{\frac{S_n - np}{\sqrt{npq}} \le x\right\} \to \Phi(x), \qquad \text{როცა } n\to\infty. \tag{2.11}$$

სადაც S_n - n ცდაში « წარმატების » აბსოლუტური სიხშირეა (წარმატებათა რიცხვია). ცნობილია, რომ $DS_n=npq$.

ამ თეორემის მეშვეობით შესაძლებელია გამოითვალოს S_n შემთხვევითი სიდიდის ზღვარითი განაწილებაც, სახელდობრ,

$$P\{S_n \le x\} - \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \to 0, \quad \text{როცა } n\to\infty. \tag{2.12}$$

აქედან უკვე მარტივად გამომდინარეობს

მუავრ-ლაპლასის თეორემა. როგორიც არ უნდა იყოს a და b ნამდვილი რიცხვები, $a<b$, S_n შემთხვევითი სიდიდისათვის ძალაშია შემდეგი მიახლოებითი ფორმულა:

$$P\{a \le S_n \le b\} \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right), \quad \text{როცა } n\to\infty. \tag{2.13}$$

თუ n საკმაოდ დიდია და წარმატების p ალბათობა არ არის ახლოს ან 1-თან ან 0-თან, ბინომური განაწილების მიახლოება ნორმალურით საკმაოდ ზუსტ შედეგებს იძლევა.

მაგალითი 2.3. 100 დამოუკიდებელ ცდაში A ხდომილობის განხორციელების ალბათობა მუდმივია და 0.8-ის ტოლია.ნორმალური აპროქსიმაციის გამოყენებით იპოვეთ ალბათობა იმისა, რომ A ხდომილობა 100 ცდაში განხორციელდება: ა) არა უმცირეს 75-ჯერ და არა უმეტეს 90-ჯერ ; ბ) არა უმცირეს 75-ჯერ ; გ)არა უმეტეს 90-ჯერ .

ამოხსნა. ამოცანის პირობის თანახმად, $n=100$, $p=0.8$, $a=75$, $b=90$ და საძიებელია შემდეგი ალბათობები: ა) $P\{75\le X\le 90\}$, ბ)$P\{X\ge 75\}$, გ) $P\{X\le 90\}$.

ა) (2.13)-ის ძალით გვაქვს

$$P\{75\le X\le 90\} = \sum_{k=75}^{90} C_n^k p^k (1-p)^{n-k} \approx \Phi\left(\frac{90-100\cdot 0.8}{\sqrt{100\cdot 0.8\cdot 0.2}}\right) - \Phi\left(\frac{75-100\cdot 0.8}{\sqrt{100\cdot 0.8\cdot 0.2}}\right) = 0.888.$$

ბ) A ხდომილობის განხორციელება არა უმცირეს 75-ჯერ ნიშნავს, რომ ის განხორციელდება ან 75-ჯერ ან 76-ჯერ და ა.შ. 100-ჯერ. ამიტომ $P\{X\ge 75\}= P\{75\le X\le 100\}$ და ამ შემთხვევაში უნდა ჩავთვალოთ, რომ $a=75$, ხოლო $b=100$. საბოლოოდ, ანალოგიურად წინა ამოცანისა, მივიღებთ $P\{X\ge 75\}=P\{75<X\le 100\}= \Phi(5)-\Phi(-1.25)= 0.89$.

გ) A ხდომილობა განხორციელდება არა უმეტეს 90-ჯერ ნიშნავს, რომ A ხდომილობა ან არ განხორციელდება ან განხორციელდება 1-ჯერ ან 2-ჯერ და ა.შ. 90-ჯერ. ამიტომ ანალოგიურად ბ) შემთხვევისა მივიღებთ, რომ $P(X\le 90)=\Phi(2.5)-\Phi(-20)=0.99$. ♦

მაგალითი 2. 4. ალბათობა იმისა, რომ მზა ხელსაწყოთა პარტიიდან შემთხვევით შერჩეული ხელსაწყო საჭიროებს დამატებით რეგულირებას 0,05-ის ტოლია. თუ შემოწმებისას აღმოჩნდა, რომ შერჩეულ ხელსაწყოთა 6%-ზე მეტი საჭიროებს დამატებით რეგულირებას, მაშინ თვლიან, რომ სტანდარტს არ შეესაბამება მთელი პარტია. იპოვეთ ალბათობა იმისა, რომ ხელსაწლოთა პარტია არ შეესაბამება სტანდარტს, თუ ამისათვის უნდა შეამოწმონ 500 ხელსაწყო .

ამოხსნა. ვისარგებლოდ ბერნულის ცდათა სქემით. ჩავთვალოთ, რომ ადგილი აქვს „წარმატებას", თუ ხელსაწყო საჭიროებს დამატებით რეგულირებას. ე.ი. პირობის თანახმად n=500, p=0,05, np=25, npq=23,8, $(npq)^{1/2}$=4,87 . (2.13)-ის ძალით

$$P\{30 \le S_n \le 500\} \approx \Phi\left(\frac{500-25}{\sqrt{23{,}8}}\right) - \Phi\left(\frac{30-25}{\sqrt{23{,}8}}\right) = \Phi(97)\text{-}\Phi(1{,}03)=1\text{-}0{,}803+0{,}192. \quad \blacklozenge$$

შენიშვნა : ბერნულის ცდათა სქემაში, როდესაც წარმატების ალბათობა p ახლოსაა ნულთან და np დიდი არაა, ნორმალურ მიახლოებას ამჯობინებენ *პუასონის მიახლოებას :*

$$P\{S_n \le x\} \approx \sum_{k=0}^{x} \frac{(np)^k}{k!} e^{-np}. \qquad (2.14)$$

ამ თანადობის მარჯვენა მხარეს მოთავსებული გამოსახულება წარმოადგენს $\lambda=np$ პარამეტრის მქონე პუასონის განაწილების ფუნქციას.

კითხვები თვითშემოწმებისათვის

- ჩამოაყალიბეთ და გამოიყვანეთ ჩებიშევის უტოლობა
- ჩამოაყალიბეთ და დაამტკიცეთ დიდ რიცხვთა კანონი.
- ჩამოაყალიბეთ დიდ რიცხვთა კანონი ბერნულის ცდათა სქემისათვის.
- ჩამოაყალიბეთ ცენტრალური ზღვარითი თეორემა.
- ჩამოაყალიბეთ ცენტრალური ზღვარითი თორემა ბერნულის ცდათა სქემისათვის
- ჩამოაყალიბეთ ცენტრალური ზღვარითი თეორემა პროპორციებისათვის მუავრ-ლაპლასის თეორემა..

.

§ 4. EXCEL

განვიხილოთ **Excel**-ის შემდეგი სტატისტიკური ფუნქციები.

NORMDIST, NORMSDIST, NORMINV, NORMSINV, EXPONDIST, CHIDIST, CHIINV, TDIST, TINV, FDIST,FINV,

რომელთა ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა მოცემულია 1.3 ცხრილში

აღნიშვნა	შინაარსი
x	განაწილების სიმკვრივის ან განაწილების ფუნქციის არგუმენტის რიცხვითი მნიშვნელობა
Mean	განაწილების მათემატიკური ლოდინი
Stanard deviation	განაწილების საშუალო კვადრატული გადახრა
Cumullative	თუ **Cumullative** პარამეტრის რიცხვითი მნიშვნელობა ნულის ტოლია, სტატისტიკური ფუნქცია გვაძლევს განაწილების სიმკვრივის რიცხვით მნიშვნელობას. თუ **Cumullative** პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, სტატისტიკური ფუნქცია გვაძლევს განაწილების ფუნქციის რიცხვით მნიშვნელობას.
Lambda	λ პარამეტრი
Probability	ალბათობის *P* დონე, რომლისთვისაც საფუძველზე გამოითვლება კვანტილი.
Deg freedom	განაწილების თავისუფლების ხარისხი
Probability	**P{X>x}** ალბათობის მნიშვნელობა
Tails	თუ **Tails** პარამეტრის რიცხვითი მნიშვნელობა ერთის ტოლია, **TDIST** ფუნქცია გვაძლევს ცალმხრივ **T(*n*)>*x*** არეში მოხვედრის ალბათობას, **P{T(*n*)>*x*}**-ს, სტიუდენტის განაწილებისათვის . თუ **Tails** პარამეტრის რიცხვითი მნიშვნელობა ორის ტოლია, **TDIST** ფუნქცია გვაძლევს ორმხრივ **|T(*n*)|>*x*** ინტრვალში მოხვედრის, ალბათობას, **P{T(*n*)>*x*}**-ს, სტიუდენტის განაწილებისათვის .

NORMSDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

NORMSDIST (*z*) .

NORMSDIST ფუნქციით გამოიანგარიშება სტანდარტული ნორმალურად განაწილებული X შემთხვევითი სიდიდისათვის $P(X \le z)$ ალბათობა ფორმულით $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$.

მაგალითი.

გამოვთვალოთ $X \sim N(0,1)$-სათვის შემდეგი ალბათობა. $P(X \le 1.37)$.

ამოხსნა.

მოცემულია- სტანდარტული ნორმალური განაწილების მქონე შემთხვევითი სიდიდე.

გამოვიყენოთ ფუნქცია **NORMSDIST,** რომელიც ასე ჩაიწერება:

NORMSDIST(1.37)=0,91.

ამოხსნის შედეგად მივიღებთ, რომ საძიებელი ხდომილობის ალბათობა არის 0.91.

NORMDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

NORMDIST (*x*, Mean, Stanard deviation, Cumullative).

NORMDIST ფუნქციის მნიშვნელობა, როცა Cumullative =0, გამოიანგარიშება ფორმულით

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

NORMDIST ფუნქციის მნიშვნელობა, როცა Cumullative =1, გამოიანგარიშება ფორმულით

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du.$$

მაგალითი.

$X \sim N(3,4)$ შემთხვევითი სიდიდესათვის გამოთვალეთ შემდეგი ალბათობა $P(X \leq 1.37)$.

ამოხსნა.

მოცემულია- ნორმალურად განაწილებული შემთხვევითი სიდიდის:

საშუალო მნიშვნელობა, **Mean**= 3;

საშუალო კვადრატული გადახრა, **Stanard deviation**=4;

Cumullative=1.

ამ შემთხვევაში **NORMDIST** ფუნქცია ასე ჩაიწერება:

NORMDIST (*x*, Mean, Stanard deviation, Cumullative)=NORMDIST(1.37,3,4,1).

ამოხსნის შედეგად მივიღებთ, რომ საძიებელი ხდომილობის ალბათობა არის 0.3418. ♦

მაგალითი.

ცნობილია, რომ საკონდიტრო ქარხნის მიერ გამოშვებული ფილა შოკოლადის წონა ნორმალურადაა განაწილებული 400 გრ-ის ტოლი საშუალოთი და 20 გრ-ის ტოლი საშუალო კვადრატული გადახრით. რას უდრის ალბათობა იმისა, რომ მაღაზიაში ნაყიდი ფილა შოკოლადის წონა არ აღემატება 410 გრ-ს.

ამოხსნა.

მოცემულია- ნორმალურად განაწილებული შემთხვევითი სიდიდის:

საშუალო მნიშვნელობა, **Mean**= 400;

საშუალო კვადრატული გადახრა, **Stanard deviation**=20;

Cumullative=1.

ამ შემთხვევაში **NORMDIST** ფუნქცია ასე ჩაიწერება:

NORMDIST (*x*, Mean, Stanard deviation, Cumullative)=NORMDIST(410,400,20,1)

ამოხსნის შედეგად მივიღებთ, რომ 69% შემთხვევაში მაღაზიაში ნაყიდი ფილა შოკოლადის წონა არ აღემატება 410 გრ-ს. ♦

NORMSINV ფუნქციის ჩაწერის სინტაქსი ასეთია:

NORMSINV (Probability),

ფუნქცია **NORMSINV** გვიბრუნებს სტანდარტული ნორმალური განაწილების p დონის კვანტილის რიცხვით მნიშვნელობას.

მაგალითი.

$X \sim N(0,1)$, იპოვეთ p=0.42 დონის კვანტილი.

ამოხსნა.

მოცემულია- სტანდარტული ნორმალური განაწილების მქონე შემთხვევითი სიდიდე.

ამ შემთხვევაში ფუნქცია **NORMSINV** ასე ჩაიწერება:

NORMSINV (Probability)=NORMSINV (0.42)

ამოხსნის შედეგად მივიღებთ, რომ სტანდარტული ნორმალური განაწილების მქონე შემთხვევითი სიდიდისათვის p=0.42 დონის კვანტილი 0.2-ის ტოლია. ♦

NORMINV ფუნქციის ჩაწერის სინტაქსი ასეთია:

NORMINV (Probability, Mean, Stanard deviation).

ფუნქცია **NORMINV** გვიბრუნებს არასტანდარტული ნორმალური განაწილების p დონის კვანტილის რიცხვით მნიშვნელობას.

მაგალითი.

$X \sim N(7,5)$, იპოვეთ $P=0.75$ დონის კვანტილის მნიშვნელობა.

ამოხსნა.

მოცემულია- ნორმალურად განაწილებული შემთხვევითი სიდიდის:

საშუალო მნიშვნელობა, Mean= 7;

საშუალო კვადრატული გადახრა, Stanard deviation=5;

ამ შემთხვევაში **NORMINV** ფუნქცია ასე ჩაიწერება:

NORMINV (Probability, Mean, Stanard deviation)=NORMINV (0.75,7,5).

ამოხსნის შედეგად მივიღებთ, რომ $X \sim N(7,5)$-სათვის $p=0.8$ დონის კვანტილი 10,37.-ის ტოლია.

♦

EXPONDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

EXPONDIST (*x*, Lambda, Cumullative).

EXPONDIST ფუნქციის მნიშვნელობა, როცა Cumullative =0, გამოიანგარიშება ფორმულით $\lambda e^{-\lambda x}$,

EXPONDIST ფუნქციის მნიშვნელობა, როცა Cumullative =1, გამოიანგარიშება ფორმულით $1-e^{-\lambda x}$.

მაგალითი.

$X \sim Exp(0.2)$. გამოთვალეთ შემდეგი ალბათობა $P(X \leq 5)$.

ამოხსნა.

მოცემულია- ექსპონენციალურად განაწილებული შემთხვევითი სიდიდის:

ინტენსივობა Lambda = 0.2.

ამ შემთხვევაში **EXPONDIST** ფუნქცია ასე ჩაიწერება:

EXPONDIST (*x*, Lambda, Cumullative)=NORMINV (5,0.2,1).

ამოხსნის შედეგად მივიღებთ, რომ $X \sim Exp(0.2)$-სათვის $P(X \leq 5)=0.63$. ♦

CHIDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

CHIDIST (*x*, Deg freedom).

CHIDIST ფუნქცია ანგარიშობს $\{X>x\}$ ხდომილობის ალბათობას $\chi^2(k)$-განაწილების მქონე შემთხვევითი სიდიდისათვის

მაგალითი.

ვთქვათ, $X \sim \chi^2(8)$. გამოთვალეთ შემდეგი ალბათობა $P(X>5)$.

ამოხსნა.

მოცემულია: "ხი კვადრატ"-განაწილების მქონე შემთხვევითი სიდიდის თავისუფლების ხარისხი, Deg freedom=8.

ამ შემთხვევაში **CHIDIST** ფუნქცია ასე ჩაიწერება:

CHIDIST (*x*, Deg freedom)= CHIDIST (5,8).

ამოხსნის შედეგად მივიღებთ, რომ $X \sim \chi^2(8)$.)-სათვის $P(X>5)=0.76$. ♦

CHIINV ფუნქციის ჩაწერის სინტაქსი ასეთია:

CHIINV(Probability, Deg freedom).

დავუშვათ, α= **Probability, k= Deg freedom.**

CHIINV ფუნქცია გვიბრუნებს *a*-ს იმ მნიშვნელობას, რომლისთვისაც $\chi^2(k)$-განაწილების მქონე შემთხვევითი სიდიდისათვის $P\{\chi^2(k)>a\}=\alpha$.

მაგალითი.

ვთქვათ, $X\sim\chi^2(13)$. იპოვეთ *a*-ს ის მნიშვნელობა, რომლისთვისაც $P\{X>a\}=0.25$.

ამოხსნა.

მოცემულია: "ხი კვადრატ"-განაწილების მქონე შემთხვევითი სიდიდის თავისუფლების ხარისხი, Deg freedom=13.

ზედა არის ალბათობის რიცხვითი მნიშვნელობა, Probability=0.25.

ამ შემთხვევაში **CHIINV** ფუნქცია ასე ჩაიწერება:

CHIINV(Probability, Deg freedom)=CHIINV (0.25,13).

ამოხსნის შედეგად მივიღებთ, რომ "ხი კვადრატ"-განაწილების მქონე შემთხვევითი სიდიდისათვის, რომლის თავისუფლების ხარისხი 13-ის ტოლია, ზედა 0.25 კრიტიკული წერტილის რიცხვითი მნიშვნელობაა 15,98. ♦

TDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

TDIST (*x*, Deg freedom, Tails).

TDIST(x,k,1) ფუნქცია ანგარიშობს *t*(*k*) განაწილების მქონე შემთხვევითი სიდიდისათვის ცალმხრივ **T(*k*)>*x*** არეში ჩავარდნის **P{T(*k*)>*x*}** ალბათობას,

TDIST(x,k,2) ფუნქცია ანგარიშობს *t*(*k*) განაწილების მქონე შემთხვევითი სიდიდისათვის ორმხრივ **|T(*k*)|>*x*** არეში ჩავარდნის **P{|T(*k*)|>*x*}** ალბათობას.

მაგალითი.

ვთქვათ, $X\sim\chi^2(8)$. გამოთვალეთ შემდეგი ალბათობა $P(X>5)$.

ამოხსნა.

მოცემულია: "ხი კვადრატ"-განაწილების მქონე შემთხვევითი სიდიდის თავისუფლების ხარისხი, Deg freedom=8.

მითითებულია, რომ არე ცალმხრივია ($P(X>5)$.

ამ შემთხვევაში **TDIST** ფუნქცია ასე ჩაიწერება:

TDIST (*x*, Deg freedom, Tails)=TDIST (5,8).

ამოხსნის შედეგად მივიღებთ, რომ $X\sim\chi^2(8)$-სათვის $P(X>5)=0.76$. ♦

TINV ფუნქციის ჩაწერის სინტაქსი ასეთია:

TINV (Probability, Deg freedom).

დავუშვათ, **α= Probability, k= Deg freedom.**

TINV ფუნქცია გვიბრუნებს *a*-ს იმ მნიშვნელობას, რომლისთვისაც t(*k*)-განაწილების მქონე *X* შემთხვევითი სიდიდისათვის $P\{X> a\}=\alpha$.

მაგალითი.

ვთქვათ, $X\sim t(13)$. იპოვეთ *a*-ს ის მნიშვნელობა, რომლისთვისაც $P\{X> a\}=0.25$.

ამოხსნა.

მოცემულია: *t*-განაწილების მქონე შემთხვევითი სიდიდის

თავისუფლების ხარისხი, Deg freedom=13.

ამ შემთხვევაში **TINV** ფუნქცია ასე ჩაიწერება:

TINV (Probability, Deg freedom)=TINV (0.25,13).

ამოხსნის შედეგად მივიღებთ, რომ *t*-განაწილების მქონე შემთხვევითი სიდიდისათვის, რომლის თავისუფლების ხარისხი 13-ის ტოლია, ზედა 0.25 კრიტიკული წერტილია 1.20. ♦

FDIST ფუნქციის ჩაწერის სინტაქსი ასეთია:

FDIST (*x*, Deg freedom, Deg freedom).

FDIST ფუნქცია ანგარიშობს ფიშერის განაწილების მქონე *X* შემთხვევითი სიდიდისათვის P{*X* >a} ალბათობას..

მაგალითი.

ვთქვათ, *X*~*F*(14,8). გამოთვალეთ შემდეგი ალბათობის *P*(*X*>2.1).

ამოხსნა.

მოცემულია: *F*-განაწილების მქონე შემთხვევითი სიდიდის:

თავისუფლების ხარისხი Deg freedom=14, თავისუფლების ხარისხი Deg freedom=8

ამ შემთხვევაში **FDIST** ფუნქცია ასე ჩაიწერება:

FDIST (*x*, Deg freedom, Deg freedom)=FDIST(2.1,14,8).

ამოხსნის შედეგად მივიღებთ, რომ *F*-განაწილების მქონე შემთხვევითი სიდიდისათვის, რომლის თავისუფლების ხარისხებია 14 და 8, საძიებელი ალბათობა 0.13 -ს ტოლია.. ♦

FINV ფუნქციის ჩაწერის სინტაქსი ასეთია:

FINV (Probability, Deg freedom, Deg freedom).

დავუშვათ, **α= Probability, k_1= Deg freedom, k_2= Deg freedom.**

FINV ფუნქცია გვიბრუნებს *a*-ს იმ მნიშვნელობას, რომლისთვისაც $F(k_1,k_2)$-განაწილების მქონე *X* შემთხვევითი სიდიდისათვის P{*X*> *a* }=α.

მაგალითი.

ვთქვათ, *X*~*F*(13,17). იპოვეთ *a*-ს ის მნიშვნელობა, რომლისთვისაც P{X> *a* }=0.1.

ამოხსნა.

მოცემულია: *F*-განაწილების მქონე შემთხვევითი სიდიდის

თავისუფლების ხარისხი Deg freedom=13, თავისუფლების ხარისხი Deg freedom=17.

ამ შემთხვევაში **FINV** ფუნქცია ასე ჩაიწერება:

FINV (Probability, Deg freedom, Deg freedom)=FINV (0.1,13,17).

ამოხსნის შედეგად მივიღებთ, რომ *F*-განაწილების მქონე შემთხვევითი სიდიდისათვის, რომლის თავისუფლების ხარისხებია 13 და 17, ზედა 0.1 კრიტიკული წერტილია 1.94. ♦

სავარჯიშოები

2.1. რას უდრის ნორმალურად განაწილებული შემთხვევითი სიდიდის მათემატიკური ლოდინი და დისპერსია, თუ განაწილების სიმკვრივეს აქვს შემდეგი სახე:

$$f_{\mu,\sigma}(x)=\frac{1}{\sqrt{8\pi}}e^{-\frac{(x-3)^2}{8}}$$

2.2. რას უდრის ნორმალურად განაწილებული შემთხვევითი სიდიდის მათემატიკური ლოდინი და დისპერსია, თუ განაწილების სიმკვრივეს აქვს შემდეგი სახე:

$$f_{\mu,\sigma}(x)=\frac{1}{\sqrt{5\pi}}e^{-\frac{x^2}{5}}$$

2.3. ამოწერეთ ნორმალურად განაწილებული შემთხვევითი სიდიდის განაწილების სიმკვრივისა და ფუნქციის გამოსახულებები, თუ მათემატიკური ლოდინი -4-ის, ხოლო დისპერსია 5-ის ტოლია.

2.4. ამოწერეთ ნორმალურად განაწილებული შემთხვევითი სიდიდის განაწილების სიმკვრივისა და ფუნქციის გამოსახულებები, თუ მათემატიკური ლოდინი –3-ის, ხოლო საშუალო კვადრატული გადახრა 2-ის ტოლია.

2.5. ვთქვათ, $X\sim N(0,1)$; ცხრილების გამოყენებით გამოთვალეთ შემდეგი ალბათობები:

ა) $P\{X<1\}$, ბ) $P\{X<-0.7\}$, გ) $P\{X<1.96\}$, დ) $P\{X<-1.6\}$, ე) $P\{X<2.33\}$, ვ) $P\{0.3<X<2.9\}$.

2.6. ვთქვათ, $X\sim N(0,1)$,

ა) $P\{X<\beta\}=0.8$, $P\{X>\alpha\}=0.7$. შეადარეთ ერთმანეთს შემდეგი რიცხვები: $|a|$ და $|\beta|$.
ბ) $P\{X<\alpha\}=0.4$, $P\{X>\beta\}=0.6$. შეადარეთ ერთმანეთს α და β რიცხვები.
გ) $P\{X>\alpha\}=0.8$, $P\{X>\beta\}=0.4$. რას უდრის $P\{\alpha<X<\beta\}$?

2.7. $X\sim N(0,1)$, იპოვეთ $p=0.42$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.8. $X\sim N(0,1)$, იპოვეთ $p=0.38$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.9. $X\sim N(0,1)$, იპოვეთ $p=0.85$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.10. $X\sim N(0,1)$, იპოვეთ $p=0.75$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.11. X შემთხვევითი სიდიდე $N(3,4)$ კანონითაა განაწილებული. გამოთვალეთ შემდეგი ხდომილობის ალბათობები ა) $P(X\le 1)$, ბ) $P(X>6)$.

2.12. X შემთხვევითი სიდიდე $N(5,16)$ კანონითაა განაწილებული. ვიპოვოთ ისეთი α და β რიცხვები, რომლებისათვისაც ა) $P\{X\le\beta\}=0.7$, ბ) $P\{X>\alpha\}=0.4$, გ) $P\{\alpha<X\le\beta\}=0.8$.

2.13. $N(8,7)$, განსაზღვრეთ b, რომლისთვისაც $P\{5<X<b\}=0.4$.

2.14. $N(13,8)$, განსაზღვრეთ b, რომლისთვისაც $P\{9<X<b\}=□□0.3$.

2.15. $N(10,9)$, განსაზღვრეთ b, რომლისთვისაც $P\{7<X<b\}=□□0.4$.

2.16. $N(4,6)$, განსაზღვრეთ b, რომლისთვისაც $P\{1<X<b\}=\square\square 0.4$.

2.17. $X\sim N(5,5)$. იპოვეთ მათემატიკური ლოდინის მიმართ ისეთი სიმეტრიული (α,β) ინტერვალი, რომლისთვისაც $P\{\alpha<X<\beta\}=0.8$.

2.18. $X\sim N(8,6)$. იპოვეთ ისეთი ცალხმრივი მარჯვენა $(a,+\infty)$ ინტერვალი, რომელშიც X შემთხვევითი სიდიდის მოხვედრის ალბათობა 0.2-ის ტოლია.

2.19. $X\sim N(13,7)$. იპოვეთ ისეთი ცალხმრივი მარცხენა $(-\infty,a)$ ინტერვალი, რომელშიც X შემთხვევითი სიდიდის მოხვედრის ალბათობა 0.1-ის ტოლია.

2.20. $X\sim N(0,1)$, იპოვეთ $p=0.95$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.21. $X\sim N(4,8)$, $Y\sim N(13,6)$. როგორია $X-Y$ შემთხვევითი სიდიდის განაწილების კანონი?

2.22. $X\sim N(34,7)$. იპოვეთ იმის ალბათობა, რომ X შემთხვევითი სიდიდე მიიღებს მნიშვნელობას (20, 50) ინტერვალიდან.

2.23. $X\sim N(20.6)$. იპოვეთ იმის ალბათობა, რომ X შემთხვევითი სიდიდის გადახრა მათემატიკური ლოდინისაგან იქნება 3-ზე ნაკლები.

2.24. დადებითია თუ უარყოფითი $N(0,1)$ განაწილების $P=0.7$ დონის $U_{0.7}$ კვანტილის რიცხვითი მნიშვნელობა?

2.25. $N(0,1)$ განაწილებისათვის ვიპოვოთ $p=0.4$ დონის კვანტილის რიცხვითი მნიშვნელობა.

2.26. X~N(7;3), Y~N(5;4), შემდეგი ალბათ ობებიდან $P\{X>5\}$, $P\{Y>7\}$რომელია მეტი?

2.27. X და Y დამოუკიდებელი ნორმალურად განაწილებული შემთხვევითი სიდიდეებია. ამასთან EX=-6 DX=9, EY=-4, DY=4. იპოვეთ ალბათობა იმისა, რომ $Z=X+2Y$ შემთხვევითი სიდიდე ოთხი დამოუკიდებელი დაკვირვების შედეგად ორჯერ მიიღებს მნიშვნელობას (-17; -12) ინტერვალიდან.

2.28. X და Y დამოუკიდებელი ნორმალურად განაწილებული შემთხვევითი სიდიდეებია. ამასთან EX= 5 DX=9, EY= -7, DY=4. იპოვეთ ალბათობა იმისა, რომ $Z=X-2Y$ შემთხვევითი სიდიდე ხუთი დამოუკიდებელი დაკვირვების შედეგად სამჯერ მიიღებს მნიშვნელობას (16; 22) ინტერვალიდან.

2.29. დაწერეთ ნორმალურად განაწილებული იმ შემთხვევითი სიდიდის განაწილების სიმკვრივე, რომლის მათემატიკური ლოდინი და დისპერსია ტოლია (-1;2) ინტერვალში თანაბრად განაწილებული შემთხვევითი სიდიდის შესაბამისი მნიშვნელობებისა.

2.30. მაღაზია ვაჭრობს მამაკაცის კოსტუმებით. ცნობილია, რომ მოთხოვნა მამაკაცის კოსტუმებზე ნორმალურად განაწილებული შემთხვევითი სიდიდეა 52-ის ტოლი საშუალოთი და 3-ის ტოლი საშუალო კვადრატული გადახრით. რისი ტოლია მოთხოვნის პროცენტული რაოდენობა 48 ზომის კოსტუმზე, თუ ამ ზომის კოსტუმებს განსაზღვრავს (47-49) ინტერვალი.

2.31. მაღაზია ვაჭრობს მამაკაცის კოსტუმებით. ცნობილია, რომ მოთხოვნა მამაკაცის კოსტუმებზე ნორმალურად განაწილებული შემთხვევითი სიდიდეა 52-ის ტოლი საშუალოთი და 3-ის ტოლი საშუალო კვადრატული გადახრით. რისი ტოლია მოთხოვნის პროცენტული რაოდენობა 50 ზომის კოსტუმზე, თუ ამ ზომის კოსტუმებს განსაზღვრავს (49-51) ინტერვალი.

2.32. X ~t(26). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X>a}=0. 975

2.33. X ~t(17). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X<a}=0.9

2.34. X ~t(15). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X<a}=0.05

2.35. X ~t(13). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X>a}=0.99

2.36. X ~χ^2(27). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X<a}=0.975

2.37. X ~χ^2(17). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X<a}=0.99

2.38. X ~χ^2(17). იპოვეთ a-ს ის მნიშვნელობა, რომლისთვისაც P{X>a}=0.95

2.39. X ნორმალური შემთხვევითი სიდიდის მათემატიკური ლოდინი და საშუალო კვადრატული გადახრა შესაბამისად ტოლია 10 და 2-ის. ვიპოვოთ: ა) $P\{12\le X\le 14\}$, ბ) $P=1/3$ დონის კვანტილი;

2.40. X ნორმალური შემთხვევითი სიდიდის მათემატიკური ლოდინი და საშუალო კვადრატული გადახრა შესაბამისად ტოლია 20 და 5-ის. ვიპოვოთ: ა) $P\{15\le X\le 25\}$, ბ) $P=1/4$ დონის კვანტილი;

2.41. X ნორმალური შემთხვევითი სიდიდის მათემატიკური ლოდინი $\mu=10$, ხოლო $P\{10\le X\le 20\}=0.3$. ვიპოვოთ: $P\{0\le X\le 10\}$

2.42. X ნორმალური შემთხვევითი სიდიდის მათემატიკური ლოდინი $\mu=25$, ხოლო $P\{10\le X\le 15\}=0.2$. ვიპოვოთ: $P\{35\le X\le 40\}$.

2.43. ვთქვათ $X\sim N(\mu,\sigma^2)$. ვიპოვოთ X ნორმალური შემთხვევითი სიდიდის $p=0.7$ დონის კვანტილი, თუ $\mu=2$, $\sigma=1$.

ლექცია IV მონაცემთა პირველადი დამუშავება

ისტორიულად ჩამოყალიბდა **სტატისტიკის** შემდეგი განსაზღვრა: ეს არის მათემატიკის დარგი, რომელიც შეისწავლის სტატისტიკური მონაცემების სისტემატიზაციის, ანალიზისა და გამოყენების მეთოდებს თეორიული და პრაქტიკული დასკვნების მიღების მიზნით. სიტყვას სტატისტიკა ("statistics ") საფუძვლად უდევს ლათინური "statista" - სახელმწიფო მოღვაწე ან "stato" - სახელმწიფო. ამ სიტყვის ქვეშ თავდაპირველად იგულისხმებოდა სახელმწიფო მოღვაწეობის რაოდენობრივი ხასიათის მონაცემების აღწერა; მოსახლეობის, აღებული მოსავლის, სახელმწიფო ხაზინის დანახარჯების აღწერა და სხვა. მონაცემთა ანალიზისა და გამოყენებითი სტატისტიკის თანამედროვე მეთოდები ემყარება და არსებითად გამოიყენებს დისკრეტული ალბათობის ფაქტებსა და ცნებებს.

როგორც დისკრეტული ალბათობა, ისე მონაცემთა ანალიზი და სტატისტიკა შეისწავლის მასობრივ მოვლენებში შემთხვევით შედეგთა (ხდომილობათა) რაოდენობრივი ხასიათის კანონზომიერებებს.

მოვლენას ეწოდება მასობრივი, თუ ამ მოვლენის დაკვირვება ერთსა და იგივე პირობებში ბევრჯერ შეიძლება იქნას განხორციელებული. შემთხვევით ხდომილობათა ქვეშ კი იგულისხმება ისეთი მასობრივი მოვლენა, რომლის დამახასიათებელი რაოდენობრივი (რიცხვითი) მნიშვნელობების წინასწარ განსაზღვრა შეუძლებელია. ამასთან დაკავშირებით განვსაზღროთ ე.წ. შემთხვევითი ექსპერიმენტი.

შემთხვევითი ექსპერიმენტი[1] (ცდა) ეწოდება პირობათა რაიმე გარკვეული S კომპლექსის განხორციელებას, რომლის დროსაც იგულისხმება, რომ:

1. ექსპერიმენტის ჩატარებამდე შესაძლებელია ყველა შესაძლო ურთიერთგამომრიცხავი შედეგის მითითება.
2. ექსპერიმენტის ჩატარებამდე შეუძლებელია მისი შესაძლო შედეგის ცალსახად განსაზღვრა.
3. შესაძლებელია ექსპერიმენტის მრავალჯერადი განმეორება.

პირველ თავში ჩვენ განვიხილავთ შემთხვევითი ექსპერიმენტის მრავალჯერადი გამეორების შედეგად ექსპერიმენტის დამახასიათებელი დაკვირვებული (რიცხვითი) მნიშვნელობების, ანუ მონაცემთა პირველადი დამუშავების (მონაცემთა ანალიზის) მეთოდებს, ხოლო მომდევნო თავებში გავეცნობით გამოყენებითი სტატისტიკის საფუძვლებს .

§ 1. მონაცემების სიხშირეთა განაწილება და მისი გრაფიკული წარმოდგენა

მონაცემების ქვეშ იგულისხმება ობიექტთა რაიმე სიმრავლის რაოდენობრივ ან თვისებრივ მახასიათებელთა დაკვირვებული მნიშვნელობების ერთობლიობა. მონაცემი შეიძლება იყოს ფიზიკური გაზომვის შედეგი (მაგალითად, წონა ან სიმაღლე და ა.შ.) პასუხი კითხვაზე (კი ან არა) ან რაიმე ნიშნის მიხედვით კლასიფიკაციის შედეგი.

დაკვირვებული რაოდენობრივი მახასიათებლების მიხედვით შეიძლება განვასხვავოთ **დისკრეტული და უწყვეტი მონაცემები**. დისკრეტული მონაცემები ღებულობს მხოლოდ ცალკე-

[1] ზედსართავი „შემთხვევითი“ ქვემოთ ხშირად გამოტოვებული იქნება.

ულ იზოლირებულ მნიშვნელობებს, ხოლო უწყვეტ მონაცემებს შეუძლია მიიღოს ნებისმიერი (ნამდვილი) რიცხვითი მნიშვნელობა რაიმე საზღვრებში. დისკრეტული მონაცემები ჩვეულებრივ თვლის შედეგად მიიღება. ასეთებია, მაგალითად, ოჯახის წევრთა რაოდენობა, მსხვილი ფირმების შვილობილი ფირმების რაოდენობა, ოთახების რაოდენობა ბინებში და ა.შ. ასეთ მონაცემებს **დისკრეტული ეწოდება**. დისკრეტული მონაცემების ალტერნატივა უწყვეტი მონაცემებია, როგორიცაა წონა, სიმაღლე, სიცოცხლის ხანგრძლივობა, შემოსავალი და სხვა. მაგალითად, წონა ბუნებითაა უწყვეტი მახასიათებელი, თორემ მოცემული სიზუსტით გაზომვისას ისიც კი შეიძლება დისკრეტულად ჩაითვალოს. ამდენად გაზომვის მოცემული საზუსტისათვის ალტერნატივა – დისკრეტული და უწყვეტი – პირობითია.

ქვემოთ ჩვენ მოვიყვანთ მონაცემთა თვალსაჩინო, კერძოდ ცხრილური და გრაფიკული წარმოდგენის მეთოდებს, რომლებსაც გავეცნობით მაგალითების საფუძველზე

მაგალითი 4.1. მსხვილი კომპანიის კადრების განყოფილების უფროსს დაევალა ჩაეტარებინა თანამშრომელთა სამსახურში არ გამოცხადების ანალიზი. ქვევით მოცემულია უკანასკნელი 21 სამუშაო დღის განმავლობაში სამსახურში არ გამოცხადებულ თანამშრომლებთა რაოდენობები.

1, 1, 2, 2, 1, 0, 2, 3, 2, 2, 1, 1, 4, 3, 1, 0, 2, 1, 4, 1, 3

ასეთ მონაცემებს დისკრეტულ მონაცემებს უწოდებენ, ვინაიდან დაკვირვებული მნიშვნელობები (სამსახურში არ გამოცხადებულ თანამშრომელთა რაოდენობა) ღებულობს მხოლოდ მთელ რიცხვით მნიშვნელობებს.

მოცემულ რიცხვთა ერთობლიობა წარმოადგენს n=21 მოცულობის შერჩევას. **შერჩევის მოცულობა** დაკვირვებათა რაოდენობის ტოლია. ასეთი სახით ჩაწერილი მონაცემების გამოყენება შემდგომი ანალიზისათვის მეტად ძნელია. ამიტომ, უპირველეს ყოვლისა, ჩვენ გავეცნობით მონაცემთა დამუშავების და წარმოდგენის შემდეგ მარტივ ხერხებს:

ა) გადავალაგოთ დაკვირვებული მნიშვნელობები ზრდის მიხედვით.

0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4

მიღებულ მიმდევრობას ვარიაციული მწკრივი ეწოდება.

ვარიაციული მწკრივიდან ჩვენ ადვილად შეგვიძლია ამოვწეროთ მასში შემავალი განსხვავებული რიცხვითი მნიშვნელობები, ანუ ვარიანტები და ამ მნიშვნელობათა სიხშირეები, ანუ ვარინტების გამეორებათა რაოდენობა. მოცემულ ვარიაციულ მწკრივში 5 განსხვავებული რიცხვითი მნიშვნელობაა. ეს მნიშვნელობებია 0,1,2,3,4. უკანასკნელი 21 სამუშაო დღის განმავლობაში მხოლოდ ორჯერ გამოცხადდა ყველა თანამშრომელი, რვაჯერ არ გამოცხადდა მხოლოდ ერთი და ა.შ. ყველა ამ ინფორმაციას თავი მოვუყაროთ 4.1 სახის ცხრილში

ცხრილი 4.1

არ გამოცხადებულთა რაოდენობა (ვარიანტი)	0	1	2	3	4	დაკვირვებული დღეების რაოდენობა
გაცდენილი დღეების რაოდენობა (სიხშირე)	2	8	6	3	2	21

ცხრილის პირველ სტრიქონში მითითებულია სამსახურში არ გამოცხადებულ თანამშრომელთა რაოდენობები, ხოლო მეორე სტრიქონში თანამშრომელთა აღნიშნული

რაოდენობის მიერ გაცდენილ დღეთა რაოდენობა ანუ სიხშირე. ცხრილი 4.1 გვაძლევს სამსახურში არ გამოცხადებულ თანამშრომელთა სიხშირული განაწილების სურათს.

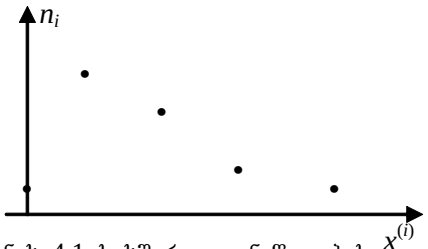
თუ გადავალთ ფარდობით სიხშირეებზე $\left(\text{ფარდობითი სიხშირე} = \frac{\text{სიხშირე}}{\text{შერჩევის მოცულობა}} \right)$,

მაშინ ცხრილი 4.1-ის მაგიერ გვექნება **ფარდობით სიხშირეთა** შემდეგი განაწილება

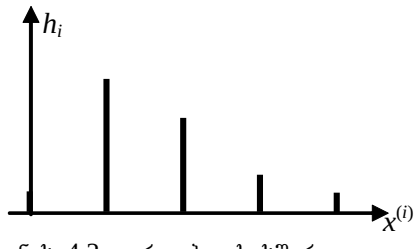
ცხრილი 4.2

ვარიანტები	0	1	2	3	4	ჯამი
ფარდობითი სიხშირე	$\frac{2}{21}$	$\frac{8}{21}$	$\frac{6}{21}$	$\frac{3}{21}$	$\frac{2}{21}$	1

4.1 და 4.2 ცხრილები შეიძლება **წერტილოვანი დიაგრამის** (ნახ. 4.1), ან **მესერული დიაგრამის** (ნახ. 4.2) სახით გამოისახოს

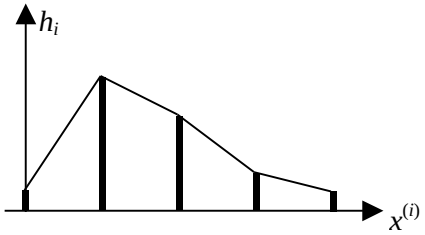


ნახ. 4.1. სიხშირეთა განაწილების წერტილოვანი დიაგრამა



ნახ. 4.2. ფარდობით სიხშირეთა განაწილების მესერული დიაგრამა

ხშირ შემთხვევაში განიხილავენ ე.წ. **პოლიგონსაც**. დისკრეტული მონაცემებისათვის პოლიგონი მიიღება წერტილოვანი დიაგრამის წერტილების შეერთებით. მას შემდეგი სახე აქვს (იხ. ნახ. 4.3).



ნახ. 4.3. პოლიგონი

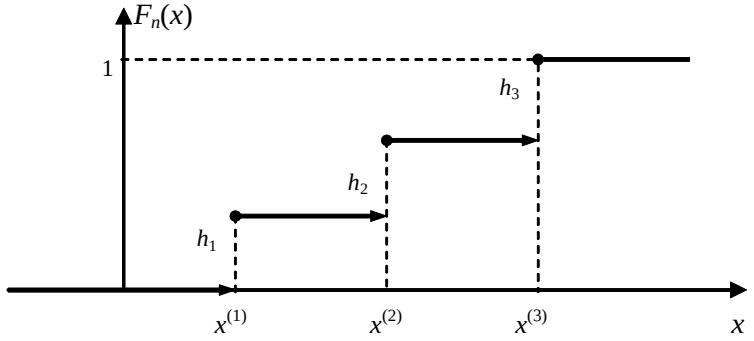
განვსაზღვროთ ე.წ. **დაგროვილი ფარდობითი სიხშირის ფუნქცია**; ცხრილი 4.2-ით განსაზღვრული ვარიანტის მოცემული მნიშვნელობისთვის **დაგროვილი (ფარდობითი) სიხშირე** განისაზღვრება როგორც იმ ვარიანტების (ფარდობითი) სიხშირეების ჯამი, რომლებიც არ აღემატება ვარიანტის მოცემულ მნიშვნელობას (იხ. *ცხრილი 4.3*)

ცხრილი 1.3ვარიანტები	0	1	2	3	4
დაგროვილი ფარდობითი სიხშირე	$\frac{2}{21}$	$\frac{2}{21}+\frac{8}{21}$	$\frac{2}{21}+\frac{8}{21}+\frac{6}{21}$	$\frac{2}{21}+\frac{8}{21}+\frac{6}{21}=\frac{3}{21}$	1

თუ $x^{(1)},\ldots,x^{(s)}$ -ით აღვნიშნავთ ზრდადობით დალაგებულ განსხვავებულ დაკვირვებულ მნიშვნელობებს, n_i-ით შესაბამის სიხშირეებს, ხოლო $h_i=n_i/n$, $i=1,2,..s$, ფარდობით სიხშირეებს, სადაც $h_1 + h_2 + \cdots + h_s = 1$, მაშინ მოცემული ნამდვილი x რიცხვისათვის დაგროვილი ფარდობითი სიხშირეებით განისაზღვრება $F_n(x)$ --- ემპირიული განაწილების ფუნქცია, როგორც იმ $x^{(i)}$ მნიშვნელობა'ა ფარდობი' სიხშირეების ჯამი, რომლებიც x-ს არ აღემატება (ანუ x-ზე ნაკლებია ან მისი ტოლია). გვაქვს

$$F_n(x) = \begin{cases} 0, & x < x^{(1)}, \\ \frac{n_1}{n}, & x^{(1)} \le x < x^{(2)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{n_1}{n}+\cdots+\frac{n_k}{n}, & x^{(k)} \le x < x^{(k+1)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 1, & x^{(s)} \le x. \end{cases}$$

ემპირიული განაწილების ფუნქციის გრაფიკი სავარაუდოდ ასეთია:



ნახ. 4.4. **კუმულატას გრაფიკი** (s=3)

როგორც აქედან ჩანს, $F_n(x)$ ფუნქცია უბან-უბან მუდმივია და მისი ზრდა ხდება $x^{(1)},\ldots,x^{(s)}$ წერტილებში. $x^{(1)}$-ამდე $F_n(x)$ ნულია, $x^{(n)}$-ის მარჯვნივ (მისი ჩათვლით) კი – 1. ყოველ $x^{(k)}$ წერტილში $F_n(x)$ ფუნქცია მატულობს $h_k=n_k/n$ ფარდობითი სიხშირის ტოლი სიდიდით, სწორედ ეს უკანასკნელი გარემოება ნიშნავს, რომ $F_n(x)$ მოიცავს ფარდობით სიხშირეთა განაწილებით მოცემულ ინფორმაციას.

დაგროვილ (კუმულაციურ) სიხშირეთა, ან ფარდობით სიხშირეთა ფუნქციას ზოგჯერ **კუმულატას** უწოდებენ.

მაგალითი 4.2. 4.1 მაგალითში მოყვანილი მონაცემებისათვის დაგროვილ ფარდობით სიხშირეთა ფუნქციაა

$$F_{21}(x)=\begin{cases} 0, & x<0, \\ 2/21, & 0\le x<1, \\ 10/21, & 1\le x<2, \\ 16/21, & 2\le x<3, \\ 19/21, & 3\le x<4, \\ 1, & x\ge 4. \end{cases}$$

♦

§ 2. მონაცემთა დაჯგუფება. სიხშირეთა ინტერვალური განაწილება

როგორც აღვნიშნეთ, **მონაცემებს ვუწოდებთ უწყვეტს**, თუ ისინი ღებულობენ ნებისმიერ ნამდვილ მნიშვნელობებს რაიმე შუალედში. ასეთი უწყვეტ მონაცემთა აღწერისას მიმართავენ მონაცემთა დაჯგუფებას და გამოთვლიან სიხშირეებს დაჯგუფებული მონაცემებისათვის.

დაჯგუფების მიზნით ინტერვალი, რომელშიც მოთავსებულია დაკვირვებული მნიშვნელობები მინიმალური მნიშვნელობიდან მაქსიმალურ მნიშვნელობამდე დაიყოფა ტოლი ან არატოლი სიგრძის რამდენიმე ქვეინტერვალად:

მაგალითი 4.3. დიდი ზომის ავტომანქანებში საწვავის დანახარჯის დასადგენად 100 ავტომანქანაზე ჩატარდა დაკვირვებები. 4.4 ცხრილი გვიჩვენებს გარბენის იმ მანძილს მილებში (1 მილი = 1.609 კმ), რომელსაც ყოფნის 1 გალონი (1 გალონი = 3.78 ლ) საწვავი (1 მილი/გალონზე = 1.426 კმ/ლ).

ცხრილი 4.4 საწვავის ხარჯზე 100 დაკვირვების შედეგები

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

ამ მონაცემების დამუშავება მდგომარეობს შემდეგში:

ა) დაკვირვებულ მონაცემებს შორის უნდა მოიძებნოს **მინიმალური და მაქსიმალური მნიშვნელობები.** წარმოდგენილ მონაცემებში უმცირესი რიცხვითი მნიშვნელობაა 14,1, ხოლო უდიდესი - 23,9. ეს ორი რიცხვი განვსაზღვრავს მონაცემთა გაფანტულობის უმარტივეს მახასიათებელს - გაბნევის დიაპაზონს. **გაბნევის დიაპაზონი** არის სხვაობა მონაცემთა მაქსიმალურ (უდიდეს) და მინიმალურ (უმცირეს) რიცხვით მნიშვნელობებს შორის, გაბნევის დიაპაზონი= 23,9 -14,1=9,8.

შენიშვნა. გაბნევის დიაპაზონი გამოთვლისა და ინტერპრეტაციის სიმარტივის გამო ფართოდ გამოიყენება ისეთი ამოცანების გადაწყვეტისას, სადაც გამოთვლის სისწრაფესა და მარტივ შინაარსს გადამწყვეტი როლი ენიჭება, მაგალითად ხარისხის კონტროლის სფეროში.

ბ) შემდგომისათვის აუცილებელია **დაჯგუფებათა ინტერვალების გამოყოფა**. ამისათვის უმეტეს შემთხვევაში გაბნევის დიაპაზონს ყოფენ ერთი და იგივი სიგრძის დაახლოებით 5–10 ქვეინტერვალად. მოცემული შერჩევისათვის დაჯგუფების 5 ტოლი სიგრძის ინტერვალი ავირჩიოთ. სიმარტივისათვის მინიმუმად ავიღოთ 14, ხოლო მაქსიმუმად – 24, ხოლო

ქვეინტერვალის სიგრძედ - 2 ($\Delta = 2$). ამრიგად, დაჯგუფების ინტერვალებია; ≤16, (16,18], (18,20], (20,22], >22. ყოველი ქვეინტერვალისათვის დავითვალოთ იმ ავტომანქანათა რაოდენობა, რომელთა გარბენს განსაზღვრავს აღნიშნული ქვეინტერვალი. ამრიგად, მივიღებთ დაჯგუფებულ მონაცემთა სიხშირული განაწილების შემდეგ სურათს. (დაჯგუფების ქვეინტერვალებს განსაზღვრავს შემდეგი რიცხვითი მნიშვნელობები: 16, 18, 20, 22.).

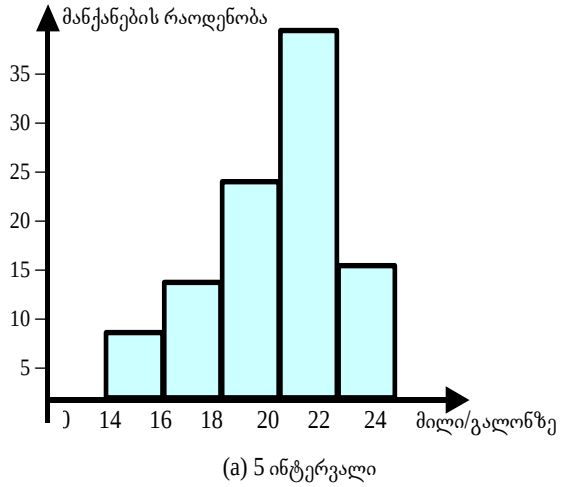
ცხრილი 4.5

გარბენი	≤16	–18	–20	–22	≥22
ავტომანქანათა რაოდენობა	9	13	24	38	16

დაჯგუფებული მონაცემების გრაფიკული წარმოდგენისათვის აგებენ ე.წ. ჰისტოგრამას. ჰისტოგრამა (histogram) არის დიაგრამა, რომელიც დაჯგუფებულ მონაცემებს წარმოადგენს (ასახავს) ცალკეული **მართკუთხედების** საშუალებით. აბსცისათა ღერძზე გადაიზომება ქვეინტერვალთა საზღვრები და თითოეულ ქვეინტერვალზე აიგება მართკუთხედი, რომლის ფუძე ქვეინტერვალის Δ სიგრძის ტოლია, ხოლო სიმაღლეა $n_i/\Delta n$, სადაც n_i ქვეინტერვალში მოხვედრილ შერჩევით მნიშვნელობათა სიხშირეა, n შერჩევის მოცულობაა.

ჰისტოგრამის აგების წესის თანახმად, ყველა ქვეინტერვალებზე აგებულ მართკუთხედების ფართობების ჯამი ერთის ტოლია; თითოეულ ქვეინტერვალზე აგებული მართკუთხედების ფართობი ამ ქვეინტერვალში ჩავარდნილ (მოხვედრილ) შერჩევით მნიშვნელობათა ფარდობითი სიხშირის ტოლია.

ქვევით, ნახ. 4.5. გამოსახულია 4.5 ცხრილის მიხედვით აგებული **სიხშირეთა ჰისტოგრამა.**



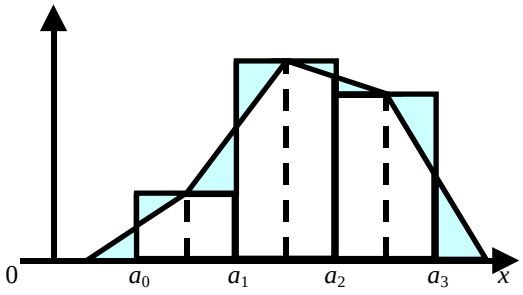
(a) 5 ინტერვალი

ნახ. 4.5. საწვავის ხარჯის სიხშირეთა ჰისტოგრამები 100 მონაცემისათვის

§ 3. პოლიგონი

ჰისტოგრამასთან ერთად განიხილავენ ე.წ. **პოლიგონსაც. ინტერვალური მონაცემებისათვის პოლიგონი მიიღება ჰისტოგრამის შემადგენელი მეზობელი მართკუთხედების**

ზედა ფუძეების შუაწერტილების შეერთებით. თუ ქვეინტერვალების სიგრძეა Δ, მაშინ პირველი ქვეინტერვალის მარცხნივ, მისი შუაწერტილიდან Δ მანძილზე და ბოლო ქვეინტერვალის შუაწერტილიდან მარჯვნივ იგივე Δ მანძილზე აიღება წერტილები და ისინიც უერთდება მიღებული ტეხილის უახლოეს წერტილებს (ნახ. 4.6). ამგვარად მიღებულ ტეხილს (რომელიც უკვე დასრულებული პოლიგონია) და *Ox* ღერძს შორის ფართობი 1-ის ტოლია.

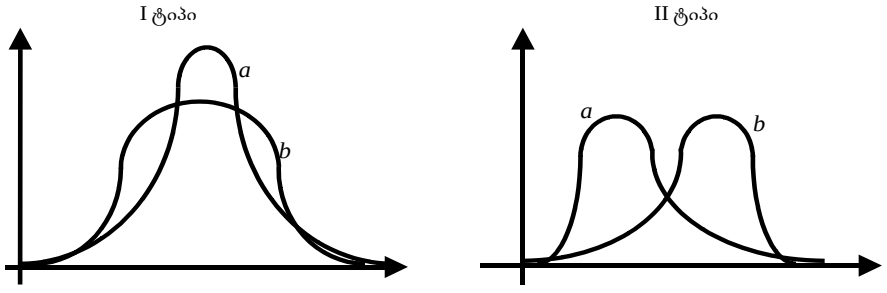


ნახ 4.6. პოლიგონი

§ 4. სიხშირეთა განაწილების ფორმები.

ჰისტოგრამისა და პოლიგონის მეშვეობით ჩვენ შეგვიძლია დავადგინოთ განაწილებათა ტიპობრივი ფორმები, რაც მნიშვნელოვანია განაწილებათა შედარებისას. (იხ. ნახ. 4.7). ცხადია, რომ რაც მეტია შერჩევის მოცულობა და დაჯგუფების ინტერვალთა რაოდენობა და რაც ნაკლებია ინტერვალის სიგრძე, მით უფრო **გლუვი** იქნებიან ჰისტოგრამა და პოლიგონი. ამიტომ სიხშირული განაწილების სავარაუდო ფორმები შეგვიძლია გლუვი წირების მეშვეობით წარმოვიდგინოთ.

პირველი ტიპის მრუდი იმ გარემოებას ასახავს, რომ კიდურა მნიშვნელობანი უფრო იშვიათად გვხვდება, ვიდრე შუალედური. ამასთან *b* მრუდი უფრო მდორედ იცვლება, ვიდრე *a* მრუდი. ორივე მრუდს ერთადერთი წვერო აქვს მნიშვნელობათა სიმრავლის შუაში და ამ მნიშვნელობის ირგვლივ ეს მრუდები სიმეტრიულია.



ნახ. 4.7. სიხშირეთა (ფარდობით სიხშირეთა) განაწილების ტიპები

მეორე ტიპის განაწილება ასიმეტრიულია წვეროს შესაბამისი მნიშვნელობის (ინტერვალის) გარშემო. *b* მრუდი მარცხნივ დამრეცია, მარჯვნივ კი ციცაბოდ ეშვება. მაქსიმალური სიხშირის მნიშვნელობის მარცხნივ მნიშვნელობათა სპექტრი უფრო განიერია, ვიდრე მარჯვნივ, და ამ მნიშვნელობისაგან ერთი და იგივე მანძილზე გადახრათაგან მარცხენას უფრო მეტი სიხშირე აქვს ვიდრე მარჯვენას. ამგვარი ასიმეტრია არის მარცხენა ასიმეტრია. *a*

მრუდი სიხშირეთა ზუსტად შებრუნებულ ყოფაქცევას გვიჩვენებს, ანუ მარჯვენა ასიმეტრია გააჩნია.

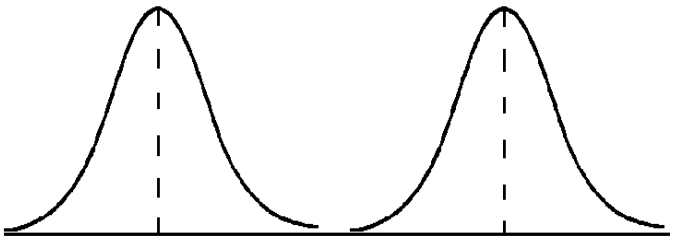
ახლა ჩვენ განვსაზღვრავთ რიცხვით მახასიათებლებს, რომლებიც საშუალებას იძლევა შეკუმშული ფორმით დავახასიათოთ რიცხვით მონაცემთა ესა თუ ის თვისება, კერძოდ, სად და როგორ არიან მონაცემები განლაგებული, სად არის **მათი განაწილების ცენტრი**, როგორია მათი **გაფანტულობა, სიმეტრიულია** თუ არა მონაცემთა განლაგება ცენტრის მიმართ, ასიმეტრიულობის რა ტიპთან გვაქვს საქმე და სხვა. გარდა ამისა, ეს რიცხვითი მახასიათებლები გვიადვილებს მონაცემთა ორი ჯგუფის, მაგალითად, ორი შერჩევის შედარებას ზემოჩამოთვლილი თვისებების მიხედვით. კერძოდ, მონაცემები, რომლებიც ხასიათდება სიხშირეთა განაწილების მსგავსი ფორმებით, შეიძლება განსხვავდებოდნენ ძირითადად ორი ნიშნით:

(1) დონით ანუ იმ ცენტრალური მნიშვნელობით, რომლის გარშემოც იყრიან თავს მონაცემები.

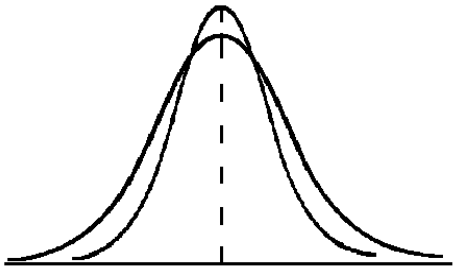
(2) დაკვირვებული მნიშვნელობების ცენტრის მიმართ გაბნევით ანუ გაფანტულობით.

სიტუაციების თვალსაჩინოდ წარმოსადგენად იხ. ნახ. 4.8–4.10, სადაც განაწილებათა ჰისტოგრამები სქემატურად, გაგლუვებული სახითაა მოცემული.

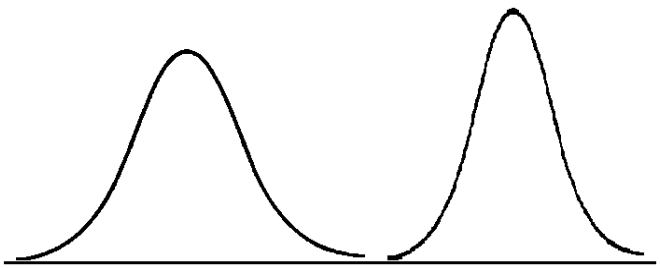
სწორედ ამ ორი ნიშნის შესაბამისად შემაჯამებელი (კრებსითი) რიცხვითი მახასიათებლები პირობითად შეგვიძლია დავყოთ ორ ძირითად ჯგუფად. პირველ ჯგუფს მიეკუთვნება მონაცემთა განლაგების ცენტრის, ანუ როგორც ამბობენ, **ცენტრალური ტენდენციის საზომები.** ესენია: **შერჩევის საშუალო** ანუ არითმეტიკული საშუალო, **შერჩევის მედიანა და მოდა.** მათ **მონაცემთა ლოკალიზაციის** საზომებსაც უწოდებენ.



ნახ. 4.8. განაწილებათა მსგავსი ფორმა, დონეთა (ცენტრების) განსხვავება.



ნახ. 4.9. განაწილებათა მსგავსი ფორმა, ერთნაირი დონე, განსხვავება გაფანტულობაში.



ნახ. 4.10. განაწილებათა მსგავსი ფორმა, დონეებისა და გაფანტულობის სხვაობა.

მეორე ჯგუფში **შედის მონაცემთა გაბნეულობის ანუ ვარიაბელობის** საზომები. ამ ჯგუფის საზომებიდან ჩვენ განვიხილავთ **გაბნევის დიაპაზონს, შერჩევის დისპერსიას და საშუალო კვადრა-ტული გადახრას, ვარიაციის კოეფიციენტს**.

თვისებრივი მონაცემებისათვის ყველაზე გავრცელებული და სასარგებლო რიცხვითი მახასიათებელია **შერჩევითი პროპორცია** (ფარდობითი სიხშირე), რომელიც მიგვითითებს, თუ რამდენად ხშირად გვხვდება რაიმე კერძო თვისების მქონე ობიექტი (ან ინდივიდი) შერჩევაში.

§ 5. არითმეტიკული საშუალო (საშუალო)

არითმეტიკული საშუალო (ანუ, უბრალოდ საშუალო-sample mean) წარმოადგენს ცენტრალური ტენდენციის ერთ-ერთ ყველაზე გავრცელებულსა და გამჭირვალე შინაარსის მქონე საზომს.

ვთქვათ, ჩვენ გაგვაჩნია რომელიმე ფირმის წლიური საპროცენტო შემოსავალი ბოლო ათი წლის განმავლობაში. ესენია

$$8.1;\ -6.2;\ 20.9;\ -2.7;\ 33.6;\ 42.9;\ 24.4;\ 5.2;\ 3.1;\ 30.5.$$

როგორია ფირმის საშუალო წლიური შემოსავალი ამ ათი წლის განმავლობაში? საშუალო შემოსავალი გამოითვლება შემდეგი წესით: შევკრებთ ყველა მონაცემს და შემდეგ გავყოფთ წლების რაოდენობაზე, ე.ი. გამოითვლება მოყვანილი რიცხვების არითმეტიკული საშუალო:

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

ამრიგად, საშუალო წლიური შემოსავალი არის 16%.

რადგანაც რიცხვითი მონაცემების ყოველი სიმრავლისათვის არითმეტიკული საშუალო გამოითვლება ერთი და იგივე გზით, ეს გამოთვლები მოსახერხებელია გამოისახოს სიმბოლურად.

ვთქვათ, მოცემულია რიცხვებით გამოხატული n დაკვირვება (n მოცულობის შერჩევითი მნიშვნელობები რაიმე პოპულაციიდან), რომლებსაც ჩვენ აღვნიშნავთ $x_1, \ldots, x_n$ სიმბოლოებით. მაშინ ამ მონაცემთა არითმეტიკული საშუალო აღინიშნება $\bar{x}_n$ -ით და გამოითვლება ფორმულით:

$$\bar{x}_n=\frac{x_1+\ldots.+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i\ , \tag{4.1}$$

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

მაგალითი 4**.4.** შესწავლილია 10 ოჯახში ბავშვთა რაოდენობა. შესაბამისი მონაცემებია

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

მაშინ ერთ ოჯახზე მოსული ბავშვების საშუალო რაოდენობა იქნება

$$\frac{3+2+1+1+2+3+3+1+2+2}{10} = 2,$$

ე.ი. 2 ბავშვი. ამრიგად, თუ 10 ოჯახს ჯამში ყავს 20 ბავშვი, მაშინ არითმეტიკული საშუალო უდრის ბავშვთა იმ რაოდენობას, რომელიც მოვიდოდა ერთ ოჯახზე ბავშვები თანაბრად რომ გაენაწილებინათ.

მოვიყვანოთ არითმეტიკული საშუალოს შემდეგი თვისება.

საშუალოდან გადახრების ჯამი ნულის ტოლია:

$$\sum_{i=1}^{n}(x_i - \bar{x}_n) = 0. \tag{4.2}$$

მართლაც,

$$\sum_{i=1}^{n}(x_i - \bar{x}_n) = \sum_{i=1}^{n} x_i - n\,\bar{x}_n = n\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \bar{x}_n\right) = n\,(\bar{x}_n - \bar{x}_n) = 0.$$

ეს თვისება იძლევა იმის ახსნას, თუ რატომ წარმოადგენს $\bar{x}$ მონაცემთა განლაგების ცენტრის მახასიათებელს. ამ მიზნით გავარკვიოთ არითმეტიკული საშუალოს ფიზიკური შინაარსი. წარმოვიდგინოთ უწონო ერთგვაროვანი ღერძი, რომელზედაც x_1, ..., x_n კოორდინატების მქონე წერტილებში განლაგებულია ერთეულოვანი მასები. ამ შემთხვევაში, მოყვანილი თვისებიდან გამომდინარე, $\bar{x}$ წარმოადგენს იმ საყრდენი წერტილის კოორდინატს, რომლის მიმართაც ღერძი წონასწორობაში იმყოფება ანუ $\bar{x}$ არის სიმძიმის ცენტრი.

ვთქვათ, x_1, ..., x_n და y_1, ..., y_n ტოლი მოცულობის ორი შერჩევაა. შევქმნათ ახალი შერჩევა $z_1 = x_1 \pm y_1$, ..., $z_n = x_n \pm y_n$. მაშინ ადვილად მოწმდება, რომ

$$\bar{z}_n = \bar{x}_n \pm \bar{y}_n .$$

თუ x_1, ..., x_n და y_1, ..., y_m ორი შერჩევაა, მაშინ

$$x_1, \ldots, x_n, y_1, \ldots, y_m$$

გაერთიანებული შერჩევის არითმეტიკული საშუალო მოიცემა ფორმულით:

$$\bar{z}_{n+m} = \frac{n}{n+m}\,\bar{x}_n + \frac{m}{n+m}\,\bar{y}_m$$

ანუ $\bar{z}_{n+m}$ არის ცალკეულ არითმეტიკულ საშუალოთა შეწონილი ჯამი, შესაბამისად, $\omega_1 = \frac{n}{n+m}$ და $\omega_2 = \frac{m}{n+m}$, წონებით. შევნიშნოთ, რომ $\omega_1 + \omega_2 = 1$.

იმის გამო, რომ საშუალო ადვილი გამოსათვლელია, გააჩნია მარტივი ინტერპრეტაცია და მასზე შეიძლება მათემატიკური ოპერაციების ჩატარება, ის წარმოადგენს მონაცემთა ცენტრის ერთ-ერთ ყველაზე გავრცელებულ რიცხვით მახასიათებელს.

§ 6. მედიანა

მედიანა (median), არითმეტიკული საშუალოს შემდეგ, წარმოადგენს ცენტრალური ტენდენციის ყველაზე გავრცელებულ საზომს.

როგორც ცნობილია „მედიანა“ არის „შუა“-ს სინონიმი. სინამდვილეშიც, მედიანა განიმარტება, როგორც მონაცემთა ვარიაციული მწკრივის (ანუ ზრდის მიხედვით დალაგებული დაკვირვებების მწკრივის) შუა მნიშვნელობა იმ აზრით, რომ მასზე ნაკლები რიცხვითი მნიშვნე-

ლობის მონაცემთა რაოდენობა მასზე დიდი რიცხვითი მნიშვნელობის მონაცემთა რაოდენობის ტოლია.

განვიხილოთ შემდეგი მაგალითი.

მაგალითი 4**.5.** 9 სამუშაო დღის განმავლობაში სამსახურში არ გამოცხადებულ თანამშრომლებთა რაოდენობებია.

3, 5, 1, 1, 0, 4, 3, 2, 0

ვარიაციულ მწკრივად დალაგებული მონაცემებია

0, 0, 1, 1, 2, 3, 3, 4, 5

მოყვანილ მაგალითში დაკვირვებათა რაოდენობა კენტია (n=9) და მედიანა, რომელიც *Me*-ით აღინიშნება, მართლაც არის ვარიაციული მწკრივის შუა წევრი, *Me* =2.

თუ დაკვირვებათა რაოდენობა ლუწია, მაგალითად $n = 2k$, მაშინ მედიანა მონაცემთა ვარიაციული მწკრივის k-ური და (k+1)-ე წევრების საშუალო არითმეტიკულია.

დავუშვათ ვარიაციულ მწკრივად დალაგებული n=10 მოცულობის მონაცემებია

1, 1, 3, 3, 4, 5, 5, 6, 7, 9

ამ შემთხვევაში მედიანა, ვარიაციულ მწკრივში რიგით მეხუთე და მეექვსე წევრების საშუალო არითმეტიკულია

$$Me = \frac{4+5}{2} = 4{,}5\,.$$

§ 7. მოდა

ცენტრალური ტენდენციის მესამე საზომს წარმოადგენს **მოდა (mode)**. ის წარმოადგენს იმ მნიშვნელობას შერჩევიდან, რომელიც ყველაზე ხშირად გვხვდება. მაგალითად, თუ ჩვენი მონაცემებია

2, 3, 5, 4, 6, 4, 7, 1,

მაშინ მოდა იქნება 4, რადგანაც ის 2-ჯერ შეგვხვდა შერჩევაში, ყველა დანარჩენი დაკვირვება კი ერთხელ.

თუ აგებული გვაქვს ასე თუ ისე მისაღები ჰისტოგრამა (ანუ ჰისტოგრამა, რომელიც ადეკვატურად აღწერს მონაცემთა სიხშირულ განაწილებას), მაშინ მოდალური ინტერვალი იქნება ის ინტერვალი, რომელსაც შეესაბამება ჰისტოგრამის მაქსიმალური მნიშვნელობა. ამრიგად, მოდალური ინტერვალი წარმოადგენს იმ ინტერვალს, რომელშიც მოხვდა დაკვირვებათა ყველაზე დიდი რაოდენობა. გარკვეულობისათვის მოდას უწოდებენ მოდალური ინტერვალის შუა წერტილს. თუ შერჩევის მოცულობა დიდია, ამჯობინებენ მონაცემთა დაჯგუფებას და მოდალური ინტერვალის მოძებნას.

ადვილი შესამჩნევია, რომ მოდის (მოდალური ინტერვალის) მოძებნა საკმაოდ რთულია. იმ ინტერვალის მითითება, რომელშიც მოხვდა დაკვირვებათა მაქსიმალური რაოდენობა, დამოკიდებულია იმაზე, თუ როგორ დავაჯგუფეთ მონაცემები. სახელდობრ, იმაზე თუ რა სიგრძის ინტერვალებად დავყავით ინტერვალი $[x_{(1)}, x_{(n)}]$. ამიტომ მოდალური ინტერვალის განსაზღვრას თან სდევს ყველა ის სირთულე, რომელიც დაკავშირებულია ჰისტოგრამების აგებასთან.

შეიძლება აღმოჩნდეს, რომ ის მნიშვნელობა, რომელიც ყველაზე ხშირად გვხვდება ორია ან ორზე მეტიც კი. შემდგომში ჩვენ ვიგულისხმებთ, რომ საქმე გვაქვს ე.წ. უნიმოდალურ განაწილებებთან ე.ი. როდესაც მოდა ერთადერთია, ანდა გვაქვს ერთადერთი მოდალური ინტერვალი.

შერჩევითი მოდის ინტერპრეტაცია წააგავს ყოველდღიურ ცხოვრებაში სიტყვა „მოდის“ მნიშვნელობას. როგორც ის, რაც „მოდაშია“ (ანუ ყველაზე ხშირია, გავრცელებულია) შეიძლება სულაც არ აღწერდეს საერთო სტილს და ასევე შერჩევითი მოდაც ახასიათებს დაკვირვებათა მხოლოდ მცირე ნაწილს.

შერჩევითი მოდა იძლევა მნიშვნელოვან ინფორმაციას სხვადასხვა საქონლის გამომშვები ფირმისათვის, მაღაზიის მფლობელისათვის, კონსტრუქტორებისათვის, რომლებიც აწოდებენ საქონელს სპეციფიკურ ბაზრებს. მაგალითად, ტანსაცმლის მაღაზიის მფლობელმა უნდა იცოდეს კოსტუმების ყველაზე გავრცელებული ზომები, რათა გაითვალისწინოს ეს ფაქტი საქონლის მომარაგებისას; საათების მწარმოებელმა უნდა იცოდეს, თუ რა კლასის საათები იყიდება ყველაზე ხშირად, იმ მიზნით, რომ გაზარდოს ამ კლასის პროდუქციის წარმოება.

ჩვენ შევისწავლეთ ცენტრალური ტენდენციის (ანუ მონაცემთა ცენტრის მდებარეობის) სამი რიცხვითი მახასიათებელი: შერჩევითი საშუალო, შერჩევითი მედიანა და შერჩევითი მოდა. აშკარაა, რომ ეს სამი მახასიათებელი სრულად ვერ აღწერს მონაცემთა სიმრავლის თვისებებს.

შემდეგი კითხვა, რომელიც დაგვებადება მონაცემთა სიმრავლის თვისებების გამოსაკვეთად იქნება შემდეგი: როგორია მონაცემთა გაფანტულობა საერთოდ და კერძოდ, საშუალო მნიშვნელობის მიმართ? როგორია მონაცემთა ცვალებადობის (ვარიაბელობის) ხარისხი და რა სიდიდეებით შეიძლება მისი დახასიათება?

§ 8. კვარტილები

§ 2-ში ჩვენ განვსაზღვრეთ მონაცემთა გაფანტულობის ერთ-ერთი უმარტივესი რიცხვითი მახასიათებელი-**გაბნევის დიაპაზონი (range**), რომელიც წარმოადგენს სხვაობას შერჩევის მაქსიმალურ (უდიდეს) და მინიმალურ (უმცირეს) წევრებს შორის, ხოლო § 6-ში მედიანა. *Me*-მედიანა, როგორც ვიცით, ვარიაციული მწკრივის შუა მნიშვნელობაა. მედიანის განსაზღვრის შემდეგ ვარიაციულ მწკრივში Me-ის მარცხნივ განლაგებული მონაცემების მედიანას, Q_1-ს, **პირველ კვარტილი** ეწოდება, ხოლო ვარიაციულ მწკრივში Me-ის მარჯვნივ განლაგებული მონაცემების მედიანას, Q_3-ს, **მესამე კვარტილი** ეწოდება. ცხადია, მედიანას მეორე კვარტილსაც უწოდებენ, $Q_2=Me$. უხეშად რომ ვთქვათ, (Q_1,Me,Q_3) კვარტილები მონაცემთა სიმრავლეს ოთხ ტოლ ნაწილად ყოფენ.

მაგალითი 4**.6.** მარკეტინგის კონსულტანტი შეისწავლის ბოსტნეულის მაღაზიის 50 მყიდველს. ერთ-ერთი სიდიდე, რომელიც მას აინტერესებს, არის თითოეულის მიერ დახარჯული თანხა. ვარიაციულ მწკრივად დალაგებული და დოლარის დამრგვალებული სიზუსტით მონაცემები შემდეგია:

2 6 6 8 9 10 11 11 12 12 13 13 14 14 14 15 15 16
17 18 18 18 19 19 20 || 20 20 21 23 26 27 28 29 30 31 32
33 33 34 36 37 39 40 43 45 52 61 63 64 69

მოვძებნოთ მედიანა. რადგან $n = 50$ ლუწია, ამიტომ

$Me = 20$.

აქ აუცილებლად უნდა შევნიშნოთ, რომ თუმცა $Me=20$, ის არ ემთხვევა ვარიაციული მწკრივის არცერთ წევრს: მისი პოზიცია არის ვარიაციული მწკრივის 25-ე და 26-ე პოზიციების არითმეტიკული საშუალო (მიღებული შეთანხმებაა კვარტილების ზემოთ შემოთავაზებული

წესით გამოთვლისას. ეს პოზიცია აღნიშნულია ||-ით), ამ პოზიციის მარცხნივ მოთავსებულია 25 წევრი, რომელთა მედიანა ანუ სრულ მონაცემთა პირველი კვარტილია (n = 25).

$$Q_1 = 14.$$

მედიანის (*Me*-ის) მარჯვნივაც 25 წევრია მოთავსებული, რომელთა მედიანა, ანუ სრულ მონაცემთა მესამე კვარტილია მედიანის პოზიციიდან მარჯვნივ მე-13-ე პოზიციაზე მყოფი წევრი:

$$Q_3 = 33.$$

ამრიგად, მოყვანილი მონაცემებისათვის Me=20, Q_1=14, Q_3=33. ♦

კვარტილები მედიანასთან ერთად, ე.ი. Me, Q_1, Q_3 სამეული გვაძლევს გარკვეულ წარმოდგენას მონაცემთა ცენტრის, გაბნევისა და განაწილების (გაგლუვებული ჰისტოგრამის ან პოლიგონის) ფორმის შესახებ. ზემოთ განხილულ მაგალითში ის ფაქტი, რომ Q_3 უფრო შორსაა მედიანისაგან, ვიდრე Q_1 ანუ 13 = Q_3– Me>Me– Q_1 = 6, მიუთითებს იმაზე, რომ განაწილება მარჯვნივ ასიმეტრიულია (ანუ მისი მარჯვენა ბოლო უფრო გაწელილია, ვიდრე მარცხენა).

მანძილი კვარტილებს შორის წარმოადგენს გაფანტულობის მარტივ საზომს, რომელიც ახასიათებს ვარიაციულ მწკრივში კვარტილებს შორის მოთავსებულ მონაცემთა გაფანტულობის ხარისხს (შუა 50%-იანი ზონის გაბნევას). მას ეწოდება **კვარტილთშორისი გაბნევის დიაპაზონი (Interquartil range – IQR):**

$$\mathbf{IQR = Q_3 - Q_1}$$

განხილულ მაგალითში IQR = 33–14 = 19.

Q_1 და Q_3 კვარტილები და კვარტილთშორისი გაბნევის დიაპაზონი მდგრადი მახასიათებლებია. არცერთი მათგანი არ რეაგირებს [Q_1, Q_3] ინტერვალის გარეთ მოთავსებულ დაკვირვებების ნებისმიერ ისეთ ცვლილებაზე, რომელიც ამ დაკვირვებებს კვლავ [Q_1, Q_3] ინტერვალის გარეთ ტოვებს, თანაც ინტერვალის იმავე მხარეს, სადაც ადრე იმყოფებოდა. განხილულ მაგალითში, პირველი 12 წევრის ნებისმიერი ისეთი ცვლილება, რომ არცერთმა არ მიიღოს 14-ზე მეტი მნიშვნელობა და ბოლო 12 წევრის ისეთი ცვლილება, რომ არცერთმა მათგანმა არ მიიღოს 33-ზე ნაკლები მნიშვნელობა, არ შეცვლის არც Q_1-ს, Q_3-ს და არც IQR-ს.

რადგან მედიანა, პირველი და მესამე კვარტილები (*Me*, Q_1, Q_3) არ შეიცავენ ინფორმაციას განაწილების (გაგლუვებული ჰისტოგრამების) კუდების შესახებ, ამიტომ განაწილების ფორმის შესახებ უფრო სრულ ინფორმაციას იძლევა ხუთეული

$$Q_0=x_{\min},\quad Q_1,\quad Q_2= Me,\quad Q_3,\quad Q_4=x_{\max}\ .$$

ბოსტნეულის მაღაზიის მყიდველთა დანახარჯების მონაცემებისათვის (მაგალითი 4.6) ეს ხუთეულია

$x_{\min}$ = 2, Q_1 = 14, Me = 20, Q_3 = 34, $x_{\max}$ = 69.

§ 9. დისპერსია

დისპერსია (variance) წარმოადგენს მონაცემთა საშუალოს მიმართ გაფანტულობის ერთ-ერთ უმნიშვნელოვანეს საზომს.

ვთქვათ, მოცემულია n მოცულობის $x_1, \ldots, x_n$ შერჩევა. გამოვთვალოთ არითმეტიკული საშუალო $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ და მისგან გადახრები

$$x_1 - \bar{x}_n\,,\ x_2 - \bar{x}_n\,,\ \ldots\,,\ x_n - \bar{x}_n$$

(შევნიშნოთ, რომ $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n) = 0$, ამიტომ, როგორც უკვე ვთქვით, გაფანტულობის საზომად $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)$ არ გამოდგება). გამოვთვალოთ გადახრათა კვადრატები

$$(x_1 - \bar{x}_n)^2,\ (x_2 - \bar{x}_n)^2,\ \ldots\,,\ (x_n - \bar{x}_n)^2$$

და მათი არითმეტიკული საშუალო. გვექნება **შერჩევითი დისპერსია**

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\,. \tag{4.3}$$

თუ $x_1, \ldots, x_N$ წარმოადგენს პოპულაციას (და არა შერჩევას), მაშინ **პოპულაციის საშუალო**

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

ხოლო

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{4.4}$$

გამოსახულებას ეწოდება **პოპულაციის დისპერსია**.

შემდეგ თავებში დისპერსიის ცნება შემოღებული იქნება უსასრულო პოპულაციისათვისაც. აქ კი შემოვიფარგლებით მხოლოდ იმ ფაქტის აღნიშვნით, რომ შერჩევითი დისპერსია წარმოადგენს პოპულაციის დისპერსიის ემპირიულ ანალოგს, რომელიც დაკვირვებათა რიცხვის უსასრულოდ ზრდისას მიისწრაფის პოპულაციის დისპერსიისაკენ.

შემდგომში ჩვენ ვნახავთ, რომ n მოცულობის შერჩევითი მნიშვნელობებისათვის, მოსახერხებელია გაფანტულობის ოდნავ მოდიფიცირებული საზომის განხილვა, რომელსაც შესწორებული შერჩევითი დისპერსია ეწოდება:

$$s_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\,s_n^2 \tag{4.5}$$

რადგან შერჩევითი დისპერსიის გამოთვლამდე $\bar{x}_n$ უკვე გამოთვლილია, შერჩევითი დისპერსიის და შესწორებული შერჩევითი დისპერსიის უფრო მარტივად გამოსათვლელი გამოსახულებებია

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}x^2 - (\bar{x}_n)^2 = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n^2}\,, \tag{4.6}$$

$$s_n'^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}_n^{\,2}\right) = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}\,. \tag{4.7}$$

შენიშვნა. თუ პოპულაციის μ საშუალო ცნობილია, მაშინ შერჩევითი დისპერსია განისაზღვრება ფორმულით

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\,. \tag{4.8}$$

აქვე აღვნიშნოთ, რომ (4.6) ფორმულით გამოითვლება პოპულაციის, ხოლო (1.7)-ით შერჩევის დის-პერსია.

მაგალითი 1.7. შემდეგი მონაცემებისათვის 2,4,7,11 გამოვთვალოთ შერჩევითი საშუალოსა და შერჩევითი დისპერსიის რიცხვითი მნიშვნელობები.

ამოხსნა. გამოთვლები ჩავატაროთ შემდეგი ცხრილის გამოყენებით:

n	x_i	x_i^2
1	2	4
2	4	16
3	7	49
4	11	121
Σ	24	190

მივიღებთ

$$\bar{x}_4 = 6,\ s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x}_n)^2 = \frac{1}{4}\cdot 190\text{-}6^2 = 12.5. \quad \blacklozenge$$

როგორც ვხედავთ, შერჩევითი დისპერსია მკაცრად განსაზღვრულია, გამოითვლება ყველა დაკვირვების საფუძველზე და ადვილად ექვემდებარება ალგებრულ ოპერაციებს. ამასთან, შერჩევითი დისპერსიის განზომილება უდრის დაკვირვებათა განზომილების კვადრატს. მაგალითად, თუ დაკვირ-ვებები იზომება წუთებში (კილოგრამებში, ლარებში, დოლარებში), მაშინ დისპერსიის საზომი ერთე-ულია წთ^2 (კგ^2, ლარი^2, დოლარი^2).

§ 10. სტანდარტული გადახრა

სტატისტიკოსს ხშირად ესაჭიროება **გაფანტულობის ისეთი საზომი**, რომელიც ამავე ერთეულებში იზომება, რაც საწყისი მონაცემები. ასეთი საზომი მარტივად მიიღება შერჩევითი დის-პერსიიდან კვადრატული ფესვის ამოღებით.

სტანდარტული გადახრა (standard deviation) წარმოადგენს კვადრატულ ფესვს შერჩევითი დისპერსიიდან

$$s' = \sqrt{s'^2}\ , \tag{4.9}$$

სტანდარტული გადახრა, ისევე როგორც დისპერსია, ზომავს გაფანტულობას საშუალოს მიმართ და გამოიყენება მაშინ, როდესაც მონაცემთა ცენტრის საზომად მიღებულია საშუალო. ცხადია, რომ $s' = 0$ მაშინ და მხოლოდ მაშინ, როდესაც ყველა x_i ტოლია.

აქვე აღვნიშნოთ, რომ სიდიდეს

$$\frac{s'}{\sqrt{n}} \tag{4.10}$$

სტანდარტულ შეცდომას (standard error) უწოდებენ.

§ 11. სტანდარტული გადახრის ინტერპრეტაცია. ჩებიშევის უტოლობა

სტანდარტული გადახრა მეტად სასარგებლო მახასიათებელია. სტანდარტული გადახრის მეშვეობით შესაძლებელი დასკვნების გაკეთება სხვადასხვა ინტერვალებში მოხვედრილ მონაცემთა წილის შესახებ. ამის შესაძლებლობას იძლევა ჩებიშევის უტოლობა, რომელიც ასე ფორმულირდება: $[\bar{x} - ks,\ \bar{x} + ks]$ ინტერვალში მოხვედრილ მონაცემთა ფარდობითი სიხშირე (ანუ ამ ინტერვალში მოხ-ვედრილ მონაცემთა რაოდენობის ფარდობა შერჩევის n მოცულობასთან) არანაკლებია $\left(1-\frac{1}{k^2}\right)$-ზე.

ჩებიშევის უტოლობა იძლევა $[\bar{x} - ks,\ \bar{x} + ks]$ ინტერვალში მოხვედრილი დაკვირვებების ფარდობითი სიხშირის ქვედა საზღვარს, როდესაც $\bar{x}$-ისა და s-ის გარდა სხვა ინფორმაცია დაკვირვებათა განაწილების შესახებ არ გაგვაჩნია.

ცხრილი 1.6. ჩებიშევის უტოლობის ილუსტრაცია

ინტერვალი	ფარდობითი სიხშირე	პროცენტული რაოდენობა
$[\bar{x} - s,\ \bar{x} + s]$	≥ 0	$\geq 0\%$
$[\bar{x} - 2s,\ \bar{x} + 2s]$	$\geq 3/4$	$\geq 75\%$
$[\bar{x} - 2.5s,\ \bar{x} + 2.5s]$	$\geq 21/25$	$\geq 84\%$
$[\bar{x} - 3s,\ \bar{x} + 3s]$	$\geq 8/9$	$\geq 89.9\%$
$[\bar{x} - 4s,\ \bar{x} + 4s]$	$\geq 15/16$	$\geq 93.7\%$

ჩებიშევის აღნიშნული უტოლობა ნათლად მიუთითებს სტანდარტული გადახრის ინტერპრეტაციაზე და მის მნიშვნელობაზე.

§ 12. შერჩევითი ვარიაციის კოეფიციენტი

მონაცემთა სიმრავლის ვარიაციის კოეფიციენტი (**შერჩევითი ვარიაციის კოეფიციენტი (sample coefficient of variation),** რომელიც აღინიშნება CV-თი მოიცემა ფორმულით

$$CV = \frac{s}{\bar{x}} \text{ ან } CV' = \frac{s'}{\bar{x}},\ \bar{x} \neq 0. \qquad (4.11)$$

ცხადია, რომ ვარიაციის კოეფიციენტი უგანზომილებო სიდიდეა.

თუ სტანდარტული გადახრა გამოიყენება მონაცემთა ერთი სიმრავლის ცვალებადობის დასახასიათებლად, ვარიაციის კოეფიციენტს უმეტეს შემთხვევაში იყენებენ მონაცემთა ორი სიმრავლის გაფანტულობის ხარისხის შესადარებლად. ვარიაციის კოეფიციენტი ითვალის-წინებს საშუალოს სიდიდესაც და ფაქტობრივად, ზომავს საშუალოს ერთეულზე მოსულ სტან-დარტულ გადახრას. შესაძლებელია, რომ მონაცემთა ერთი სიმრავლის სტანდარტული გადახრა უფრო დიდი აღმოჩნდეს მეორე სიმრავლის სტანდარტულ გადახრაზე, მაგრამ ვარიაციის კოეფი-ციენტებისათვის გვქონდეს შებრუნებული თანაფარდობა. ვარიაციის კოეფიციენტი საშუალოთა განსხვავების ფაქტორის გამორიცხვის საშუალებას იძლევა.

მაგალითი 4.8. განვიხილოთ რაიმე ორი ფირმა, რომელიც უშვებს ერთი და იგივე ტიპის პროდუქციას (ვთქვათ, სათამაშოებს). ვთქვათ, ამ ფირმების წლიურ გაყიდვათა ზრდის ტემპი ცვალებადია წლიდან წლამდე და მოცემულია 4.7 ცხრილის მეშვეობით (მონაცემები გამოხატულია პროცენტებში).

ცხრილი 4.7.

	I წელი	II წელი	III წელი	IV წელი	V წელი
ფირმა M	13.7	23.9	18.3	–52.8	39.1
ფირმა T	8.4	14.0	28.9	–19.9	50.0

დავრწმუნდეთ, რომ T ფირმის გაყიდვათა ზრდის ტემპი უფრო სტაბილურია, ვიდრე M ფირმისა, ანუ M ფირმის გაყიდვათა ზრდის ტემპის ვარიაბელობა (ცვალებადობა) უფრო ძლიერია T ფირმასთან შედარებით.

ამოხსნა. რადგან დისპერსია მონაცემთა ვარიაბელობის ყველაზე თვალსაჩინო საზომია. ყოველი ტრესტის მონაცემების მიხედვით გამოვთვალოთ თითოეული მათგანის საშუალოები, შესწორებული შერჩევითი დისპერსიები, ვარიაციის კოეფიციენტები და შევადაროთ ისინი ერთმანეთს. მივიღებთ

$$\bar{x}_M = 8.44\%,\ s_M'^2 = \tfrac{1}{4}(5.26^2 + 15.46^2 + 9.86^2 + 61.24^2 + 30.66^2) \approx 1263.57(\%)^2,\ s_M' = 35.55\%$$

$$\bar{x}_T = 16.28\%,\ s_T'^2 = \tfrac{1}{4}(9.88^2 + 2.28^2 + 12.62^2 + 36.18^2 + 33.72^2) \approx 668(\%)^2,\ s_T' = 25.85\%$$

$$CV_M' = \frac{s_M'}{\bar{x}_M} \approx 4.21,\quad CV_T' = \frac{s_T'}{\bar{x}_T} \approx 1.59,\quad \Rightarrow\quad CV_M' > CV_T'$$

რადგან $CV_M' > CV_T'$, ამიტომ შეგვიძლია დავასკვნათ, რომ M ფირმის გაყიდვათა ზრდის ტემპის ვარიაბელობა წლების მიხედვით უფრო მაღალია, ვიდრე T ფირმისა.

§ 13. ასიმეტრიისა და ექსცესის შერჩევითი კოეფიციენტები

ასიმეტრიის შერჩევითი კოეფიციენტი ეწოდება სიდიდეს

$$\hat{\alpha}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}. \qquad (4.12)$$

ეს კოეფიციენტი იძლევა **სიხშირეთა ჰისტოგრამის სიმეტრიულობის (ან ასიმეტრიულობის) შესახებ დასკვნის გაკეთების საშუალებას მისი აგების გარეშე.** კერძოდ, თუ

$$\hat{\alpha}_n \neq 0,$$

შეგვიძლია დავასკვნათ, რომ სიხშირეთა ჰისტოგრამა არაა სიმეტრიული $\bar{x}$-ის მიმართ, უფრო მეტიც, თუ

$$\hat{\alpha}_n > 0 \qquad (\hat{\alpha}_n < 0)$$

მაშინ სიხშირეთა განაწილება მარჯვნივ (მარცხნივ) ასიმეტრიულია. მართლაც, თუ $\hat{\alpha}_n > 0$, მაშინ $\sum_{i=1}^{n}(x_i - \bar{x})^3 > 0$, საიდანაც გამომდინარეობს, რომ $\bar{x}$-დან დადებით გადახრათა კუბების ჯამი აღემატება $\bar{x}$-დან უარყოფით გადახრათა კუბების ჯამს, რაც მიუთითებს იმაზე, რომ $\bar{x}$-ის მარცხნივ მყოფი დაკვირვებები უფრო კონცენტრირებულნი არიან $\bar{x}$–თან, ვიდრე მარჯვნივ მყოფნი, ანუ სიხშირეთა განაწილება მარჯვნივ ასიმეტრიულია.

ექსცესის შერჩევითი კოეფიციენტი ეწოდება სიდიდეს

$$\hat{e}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3. \qquad (4.13)$$

შემდგომ თავებში ჩვენ განვიხილავთ ასიმეტრიისა და ექსცესის კოეფიციენტებს პოპულაციისათვის და მათზე საუბარი დაწვრილებით იქ გვექნება. აქ შემოვიფარგლებით მხოლოდ იმის აღნიშვნით, რომ ექსცესის კოეფიციენტი არის სიხშირეთა ჰისტოგრამის $\bar{x}$-ის მიდამოში ზევით გაწელილობის (ანუ წვეტიანობის) საზომი.

§ 14. EXCEL

წარმოგიდგენთ **Excel**-ის სტატისტიკურ ფუნქციებს, რომლებიც ხშირად გამოიყენება *მონაცემთა პირველადი დამუშავების* დროს:

AVERAGE, COUNT, FREQUENCY, KURT, LARGE, MAX, MEDIAN, MIN, MODE, QUARTILE, SMALL, STDEV, STDEVP, SKEW, VAR, VARP.

ცხრილში მოცემულია ამ ფუნქციების ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა

აღნიშვნა	შინაარსი
Number	მნიშვნელობათა (რიცხვითი) სიმრავლე
Value	მნიშვნელობათა სიმრავლე
Data array	მონაცემები
Bins array	მონაცემთა დაჯგუფების ინტერვალები
Array	სიმრავლე, ცხრილი
Quart	მეოთხედი

ქვევით ჩვენ მოვიყვანთ აღნიშნული ფუნქციების გამოყენების მაგალითებს შემდეგი მონაცემების საფუძველზე:

18.5	29.22	28.67	18.27	30	12.3	9.41	9.69	
13.61	23.22	9.08	36.22	11.55	40.92	17.16	39.48	(1)
21.22	30	18.83	27.19	10.77	29.74	17.98	5.99	
26.38	24.11	25.48	33.38	15.11	22.73	20.67	20.75	
25.99	9.49	14.57	20.04	16.13	30	18.17	59.27	

ვიგულისხმებთ, რომ მონაცემები ჩაწერილია Excel-ის დავთრის ფურცლის **B2:I6** არეში.

AVERAGE ფუნქციის ჩაწერის სინტაქსი ასეთია:

AVERAGE(Number1, Number2,..).

AVERAGE - გამოიანგარიშება რიცხვითი მონაცემების არითმეტიკული საშუალო.

მონაცემთა (1) სიმრავლისათვის **AVERAGE** ფუნქცია ასე ჩაიწერება:

AVERAGE(Number1)=AVERAGE(B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ მონაცემების არითმეტიკული საშუალო 22,28-ის ტოლია.

COUNT ფუნქციის ჩაწერის სინტაქსი ასეთია:

COUNT(Value 1, Value 2,..).

COUNT– გამოიანგარიშება რიცხვითი ტიპის მონაცემთა რაოდენობა.

მონაცემთა (1) სიმრავლისათვის **COUNT** ფუნქცია ასე ჩაიწერება:

COUNT(Value1)= COUNT (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ მონაცემთა რაოდენობა (1) სიმრავლეში 40-ის ტოლია.

FREQUENCY ფუნქციის ჩაწერის სინტაქსი ასეთია:

FREQUENCY(Data array, Bins array).

FREQUENCY– გამოიანგარიშება მონაცემთა რაოდენობა, რომლებიც ეკუთვნიან წინასწარ განსაზღვრულ დაჯგუფების ინტერვალებს.

მონაცემთა (1) სიმრავლისათვის გამოვიანგარიშოთ მონაცემთა რაოდენობა, რომლებიც ეკუთვნიან დაჯგუფების შემდეგ ინტერვალებს: (-∞,12], (12, 25],(25,40},(40,+∞).

შენიშვნა. დაჯგუფების აღნიშნულ ოთხ ინტერვალს განსაზღვრავს შემდეგი სამი რიცხვითი მნიშვნელობა: 12, 25, 40. მართლაც, 12 არის პირველი ინტერვალის ზედა საზღვარიბ რომელიც ამავე დროს მომდევნო ინტერვალის ქვედა საზღვარია და ა.შ., 40 არის ბოლო ინტერვალის ქვედა საზღვარი. დავუშვათ, რიცხვითი მნიშვნელობები 12, 25, 40 ჩაწერილია Შჩსუდ-ის დავთრის ფურცლის B4:B6 არეში. **Bins array (B8:B10)** არის განსაზღვრის შემდეგ, საჭიროა განისაზღვროს იმ არის მისამართი, რომელშიც ჩაიწერება გამოთვლის შედეგები (დავუშვათ ეს არეა **C8:C11**). რომლებიც ეკუთვნიან B4:B6 ჩანაწერით განსაზღვრულ დაჯგუფებათა ინტერვალებს. გამოთვლების ჩასატარებლად ერთდროულად დავაჭიროთ კლავიშებს: Shift+Ctrl+Enter.

მონაცემთა (1) სიმრავლისათვის **FREQUENCY** ფუნქცია ასე ჩაიწერება:

FREQUENCY(Data array, Bins array)= FREQUENCY(B2:I6, B8:B10).

გამოთვლის შედეგად C8:C11 არეში ჩაიწერება გამოთვლის შედეგები

7
18
13
2

KURT ფუნქციის ჩაწერის სინტაქსი ასეთია:

KURT(Number1, Number2,..).

KURT- გამოიანგარიშება შერჩევითი ასიმეტრიის კოეფიციენტი ფორმულით $\frac{\sum(x-\bar{x})^3}{ns^3}$.

მონაცემთა (1) სიმრავლისათვის **KURT** ფუნქციის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება: **KURT(Number1)= KURT(B2:I6).**

გამოთვლის შედეგად მივიღებთ, რომ მონაცემების ასიმეტრიის კოეფიციენტი 2.534ის ტოლია.

LARGE ფუნქციის ჩაწერის სინტაქსი ასეთია:

LARGE(Array,k.)

LARGE– გამოიანგარიშება უდიდესი *k*-ური წევრი.

მონაცემთა (1) სიმრავლისათვის სიდიდით მესამე წევრის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

LARGE(Array,k.)= LARGE(B2:I6,3).

გამოთვლის შედეგად მივიღებთ, რომ სიდიდით მესამე წევრი მონაცემთა სიმრავლეში 39,48-ის ტოლია.

MAX ფუნქციის ჩაწერის სინტაქსი ასეთია:

MAX(Number1, Number,..).

MAX – გამოიანგარიშება უდიდესი მნიშვნელობა.

მონაცემთა (1) სიმრავლისათვის უდიდესი მნიშვნელობის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

MAX(Number1)=MAX (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ უდიდესი რიცხვითი მნიშვნელობა 59.27-ის ტოლია.

MEDIAN ფუნქციის ჩაწერის სინტაქსი ასეთია:

MEDIAN(Number1, Number2,..).

MEDIAN– გამოიანგარიშება მონაცემთა მედიანა.

მონაცემთა (1) სიმრავლისათვის მედიანას გამოსაანგარიშებლად **MEDIAN** ფუნქცია ასე ჩაიწერება:

MEDIAN(Number1)= MEDIAN(B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ მედიანა 20.71-ის ტოლია.

MIN ფუნქციის ჩაწერის სინტაქსი ასეთია:

MIN(Number1, Number2,..).

MIN–გამოიანგარიშება უმცირესი მნიშვნელობა.

მონაცემთა (1) სიმრავლისათვის უმცირესი მნიშვნელობის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

MIN(Number1)=MIN (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ უმცირესი მნიშვნელობა 5.99-ის ტოლია.

MODE ფუნქციის ჩაწერის სინტაქსი ასეთია:

MODE(Number1, Number2).

MODE– გამოიანგარიშება მონაცემთა მოდა.

მონაცემთა (1) სიმრავლისათვის მოდას გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

MODE(Number1)=MODE (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ მოდა 30-ის ტოლია.

QUARTILE ფუნქციის ჩაწერის სინტაქსი ასეთია:

QUARTILE(Array,Quart).

QUARTILE(x_1:x_n,k) - გამოიანგარიშება მონაცემთაკვარტილი;

როცა k=0 გამოიანგარიშება უმცირესი მნიშვნელობა,

როცა k=1 გამოიანგარიშება პირველი კვარტილი (25 პროცენტილი),

როცა k=2 გამოიანგარიშება მეორე კვარტილი, მედიანა (50 პროცენტილი),

როცა k=3 გამოიანგარიშება მესამე კვარტილი (75 პროცენტილი),

როცა k=4 გამოიანგარიშება უდიდესი მნიშვნელობა.

შენიშვნა. Q_1 პირველი და Q_3 მესამე კვარტილის რიცხვით მნიშვნელობებს Excel-ი გამოითვლის შემდეგი ალგორითმით: ვარიაციული მწკრივის k-ური ელემენტის რიცხვითი მნიშვნელობა აღვნიშნოთ $x[k]$-თი. თუ მონაცემთა რაოდენობაა n, მაშინ

$$Q_1 = x\left[\frac{n}{4}+\frac{3}{4}\right] + \frac{\left(x\left[\left(\frac{n}{4}+\frac{3}{4}\right)+1\right] - x\left[\left(\frac{n}{4}+\frac{3}{4}\right)\right]\right)}{4}$$

$$Q_3 = x\left[\frac{3n}{4}+\frac{1}{4}\right] + \frac{3\left(x\left[\left(\frac{3n}{4}+\frac{1}{4}\right)+1\right] - x\left[\left(\frac{3n}{4}+\frac{1}{4}\right)\right]\right)}{4}$$

მონაცემთა (1) სიმრავლისათვის მესამე რიგის კვარტილის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

QUARTILE(Array,Quart)=QUARTIL(B2:I6,3).

გამოთვლის შედეგად მივიღებთ, რომ მესამე რიგის კვარტილი, ანუ 75 პროცენტილი 28.8075-ის ტოლია.

SMALL ფუნქციის ჩაწერის სინტაქსი ასეთია:

SMALL(Array,k).

SMALL– გამოიანგარიშება უმცირესი k-ური წევრი.

მონაცემთა (1) სიმრავლისათვის რიგით მეორე უმცირესი მნიშვნელობის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

SMALL(Array,k)=SMALL (B2:I6,2).

გამოთვლის შედეგად მივიღებთ, რომ კლებად ვარიაციულ მწკრივში მეორე ელემენტი 9.08-ის ტოლია.

STDEV ფუნქციის ჩაწერის სინტაქსი ასეთია:

STDEV (Number1, Number2,..).

STDEV – გამოიანგარიშება შერჩევის სტანდარტული გადახრა (კვადრატული ფესვი შესწორებული შერჩევითი დისპერსიიდან) ფორმულით $\left(\frac{1}{n-1}\sum(x_i-\bar{x})^2\right)^{1/2}$

მონაცემთა (1) სიმრავლისათვის შერჩევის სტანდარტული გადახრის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

STDEV (Number1)= STDEV (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ შერჩევის სტანდარტული გადახრა 10.54-ის ტოლია.

STDEVP ფუნქციის ჩაწერის სინტაქსი ასეთია:

STDEVP(Number1, Number2,..).

STDEVP – გამოიანგარიშება პოპულაციის სტანდარტული გადახრა (კვადრატული ფესვი შერჩევითი დისპერსიიდან) ფორმულით $\left(\frac{1}{n}\sum(x_i-\bar{x})^2\right)^{1/2}$.

მონაცემთა (1) სიმრავლისათვის პოპულაციის სტანდარტული გადახრის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

STDEVP(Number1)=STDEVP (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ პოპულაციის სტანდარტული გადახრა 10.41-ის ტოლია.

SKEW ფუნქციის ჩაწერის სინტაქსი ასეთია:

SKEW Number1, Number2,..).

SKEW - გამოიანგარიშება შერჩევითი ექსცესის კოეფიციენტი ფორმულით $\frac{\sum(x-\bar{x})^4}{ns^4}-3$.

მონაცემთა (1) სიმრავლისათვის ექსცესის კოეფიციენტის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

SKEW (Number1)=SKEW (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ შერჩევითი ექსცესის კოეფიციენტი 1.154-ის ტოლია.

VAR ფუნქციის ჩაწერის სინტაქსი ასეთია:

VAR(Number1, Number2,..).

VAR– გამოიანგარიშება შესწორებული შერჩევითი დისპერსია ფორმულით:

$$s'^2_n=\frac{1}{n-1}\sum(x_i-\bar{x})^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n(n-1)}.$$

მონაცემთა (1) სიმრავლისათვის შესწორებული შერჩევითი დისპერსიის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

VAR(Number1)=VAR(B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ შესწორებული შერჩევითი დისპერსია 111.38-ის ტოლია.

VARP ფუნქციის ჩაწერის სინტაქსი ასეთია:

VARP (Number1, Number2).

VARP – გამოიანგარიშება შერჩევითი დისპერსია (პოპულაციის დისპერსია) ფორმულით:

$$s_n^2=\frac{1}{n}\sum(x_i-\bar{x})^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n^2}.$$

მონაცემთა (1) სიმრავლისათვის პოპულაციის დისპერსიის გამოსაანგარიშებლად ფუნქცია ასე ჩაიწერება:

VARP (Number)=VARP (B2:I6).

გამოთვლის შედეგად მივიღებთ, რომ შერჩევითი დისპერსია 108.38-ის ტოლია.

ქვევით, ცხრილში წარმოდგენილია სტატისტიკური ქვეპროგრამების **Descriptive Statistics, Histogram** ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა

Array	სიმრავლე, ცხრილი
Data array	მონაცემთა სიმრავლე
Quart	მეოთხედი
Number	მნიშვნელობათა (რიცხვითი) სიმრავლე
Value	მნიშვნელობათა სიმრავლე
Grouped	დაჯგუფებული
Labels	წარწერა, მონიშვნა
Chart Output	გადამრთველების გააქტიურებით ამოიბეჭდება მონაცემთა სიხშირული განაწილების ცხრილი და ჰისტოგრამის გრაფიკი
Cumulative Percentage	გადამრთველების გააქტიურებით დამატებით ამოიბეჭდება დაგროვილი სიხშირული განაწილების ცხრილი და კუმულატას გრაფიკი
Pareto(Sorter histogram)	გადამრთველების გააქტიურების შემთხვევაში აღნიშნული მონაცემები და შესაბამისი გრაფიკი გადალაგებული იქნება სიხშირეთა ზრდადობით

Descriptive Statistics ქვეპროგრამა (1) მონაცემების საფუძველზე გამოითვლის მონაცემთა შემდეგ რიცხვით მახასიათებლებს

Column1	
Mean / საშუალო არითმეტიკული	22.28
Standard Error / სტანდარტული შეცდომა	1.67
Median / მედიანა	20.71
Mode / მოდა	30.00
Standard Deviation / სტანდარტული გადახრა-	10.54
Sample Variance / შესწორებული შერჩევითი დისპერსია	111.16
Kurtosis / ასიმეტრიის შერჩევითი კოეფიციენტი	2.53
Skewness / ექსცესის შერჩევითი კოეფიციენტი	1.15
Range / გაბნევის დიაპაზონი	53.28
Minimum / მინიმუმი	5.99
Maximum / მაქსიმუმი	59.27
Sum / ჯამი	891.29
Count / მონაცემთა რაოდენობა	40.00
Largest(3) / მესამე უდიდესი მნიშვნელობა)	39.48
Smallest(2) / მეორე უმცირესი მნიშვნელობა	9.08
Confidence Level(95.0%) ნდობის ინტერვალი*	3.37

ახლა ავაგოთ ჰისტოგრამა. ჰისტოგრამის ასაგებად ავირჩიოთ მონაცემთა დაჯგუფების შემდეგი ექვსი ინტერვალი

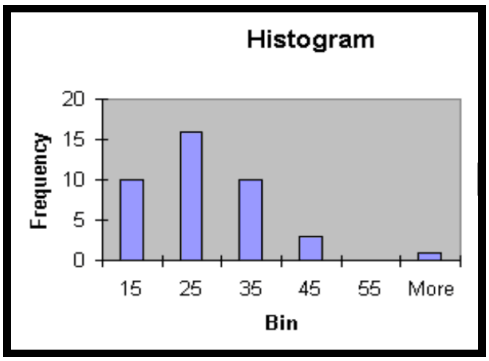
[min, 15],(15,25],(25,35],(35,45],(45,55],(55,max].

აქვე აღვნიშნოთ, რომ არაა აუცილებელი გაბნევის დიაპაზონი დაიყოს ერთი და იგივე სიგრძის ქვეინტერვალებად. Excel-ის სამუშაო დავთრის ფურცლში ჩავწეროთ აღნიშნული დაჯგუფების განმსაზღვრავი შემდეგი (ხაზგასმული) რიცხითი მნიშვნელობები: ვთქვათ 15,25,35,45,55. ქვეპროგრამის რეალიზაციით ვღებულობთ სიხშირული განაწილების ცხრილს და ჰისტოგრამის შემდეგ გრაფიკულ გამოსახულებას.

Bin	Frequency
15	10
25	16
35	10
45	3
55	0
More	1

სიხშირული განაწილების ცხრილი გვაძლევს შემდეგ ინფორმაციას:
მონაცემთა რაოდენობა, რომლებიც ნაკლები ან ტოლია 15-ზე 10-ის ტოლია;
მონაცემთა რაოდენობა, რომლებიც მეტია 15-ზე და ნაკლები ან ტოლია 25-ზე 16-ის ტოლია;
....
მონაცემთა რაოდენობა, რომლებიც მეტია 55-ზე ერთის ტოლია.

ხოლო ჰისტოგრამის გრაფიკს შემდეგი სახე აქვს



კითხვები თვითშემოწმებისათვის

რა იგულისხმება მონაცემების ქვეშ?
რას წარმოადგენს სტატისტიკური პოპულაცია?
როგორ განისაზღვრება შერჩევა, შერჩევის მოცულობა? რას წარმოადგენს ფარდობით სიხშირეთა განაწილება?
როგორ გამოითვლება შერჩევითი საშუალო და რა თვისებები აქვს მას?
როგორ გამოითვლება შერჩევითი მედიანა, შერჩევითი მოდა?
როგორ განისაზღვრება შერჩევის გაბნევის დიაპაზონი?
რას წარმოადგენს შერჩევის დისპერსია და როგორ გამოითვლება იგი?
როგორ აიგება ჰისტოგრამა, პოლიგონი?

სავარჯიშოები

4.1. ბავშვთა ფსიქოლოგს აინტერესებს, რა დროს ხარჯავენ ბავშვები ტელეგადაცემათა საყურებლად. ის აგროვებს მონაცემებს 100 შემთხვევით შერჩეული ბავშვისაგან. ეს შერჩევაა, თუ პოპულაციაა?

4.2. ახსენით, შერჩევაა თუ პოპულაცია ქვემოთ ჩამოთვლილი მონაცემები:

(a) 40 სტუდენტს, რომლებიც შემთხვევით გააჩერეს უნივერსიტეტის ეზოში, შეკითხვები დაუსვეს და ჩაწერეს პასუხები;

(b) კითხვები დაეგზავნა 200 ადამიანს, რომლებიც შემთხვევით აარჩიეს ტელეფონების ნომრების წიგნიდან და მიღებულია პასუხები;

(c) ორი კარტის ყველა შესაძლო არჩევათა სია 52 კარტიდან.

4.3. ახსენით, უწყვეტია თუ დისკრეტული შემდეგი მონაცემები:

(a) 20 მასწავლებლის წლიური შემოსავლები;

(b) საქალაქთაშორისო სატელეფონო საუბრების ოდენობა მოცემული ფოსტიდან ყოველ თვეში;

(c) თითოეული საუბრის ხანგრძლივობა.

4.4. ვთქვათ, მონაცემთა რაიმე სიმრავლისათვის მოცემულია

$$\sum_{i=1}^{n} x_i = 110, \qquad \sum_{i=1}^{n} x_i^2 = 3150, \qquad n=10.$$

ა) გამოთვალეთ $\bar{x}$, s^2, s;

ბ) შესაძლებელია თუ არა ამ მონაცემებით მედიანის გამოთვლა?

4.5. ა) ვთქვათ, $n=9$ და მონაცემთა სიმრავლის შესახებ გაგვაჩნია შემდეგი ინფორმაცია: სიდიდით მეხუთე ელემენტი ე.ი. $x_{(5)}=17$. შესაძლებელია თუ არა მედიანის გამოთვლა?

ბ) ვთქვათ, $n=20$ და $x_{(10)}=14$, $x_{(11)}=17$. რას უდრის მედიანა და რა პოზიცია უკავია მას?

4.6. განვიხილოთ მონაცემთა ორი A და B სიმრავლე

A	1	1	2	2	2	2	2	3	10
B	1	1	2	2	2	2	2	3	100

გამოთვალეთ თითოეული მათგანისათვის საშუალო, მედიანა, გაბნევის დიაპაზონი, Q_1, Q_3, სტანდარტული გადახრა;

4.7. ვთქვათ, მონაცემთა რაიმე სიმრავლისათვის გამოთვლილია შემდეგი სიდიდეები:

$x_{\min}=20$, $Q_1=40$, $Q_2=45$, $Q_3=50$, $x_{\max}=60$.

გააკეთეთ სქემატური დასკვნა ჰისტოგრამის ფორმის შესახებ.

შემდგომი ამოცანები შეგიძლიათ ამოხსნათ Excel-ის შესაბამისი ფუნქციების გამოყენებით

4.8. ერთ-ერთი უნივერსიტეტის გამოყენებითი მათემატიკის ფაკულტეტის 25 თანამშრომლის განაწილება თანამდებობების მიხედვით შემდეგია:

ლაბორანტი	8
ადმინისტრაციული პერსონალი	4
მასწავლებელი	6
უფროსი მასწავლებელი	4

დოცენტი	2
პროფესორი	1

ააგეთ თანამშრომელთა თანამდებობების განაწილების წრიული დიაგრამა.

4.9. შემდეგი მონაცემები იძლევიან გარკვეული ასაკის ბავშვების სიტყვათა მარაგის მოცულობას:

205, 377, 292, 300, 179, 240, 300, 190, 680, 250,
280, 170, 211, 266, 303, 350, 375, 288, 360, 225.

ააგეთ სიხშირეთა და ფარდობით სიხშირეთა განაწილება.

4.10. ტურისტულმა ფირმამ გამოკითხა სამშობლოში დაბრუნებული ტურისტები, თუ რამდენი ფოტო-ფირი დახარჯეს მოგზაურობის დროს, პასუხები შემდეგია:

8 6 3 11 14 8 9 16 9 10 5 11 7 8 10 9 12 13 9

ა) გამოთვალეთ საშუალო, მედიანა, Q_1, Q_3;
ბ) გამოთვალეთ გაბნევის დიაპაზონი, IQR, შერჩევითი დისპერსია, სტანდარტული გადახრა;

4.11. NED-ელექტრონიკსის მუშების მცირე რაოდენობა გამოკითხული იქნა იმ მიზნით, რომ დაედგინათ თუ როგორია ტიპიური საათობრივი ანაზღაურება ($-ში), შედეგად მიიღეს:

9.50, 9.00, 11.70, 14.80, 13.00.

ა) ეს შერჩევაა თუ პოპულაცია?
ბ) რას უდრის არითმეტიკული საშუალო, მედიანა, დისპერსია.

4.12. იმისათვის, რომ ამერიკულმა საავადმყოფოებმა მიიღონ დახმარება სხვადასხვა სამთავრობო თუ კერძო პროგრამისაგან, საჭიროა წლიური ფინანსური ანგარიშების წარდგენა, რომლის ერთ-ერთი მთავარი მახასიათებელია სრული რაოდენობა საწოლებისა, რომლებიც განკუთვნილია პაციენტებისათვის. ქვემოთ მოყვანილია მონაცემები იმ 54 საავადმყოფოში საწოლების რაოდენობის შესახებ, რომელთა ფინანსური ანგარიშიც დამაკმაყოფილებლად იქნა მიჩნეული.

303	550	243	282	195	310	288	188	190
335	473	169	292	492	200	478	182	172
231	375	171	262	198	313	600	264	311
371	145	242	278	183	215	719	519	382
249	350	99	218	300	450	337	330	252
400	514	427	533	930	319	210	550	488

ააგეთ ფარდობით სიხშირეთა განაწილება დაჯგუფების 10 ინტერვალით.

4.13. ამოხსენით 1.12 სავარჯიშო დაჯგუფების სამი ინტერვალისათვის. შეადარეთ მიღებული განაწილება წინას. რომელი უფრო მეტ ინფორმაციას შეიცავს? რატომ ზღუდავს ინტერვალების შეუსაბამო რაოდენობა ფარდობით სიხშირეთა განაწილებაში არსებულ ინფორმაციას?

4.14. ამოხსენით 1.12 სავარჯიშო, ოღონდ უკვე 25 დაჯგუფების ინტერვალისათვის. შეადარეთ შედეგი წინა ორ სავარჯიშოს.

4.15. ქვემოთ მოყვანილია 32 სააქციო საზოგადოების აქციათა მფლობელების მოგებათა პროცენტების ცხრილი გადასახადების გადახდის შემდეგ.

10.6	10.8	14.8	10.8
12.5	6.0	10.7	11.0
14.6	6.0	12.8	10.1
7.9	5.9	10.0	10.6
10.8	16.2	18.4	10.7
10.6	13.3	8.7	15.4
6.5	10.1	8.7	7.5
11.9	9.0	12.0	9.1

(a) ამ მონაცემებით ააგეთ ფარდობით სიხშირეთა ჰისტოგრამა შვიდი ტოლი ინტერვალით, რომელთაგან თითოეულის სიგრძეა 2 და პირველის მარცხენა ბოლოა 5.55;

(b) ააგეთ ფარდობით სიხშირეთა პოლიგონი;

(c) როგორია თქვენი აზრით იმის შანსი, რომ შემთხვევით არჩეული სააქციო საზოგადოების წევრის მოგების პროცენტი 15.55-ზე მეტია?

(d) როგორია იმის შანსი, რომ შემთხვევით არჩეული სააქციო საზოგადოების წევრის მოგების პროცენტი 9.55-ზე ნაკლებია?

(e) რა შანსი არსებობს იმისა, რომ შემთხვევით არჩეული სააქციო საზოგადოების წევრის მოგების პროცენტი 9.55-სა და 15.55-ს შორისაა?

4.16. შემდეგი მონაცემები შეადგენენ მსოფლიოს 57 უდიდესი ჰიდროელექტროსადგურის სიმძლავრეებს მეგავატებში:

12600	4500	2700	2400	2100	1824	1950
10080	4150	2700	2400	2069	1807	1890
10060	4050	2700	2304	2031	1800	2100
6480	3600	2680	2300	2030	1800	2100
6400	3575	2650	2300	2000	1800	2500
6096	3409	2637	2250	2000	1800	2416
6000	3300	2610	2124	2000	1750	2715
5328	3200	2560	2100	1979	4600	2820
5225						

(a) ააგეთ ფოთლებიანი ღეროების მსგავსი დიაგრამა (ფოთლის ერთეული იყოს 10 მეგავატი, დაამრგვალეთ, ორი ნიშანი დაუტოვეთ ფოთოლს, დანარჩენი ღეროს მიაკუთვნეთ), მიიღებთ 12 ჰორიზონტალურ ღეროს. რა ინტერვალშია მოთავსებული მონაცემთა ძირითადი უმრავლესობა?

(b) ააგეთ ფოთლებიანი ღეროების მსგავსი დიაგრამა ისე, რომ (a) დავალების პასუხად უკვე აგებული თითოეული ჰორიზონტალური ღერო დაყოთ ორ ღეროდ: პირველი უნდა შედგებოდეს ფოთლებისაგან, რომელთა პირველი ციფრია რომელიმე 0-იდან 4-ის ჩათვლით ხოლო მეორე – ფოთლებისაგან, რომელთა საწყისი ციფრია რომელიმე დანარჩენთაგან. პირველი ღერო აღვნიშნოთ L-ით, მეორე – H-ით, მაგალითად, 2L და 2H ღეროები.

4.17. ოცდარვა პირი მონაწილეობდა სახელმწიფო პროგრამით სოციალურ სამუშაოზე თანამდებობათა დასაკავებელ კონკურსში. ტესტირების ჩატარების შედეგად მიღებულია შემდეგი ქულები:

79	97	86	76
93	87	98	68

84	88	81	91
86	87	70	94
77	92	66	85
63	68	98	88
46	72	59	79

(a) ააგეთ ფარდობით სიხშირეთა ჰისტოგრამა 9 ქვეინტერვალით, პირველის მარცხენა ბოლო იყოს 44.5;

(b) ააგეთ ფოთლებიანი ღეროების მსგავსი დიაგრამა. იგივეა თუ არა ფორმა? რა დამატებით ინფორმაციას შეიცავს ფოთლებიანი ღეროების მსგავსი დიაგრამა ჰისტოგრამასთან შედარებით?

4.18. ქვემოთ მოყვანილ ცხრილში მოცემულია აშშ-ს 13 შტატის შემოსავლები ერთ სულ მოსახლეზე 1983 წელს (მონაცემები გამოსახულია დოლარებში)

9999, 9342, 11769, 12580, 9560, 10719, 9031, 12516, 12051,10920,13239,16820, 12101.

ა) მოძებნეთ საშუალო, მედიანა, Q_1, Q_3, $x_{\min}$, $x_{\max}$;
ბ) ფიქრობთ თუ არა, რომ ამ შემთხვევაში საშუალო უკეთ ახასიათებს ტიპობრივ შემოსავლებს, ვიდრე მედიანა?
გ) გამოთვალეთ $Q_1 - x_{\min}$, $Q_2 - Q_1$, $Q_3 - Q_2$, $x_{\max} - Q_3$ (გავიხსენოთ, რომ Q_2 არის შერჩევითი მედიანა);
დ) როგორი იქნებოდა ეს სხვაობები, მონაცემები რომ ($\bar{x}$-ის მიმართ) ზუსტად სიმეტრიულად ყოფილიყო განაწილებული?
ე) გ) პუნქტში გამოთვლილი სიდიდეების მიხედვით შეგიძლიათ თუ არა დაასკვნათ, რომ მონაცემები სიმეტრიულადაა განაწილებული?

4.19. აშშ-ში შტატების მიხედვით წლიური შემოსავალი ერთ სულ მოსახლეზე ($1000) არის

11.1, 17.7, 13.2, 10.7, 16.8, 15.1, 19.2, 15.1, 18.9, 14.3, 13.2, 14.7, 11.4, 15.4, 12.9, 13.2, 14.4,

11.1, 11.2, 12.7, 16.6, 17.5, 14.1, 14.7, 9.5, 13.6, 11.9, 13.8, 15.1, 15.9, 18.3, 11.1, 17.1, 12.2,

12.3, 18.7, 12.4, 12.2, 13.9, 14.7, 11.1, 11.9, 11.8, 13.5, 10.7, 12.8, 15.4, 14.5, 10.5, 13.8, 13.2.

ა) ააგეთ სიხშირეთა განაწილება;
ბ) გამოთვალეთ საშუალო, მედიანა, Q_1, Q_3, $x_{\min}$, $x_{\max}$; 20-პროცენტილი, პირველი და ბოლო დეცილები;
გ) გამოთვალეთ გაბნევის დიაპაზონი, დისპერსია და სტანდარტული გადახრა;
დ) გამოთვალეთ ასიმეტრიისა და ექსცესის კოეფიციენტები;
ე) არის თუ არა განაწილება სიმეტრიული?
შეაჯამეთ მიღებული შედეგები.

4.20. შემთხვევით შერჩეული 100 მანქანისათვის დადგენილია მანძილი (მილებში), რომელსაც მანქანა გაივლის 1 გალონი ბენზინის მოხმარებისას. მონაცემები დაჯგუფებულია შემდეგი ცხრილის მიხედვით:

ბენზინის მოხმარება	[14,16)	[16,18)	[18,20)	[20,22)	[22,24)
მანქანათა რაოდენობა	9	18	24	38	16

ააგეთ სიხშირეთა ჰისტოგრამა.

4.21. გარკვეულ რეგიონში ოჯახური შემოსავლების შესწავლის მიზნით შემთხვევით შეარჩიეს 12 ოჯახი, რომელთა წლიური შემოსავლების მონაცემები (ათასობით $-ში) შემდეგია:

19.9 23.5 33.5 17.45 19 21 37.7 41.4 17.25 18 19.24 22.21

ა) გამოთვალეთ საშუალო, მედიანა. რომელი მათგანი უფრო გამოდგება ტიპიური შემოსავლების დასახასიათებლად?

ბ) გამოთვალეთ Q_1, Q_3, $x_{\min}$, $x_{\max}$; პირველი და მეცხრე დეცილები;

გ) გამოთვალეთ გაბნევის დიაპაზონი, IQR, დისპერსია, სტანდარტული გადახრა;

4.22. NYSE-ს აქციების შერჩევისათვის მიღებული იქნა ფასი/ამონაგები ფარდობის შემდეგი მონაცემები:

15.3 9.7 12.4 21.0 62.3 9.5 5.3 8.6 29.7 43.4

გამოთვალეთ საშუალო, მედიანა, გაბნევის დიაპაზონი, Q_1, Q_3, $\bar{x}$, $\tilde{x}$, დისპერსია, სტანდარტული გადახრა, ასიმეტრიისა და ექსცესის კოეფიციენტები;

4.23. დიდ საინვესტიციო ფირმას სურს შეისწავლოს მისი აქციების ბროკერთა ასაკების განაწილება. ფირმა ფიქრობს, რომ ეს ინფორმაცია სასარგებლო იქნება ახალი კადრებით შევსებისა და ძველი კადრების გათავისუფლების პოლიტიკის შესამუშავებლად. 25 ბროკერისაგან შედგენილ შერჩევაში ასაკების მონაცემები შემდეგია:

50 64 32 55 41 44 24 46 58 47 36 52 55 54 44 66 47 59 51 57 49 28 42 38 45

ა) გამოთვალეთ მედიანა, Q_1, Q_3, მოდა, $x_{\min}$, $x_{\max}$;

ბ) გამოთვალეთ საშუალო, დისპერსია, სტანდარტული გადახრა;

გ) ააგეთ სიხშირეთა განაწილება (მონაცემები დაყავით 5 ინტერვალად a_0=20).

4.24. ფასი/ამონაგები ფარდობა (P/E) დიდ როლს თამაშობს საინვესტიციო ანალიზში, რადგან ის ერთდროულად შეიცავს ინფორმაციას აქციის რისკიანობისა და მისი გაძლიერების შესაძლებლობების შესახებ. ქვემოთ მოყვანილია ავსტრალიის 30 ინდუსტრიული აქციის P/E ფარდობა (1992წ., 31 მაისი):

15.1 15.3 20.4 14.1 11.7 17.2 30.3 17.2 11.2 8.6
25.1 17.8 10.6 22.1 18.1 18.7 21.2 18.8 17.6 19.2
14.4 11.3 21.1 20.4 20.5 19.4 17.2 9.7 15.7 21.9

ა) გამოთვალეთ საშუალო და სტანდარტული გადახრა;

ბ) ააგეთ ჰისტოგრამა;

4.25. შემდეგ ცხრილში მოყვანილია აშშ-ს ერთობლივი ნაციონალური პროდუქტის მონაცემები 1970–1984 წლებში (მილიარდ $-ში). საბაზისო წელია 1972.

1970	1085.6	1975	1231.6	1980	1475.0
1971	1122.4	1976	1298.2	1981	1512.8
1972	11.85.9	1977	1369.7	1982	1480.0
1973	1254.3	1978	1438.6	1983	1534.7
1974	12.46.3	1979	1479.4	1984	1639.0

ა) გამოთვალეთ საშუალო, მედიანა, დისპერსია და სტანდარტული გადახრა;

ბ) გამოთვალეთ ასიმეტრიის კოეფიციენტი, არის თუ არა მონაცემები მარცხნივ ასიმეტრიული?

ლექცია V,VI სტატისტიკური პოპულაცია და სტატისტიკა

§ 1. ძირითადი ცნებები

დავუშვათ გვაქვს ელემენტთა რაიმე სიმრავლე, რომლის ყოველი წევრი ხასიათდება რაოდენობრივი ან თვისებრივი მახასიათებლის კონკრეტული მნიშვნელობით.

სტატისტიკური პოპულაცია არის კონკრეტული მახასიათებლის ყველა შესაძლო მნიშვნელობის ერთობლიობა.

ვიტყვით, რომ პოპულაცია **სრულადაა მოცემული**, თუ შესაძლებელია ელემენტთა აღნიშნული სიმრავლის ყველა ელემენტის გადარჩევა და მათთვის მახასიათებლის რიცხვითი მნიშვნელობის განსაზღვრა.

უნდა აღინიშნოს, რომ სტატისტიკურ პოპულაციაზე საუბრისას (ზედსართავი სტატისტიკური ქვემოთ ხშირად გამოტოვებული იქნება) არ იგულისხმება, რომ ყველა შესაძლო მნიშვნელობა მართლაც დაკვირვებულია. პოპულაცია შეიძლება იყოს როგორც რეალური, ისე წარმოსახვითი ანუ კონცეპტუალური.

პოპულაციის საპირისპიროდ **შერჩევა ანუ ამოკრეფა** მხოლოდ ზოგიერთ დაკვირვებულ მნიშვნელობას შეიცავს.

შერჩევა ანუ ამოკრეფა არის დაკვირვებათა ერთობლიობა, რომელიც შეადგენს პოპულაციის მხოლოდ რაიმე ნაწილს.

სტატისტიკური კვლევის საბოლოო მიზანი პოპულაციის შესწავლაა.
პოპულაციასა და შერჩევაში შემავალ მონაცემთა რაოდენობას მისი **მოცულობა** ჰქვია.
პოპულაციის მოცულობა **სასრულიც** შეიძლება იყოს და, როცა პოპულაცია წარმოსახვითია, **უსასრულოც**. შერჩევა ყოველთვის სასრული მოცულობისაა.

შემთხვევითი სიდიდის ცნების შემოღების შემდეგ შესაძლებელი ხდება პოპულაციის აღწერა რაიმე ***X* შემთხვევითი სიდიდის** მეშვეობით. მაგალითად, თუ გვაქვს სასრული სიმრავლე, რომლის სრული აღწერა გვაძლევს შემდეგ სურათს: პოპულაცია შედგება რაოდენობრივი მახასიათებლის N მნიშვნელობისაგან, რომელთა შორის არის k განსხვავებული მნიშვნელობა. ეს მნიშვნელობები იყოს $x_1,x_2,...,x_k$. ვთქვათ, x_i-ის სიხშირეა n_i (ე.ი. n_i არის პოპულაციის იმ ელემენტების რაოდენობა, რომელთა მნიშვნელობებიც x_i-ის ტოლია), $\sum_{i=1}^{k} n_i = N$, მაშინ პოპულაცია აღიწერება დისკრეტული X შემთხვევითი სიდიდით, რომლის განაწილების კანონია

X	x_1	x_2	. . .	x_k
	p_1	p_2	. . .	p_k

სადაც

$$p_i = P\{X=x_i\} = \frac{n_i}{N},\ i=1,2,...,k.$$

ეს დაშვება ბუნებრივია იმ თვალსაზრისით, რომ პოპულაციიდან შემთხვევით არჩეული ელემენტის მიერ x_i მნიშვნელობის მიღების ალბათობა n_i/N შეფარდების ტოლია.

შემდგომში **პოპულაციას გავაიგივებთ მის აღმწერ *X* შემთხვევით სიდიდესთან და პოპულაციის განაწილების ქვეშ ვიგულისხმებთ *X*-ის განაწილების კანონს**. ამასთან, განაწილების კანონი შეიძლება მოცემული იყოს ნებისმიერი ფორმით: დისკრეტული

განაწილების კანონით, განაწილების ფუნქციით, განაწილების სიმკვრივით და სხვა. პოპულაციას აღმწერი X შემთხვევითი სიდიდის რიცხვით მახასიათებლებს (საშუალოს, დისპერსიას, სტანდარტულ გადახრას და სხვა) **პოპულაციის პარამეტრები** ეწოდებათ.

დავუშვათ X შემთხვევითი სიდიდის დაკვირვებული მნიშვნელობებია $x_1,x_2,...,x_n$: x_1 არის პირველი დაკვირვებული მნიშვნელობა, x_2 – მეორე და ა.შ. დაკვირვებების მოპოვებამდე ყოველი კერძო დაკვირვება განუსაზღვრელია (დაკვირვების ანუ ცდის შედეგის წინასწარ განჭვრეტა შეუძლებელია), რის გამოც, მონაცემების მოპოვებამდე დაკვირვებებს განვიხილავთ, როგორც შემთხვევით სიდიდეებს და აღვნიშნავთ $X_1,X_2,...,X_n$ სიმბოლოებით, რეალურად დაკვირვებულ $x_1,x_2,...,x_n$ მნიშვნელობებს კი მივიჩნევთ $\boldsymbol{X}=(X_1,X_2,...,X_n)$ შემთხვევითი ვექტორის კონკრეტულ რეალიზაციად.

ვიტყვით, რომ $\boldsymbol{X}=(X_1,X_2,...,X_n)$ **შემთხვევითი ვექტორი** წარმოადგენს n მოცულობის შემთხვევით შერჩევას (ან, უბრალოდ, შერჩევას) მოცემული პოპულაციიდან (გენერალური ერთობლიობიდან), თუ $X_1,X_2,...,X_n$ წარმოადგენს დამოუკიდებელ ერთნაირად განაწილებულ შემთხვევით სიდიდეებს, რომელთა საერთო განაწილება ემთხვევა პოპულაციის განაწილებას.

ხშირად გამოიყენება შემდეგი ტერმინოლოგიაც: $X_1,X_2,...,X_n$ მიმდევრობა წარმოადგენს დამოუკიდებელ დაკვირვებებს X შემთხვევით სიდიდეზე, რომლის განაწილების კანონია $F(x)$.

პირველ თავში ჩვენ გავეცანით რიცხვითი მონაცემების (შერჩევითი მნიშვნელობების) წარმოდგენისა და დამუშავების ხერხებს. შერჩევითი მნიშვნელობების (მონაცემების) საფუძველზე ვითვლიდით სხვადასხვა მახასიათებლებს; ყოველი მახასიათებელი იყო $\boldsymbol{x}=(x_1,x_2,...,x_n)$ შერჩევით მნიშვნელობათა საფუძველზე გამოთვლილი $T(x_1,x_2,...,x_n)$ რიცხვითი ფუნქცია. მაგრამ, როგორც აღვნიშნეთ, $\boldsymbol{x}=(x_1,x_2,...,x_n)$ ვექტორი არის $\boldsymbol{X}=(X_1, X_2,...,X_n)$ შემთხვევითი ვექტორის ერთ-ერთი შესაძლო მნიშვნელობა, ამიტომ $T_n(\boldsymbol{x})=T(x_1,x_2,...,x_n)$ სიდიდე წარმოადგენს $T_n(\boldsymbol{X})=T(X_1,X_2,...,X_n)$ შემთხვევითი სიდიდის ერთ-ერთ დაკვირვებულ მნიშვნელობას $T_n(\boldsymbol{X})$-ის მნიშვნელობათა სიმრავლიდან.

შერჩევითი მნიშვნელობების ნებისმიერ $T_n(\boldsymbol{x})=T(x_1,x_2,...,x_n)$ ფუნქციას, რომლის აგების ალგორითმი ცნობილია, **<u>სტატისტიკას (შემფასებელს)</u>** უწოდებენ.

<u>სტატისტიკას ჩვენ შევხედავთ ორნაირად</u>:

- როგორც რიცხვს; თუ T_n ფუნქციის არგუმენტად ავიღებთ $x_1,x_2,...,x_n$ შერჩევით მნიშვნელობებს. $t_n=T(x_1,x_2,...,x_n)$ სტატისტიკის <u>დაკვირვებული (რეალიზებული) მნიშვნელობაა.</u>
- როგორც შემთხვევით სიდიდეს; თუ T_n ფუნქციის არგუმენტად ავიღებთ $X_1,X_2,...,X_n$ შერჩევას, მაშინ იგივე $T_n=T(X_1,X_2,...,X_n)$ ფუნქცია წარმოადგენს <u>შემთხვევით სიდიდეს</u>.

სხვა სიტყვებით რომ ვთქვათ, სტატისტიკა წარმოადგენს სიდიდეს, რომლის მნიშვნელობა შეიძლება გამოითვალოს შერჩევითი მონაცემების საშუალებით. მონაცემების მიღებამდე სტატისტიკები განიხილება როგორც შემთხვევითი სიდიდეები, დაკვირვებების ჩატარების შემდეგ კი როგორც რიცხვები.

ასე მაგალითად, ყოველი დაკვირვებული $x_1,x_2,...,x_n$ მნიშვნელობებისათვის $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ -ს ($s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$) რიცხვითი მნიშვნელობა არის შესაბამისი წესით აგებული $\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n}X_i$ ($S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$) შემთხვევითი სიდიდის დაკვირვებული მნიშვნელობა.

სტატისტიკას, განხილულს როგორც შემთხვევით სიდიდეს, გააჩნია **თავისი განაწილების კანონი**, რომელსაც ეწოდება **სტატისტიკის შერჩევითი განაწილება**. ყოველი კერძო სტატისტიკის განაწილება დამოკიდებულია არა მხოლოდ შერჩევის n მოცულობასა და პოპულაციის განაწილებაზე, არამედ იმ მეთოდზეც, რომლითაც წარმოებს შერჩევა.

შემდგომში დავუშვებთ, რომ პოპულაციიდან ამოკრეფა ხდება ისეთი წესით, რომ n მოცულობის $X_1,X_2,...,X_n$ შერჩევა წარმოადგენს **დამოუკიდებელ ერთნაირად განაწილებულ შემთხვევით სიდიდეებს, რომელთა საერთო განაწილება ემთხვევა პოპულაციის განაწილებას.** მაგალითად, ეს პირობები სრულდება, როდესაც შერჩევა წარმოებს ან უსასრულო (წარმოსახვითი) პოპულაციიდან ანდა სასრული პოპულაციიდინ დაბრუნებით (მონეტის ან კამათლის n-ჯერ აგდება, რაიმე შემთხვევითი ექსპერიმენტის n-ჯერ გამეორება და სხვა).

$\boldsymbol{X}$ შემთხვევითი **ვექტორის განაწილების** მეშვეობით უკვე შესაძლებელია სტატისტიკის განაწილებისა და მისი რიცხვითი მახასიათებლების გამოთვლა. ზოგიერთ შემთხვევაში ხერხდება <u>სტატისტიკის ზუსტი განაწილების დადგენა</u>, ხოლო ზოგიერთ შემთხვევაში კი მხოლოდ <u>ზღვარითი განაწილებისა</u>.

განვიხილოთ შემდეგი 2 შემთხვევა.

1. დავუშვათ, $X_1,X_2,...,X_n$ წარმოადგენს შერჩევას ბერნულის კანონით განაწილებული გენერალურ ერთობლიობიდან, $X_i \in Bi(1, p)$, სადაც p წარმატების ალბათობაა ერთ ცდაში (იხ. შემდეგი ლექციის შესავალი), $i=1,2,...,n$. განვიხილოთ სტატისტიკა (ერთიანების გამოჩენის ფარდობითი სიხშირე n დამოუკიდებელ ცდაში)

$$\hat{P}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i = \frac{S_n}{n}$$

და დავადგინოთ მისი განაწილების კანონი. .

რადგან $X_1,X_2,...,X_n$ შემთხვევითი სიდიდეები ღებულობენ მხოლოდ ორ მნიშვნელობას, ნულსა და ერთს, მათი ჯამი $S_n=X_1+X_2+..+X_n$, როგორც ვიცით, ბინომურად განაწილებული შემთხვევითი სიდიდეა, ე.ი.

$$P\{S_n = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0,1,..., n,$$

ამიტომ $\hat{P}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$ -ის შესაძლო მნიშვნელობებია 0, $\frac{1}{n}$, $\frac{2}{n}$,... , 1, ამასთან

$$P\left\{\hat{P}_n = \frac{k}{n}\right\} = C_n^k p^k (1-p)^{n-k}, \quad k=0,1,...,n, \quad E\hat{P}_n = p, \quad D\hat{P}_n = \frac{p(1-p)}{n}$$

$\hat{P}_n$ სიდიდის სტანდარტიზაციის შედეგად

$$P\left\{\frac{\hat{P}_n - p}{\sqrt{p(1-p)/n}} < x\right\} = \sum_{k;k<x} C_n^k p^k (1-p)^{n-k}$$

ამავე დროს, მუავრ-ლაპლასის თეორემის ძალით (იხ. ლექცია 2-3)) როცა $n\to\infty$

$$P\left\{\frac{\hat{P}_n - p}{\sqrt{p(1-p)/n}} \le x\right\} \to \Phi(x),$$

ე.ი. საკმარისად დიდი n-ისათვის $\hat{P}_n$ შემთხვევითი სიდიდე (სტატისტიკა) მიახლოებით ნორმალურადაა განაწილებული p საშუალოთი და pq/n დისპერსიით. ამრიგად ჩვენ დავადგინეთ $\hat{P}_n$-ის როგორც ზუსტი, ასევე მიახლოებითი განაწილების კანონი.

2. ახლა დავუშვათ, რომ $X_1,X_2,...,X_n$ შერჩევა მიღებულია ნორმალურად განაწილებული გენერალური ერთობლიობიდან. ამ შემთხვევაში, როგორც ადრე აღვნიშნეთ (იხ. ლექცია 2-3), დამოუკიდებელი ნორმალურად განაწილებული შემთხვევითი სიდიდეების წრფივი კომბინაცია კვლავ ნორმალურადაა განაწილებული, ანუ

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ნორმალურად განაწილებული შემთხვევითი სიდიდეა.

გავაკეთოთ ერთი მნიშვნელოვანი შენიშვნა: თუ შერჩევა **არ არის** მიღებული ნორმალური კანონით განაწილებული გენერალური ერთობლიობიდან, მაგრამ n საკმარისად დიდია (n>30) ცენტრალური ზღვარითი თეორემის თანახმად (იხ. ლექცია 2-3) შეგვიძლია ჩავთვალოთ, რომ

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

შემთხვევითი სიდიდე მიახლოებით ნორმალურადაა განაწილებული.

ამრიგად ამ $\overline{X}_n$ სტატისტიკისათვის (შერჩევითი საშუალოსათვის) ჩვენ დავადგინეთ, როგორც ზუსტი, ასევე მიახლოებითი განაწილების კანონი სხვადასხვა დაშვებებში. $\overline{X}_n$ სტატისტიკის ზუსტი ან ზღვარითი განაწილების კანონის ცოდნა **საშუალებას მოგვცემს ვისაუბროთ $\overline{X}_n$-ის ალბათურ თვისებებზე, რასაც გადამწყვეტი მნიშვნელობა აქვს სტატისტიკური დასკვნების მიღებისათვის.**

§ 2. შერჩევითი საშუალოს მათემატიკური ლოდინი და დისპერსია

ვთქვათ $X_1,X_2,...,X_n$ დამოუკიდებელი შერჩევაა პოპულაციიდან, რომლის საშუალოა μ და დისპერსიაა σ^2. გამოვთვალოთ შერჩევითი საშუალოს, $\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ -ს მათემატიკური ლოდინი და დისპერსია. მათემატიკური ლოდინისა და დისპერსიის თვისებების გათვალისწინებით მივიღებთ, რომ:

$$E\,\overline{X}_n = E\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{1}{n}\cdot n\mu = \mu,$$

ხოლო

$$D\,\overline{X}_n = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{\sigma^2}{n}.$$

ამრიგად, შერჩევითი საშუალოს მათემატიკური ლოდინი ტოლია გენერალური ერთობლიობის მათემატიკური ლოდინისა, ხოლო შერჩევითი საშუალოს დისპერსია n-ჯერ

ნაკლებია გენერალური ერთობლიობის დისპერსიაზე. შევნიშნოთ, რომ მიღებული შედეგი სამართლიანია ნებისმიერად განაწილებული პოპულაციისათვის. ამასთან, დიდ რიცხვთა კანონის ძალით, $\overline{X}_n \xrightarrow{P} \mu$, როცა $n\to\infty$.

დასკვნა:

თუ $X_1,X_2,\ldots,X_n$ დამოუკიდებელი ერთნაირად განაწილებული შემთხვევითი სიდიდეებია და $EX_i=\mu$, $DX_i=\sigma^2$ და $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$, მაშინ

$$E\,\overline{X}_n=\mu, \qquad D\,\overline{X}_n=\frac{\sigma^2}{n}, \quad \overline{X}_n \xrightarrow{P} \mu$$

§ 3. შერჩევითი დისპერსიის მათემატიკური ლოდინი

გამოვთვალოთ

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

შერჩევითი დისპერსიის მათემატიკური ლოდინი. ვინაიდან რაიმე Y შემთხვევითი სიდიდის დისპერსია $DY=EY^2-(EY)^2$ ტოლობით მოიცემა. გვაქვს:

$$E\,S_n^2=E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right).$$

როგორც ვიცით

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i^2-2\overline{X}_nX_i+\overline{X}_n^2\right)=\frac{1}{n}\sum_{i=1}^{n}X_i^2-2\overline{X}_n\frac{1}{n}\sum_{i=1}^{n}X_i+\overline{X}_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}_n^2$$

ამიტომ

$$E\,S_n^2=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}_n^2\right)=\frac{1}{n}\sum_{i=1}^{n}EX_i^2-E\overline{X}_n^2,$$

$$EX_i^2=DX_i+(EX_i)^2=\sigma^2+\mu^2,$$

$$E\,\overline{X}_n^2=D\,\overline{X}_n+(E\,\overline{X}_n)^2=\frac{\sigma^2}{n}+\mu^2,$$

საიდანაც

$$ES_n^2=\sigma^2-\frac{\sigma^2}{n}=\sigma^2\,\frac{n-1}{n}.$$

ამრიგად $E\,\overline{X}_n=\mu$ ტოლობისაგან განსხვავებით, $ES_n^2\neq\sigma^2$. ამ მიზეზის გამო განიხილავენ **შესწორებულ შერჩევით დისპერსიას**, რომელიც შემდეგი ფორმულით განისაზღვრება

$$S'^2_n=\frac{n}{n-1}S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2. \qquad (5.1)$$

ადვილი სანახავია, რომ

$$ES'^2_n=E\left(\frac{n}{n-1}S_n^2\right)=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2.$$

აქვე აღვნიშნოთ, რომ $S'^2_n \xrightarrow{P} \sigma^2$, როცა $n\to\infty$.

დასკვნა:

თუ $X_1,X_2,...,X_n$ დამოუკიდებელი ერთნაირად განაწილებული შემთხვევითი სიდიდეებია და $EX_i=\mu$, $DX_i=\sigma^2$ და S'^2 განსაზღვრულია (7.1) თანაფარდობით, მაშინ

$$E\,S'^2_n = \sigma^2, \qquad S'^2_n \xrightarrow{P} \sigma^2, \text{ როცა } n\to\infty.$$

§ 4. შერჩევითი საშუალოსა და დისპერსიის განაწილების კანონი ნორმალური პოპულაციისათვის

როგორც აღვნიშნეთ, ნებისმიერი T_n სტატისტიკის სავარაუდო ყოფაქცევას აღწერს მისი განაწილების კანონი. ამ პარაგრაფში ჩვენ მოვიყვანთ შერჩევითი საშუალოსა და შერჩევითი დისპერსიის განაწილების კანონებს მხოლოდ იმ შემთხვევისათვის, როდესაც პოპულაცია (გენერალური ერთობლიობა) ნორმალურადაა განაწილებული.

ვთქვათ, $X_1,X_2,...,X_n$ წარმოადგენს შერჩევას ნორმალური პოპულაციიდან, $X_i \sim N(\mu,\sigma^2)$, $i=1,2,...,n$. ცნობილია, რომ $\overline{X}_n$ წარმოადგენს დამოუკიდებელი ნორმალურად განაწილებული შემთხვევითი სიდიდეების წრფივ კომბინაციას

$$\overline{X}_n \sim N(\mu, \sigma^2/n). \tag{5.2}$$

გადავიდეთ S_n^2 შერჩევითი დისპერსიის განაწილების კანონის დადგენაზე. ამისათვის უპირველეს ყოვლისა დავადგინოთ, თუ როგორი განაწილება აქვს შერჩევით დისპერსიას $\overline{S}_n^2$ იმ შემთხვევაში, როდესაც პოპულაციის საშუალო ცნობილია.

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 . \tag{5.3}$$

რადგან $X_i \sim N(\mu,\sigma^2)$ მივიღებთ, რომ $(X_i-\mu)/\sigma \sim N(0,1)$, ე.ი. $\dfrac{n\overline{S}_n^2}{\sigma^2} = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2$ წარმოადგენს $N(0,1)$ კანონით განაწილებული დამოუკიდებელი შემთხვევითი სიდიდეების კვადრატების ჯამს და მაშასადამე, მას აქვს χ^2-განაწილება თავისუფლების ხარისხით n(იხ. ლექცია 1) :

$$\frac{n\overline{S}_n^2}{\sigma^2} \sim \chi^2(n). \tag{5.4}$$

შევნიშნოთ, რომ $\overline{S}_n^2$ სტატისტიკის (5.3) გამოსახულებაში $(X_i-\mu)$, $i=1,...,n$, არიან დამოუკიდებელი შემთხვევითი სიდიდეები. სულ სხვანაირადაა საქმე S^2-ისათვის: $\dfrac{nS_n^2}{\sigma^2}$ არ წარმოადგენს დამოუკიდებელი შემთხვევითი სიდიდეების კვადრატების ჯამს, $(X_i-\overline{X}_n)$ შემთხვევით სიდიდეებს **გააჩნიათ ერთი „ბმა“**; კერძოდ, $\sum_{i=1}^{n}(X_i-\overline{X}_n)=0$. ამ შემთხვევაში სამართლიანია შემდეგი ფაქტი $\dfrac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$. ამასთან, შეიძლება დამტკიცდეს, **რომ $\overline{X}_n$ და S_n^2 დამოუკიდებელი შემთხვევითი სიდიდეებია.** საბოლოოდ, ყველა ჩამოთვლილი ფაქტი შეგვიძლია ჩამოვაყალიბოთ შემდეგი **თეორემის სახით, რომელიც ეკუთვნის ფიშერს (R. A. Fisher):**

თუ $X_1,X_2,...,X_n$ შემთხვევითი შერჩევაა ნორმალურად განაწილებული გენერალური

ერთობლიობიდან μ, σ^2 პარამეტრებით, მაშინ

ა) $\overline{X}_n$ **და** S_n^2 ($S_n'^2$) **შემთხვევითი სიდიდეები დამოუკიდებელია;**

ბ) $\overline{X}_n \sim N(\mu,\sigma^2/n)$;

გ) $\frac{(n-1)S_n'^2}{\sigma^2} \sim \chi^2(n-1)$.

ნორმალური პოპულაციიდან მიღებული შერჩევისათვის განვიხილოთ შემდეგი ორი სტატისტიკა: ე.წ. ***Z* სტატისტიკა და *T* სტატისტიკა**

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}, \tag{5.5}$$

$$T_n = \frac{\overline{X}_n - \mu}{S_n'/\sqrt{n-1}}, \tag{5.6}$$

ფიშერის თეორემის ძალით ადვილად მივიღებთ, რომ Z_n სტატისტიკას აქვს სტანდარტული ნორმალური განაწილება, ხოლო T_n სტატისტიკას აქვს სტიუდენტის t განაწილება $n-1$ თავისუფლების ხარისხით. მართლაც წარმოვიდგინოთ T_n სტატისტიკა შემდეგი სახით

$$T_n = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_n'^2}{\sigma^2}\Big/(n-1)}}.$$

მაშინ ფიშერის თეორემის ძალით

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{(n-1)S_n'^2}{\sigma^2} \sim \chi^2(n-1),$$

ამასთან $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ და $\frac{(n-1)S_n'^2}{\sigma^2}$ **შემთხვევითი სიდიდეები დამოუკიდებელია**. თუ ახლა გავიხსენებთ სტიუდენტის t განაწილების განმარტებას(იხ. ლექცია 1), მივიღებთ სასურველ შედეგს.

§ 5 სტატისტიკური მოდელები და წერტილოვანი შეფასების ამოცანა

ამოცანის დასმა. დავუშვათ პოპულაციის აღმწერი განაწილების ფუნქციის ანალიზური სახე ცნობილია. ჩავთვალოთ, რომ ის შეიცავს **უცნობ θ პარამეტრს**. უცნობი θ პარამეტრის **დასაშვებ მნიშვნელობათა სიმრავლე აღვნიშნოთ Θ** – თი. ამით ჩვენ ვგულისხმობთ, რომ **$\{F(x;\theta);\theta\in\Theta\}$ განსაზღვრავს განაწილების ფუნქციათა პარამეტრულ ოჯახს.** ჩვენ ხშირ შემთხვევაში ვიხმართ ცნებას **„მოცემულია სტატისტიკური მოდელი“-- $\{F(x;\theta);\theta\in\Theta\}$ პარამეტრული ოჯახისათვის**. მოდელს ეწოდება უწყვეტი ან დისკრეტული, თუ პოპულაციის განაწილება შესაბამისად უწყვეტია ან დისკრეტული.

ჩამოვთვალოთ ძირითადი პარამეტრული მოდელები, რომლებსაც ჩვენ შევხვდებით მომავალში და მივუთითოთ, თუ რა სიტუაციაში იყენებენ აღნიშნულ მოდელებს.

1) **ბინომური მოდელი** ($X\in Bi(n,\ p)$ მოდელი) : X შემთხვევითი სიდიდე განაწილებულია ბინომურად ცდათა n რიცხვითა და წარმატების უცნობი p ალბათობით, $0< p <1$. $EX=np$.

ამ მოდელის კერძო შემთხვევაა ბერნულის სქემა - $X\in Bi(1,\ p)$ მოდელი. $EX=p$.

ბინომური მოდელის წარმოშობა კარგადაა ცნობილი დისკრეტული ალბათობიდან.

2) **პუასონის მოდელი** ($X\in\Pi(\lambda)$ მოდელი): X შემთხვევითი სიდიდე განაწილებულია პუასონის კანონით, რომლის λ პარამეტრი უცნობია, $\lambda = EX$.

პუასონის მოდელი აღწერს იშვიათ ხდომილობათა სქემას. როგორც გვახსოვს, პუასონის განაწილება წარმოადგენს ზღვარით შემთხვევას ბინომური განაწილებისათვის, როდესაც წარმატების ალბათობა, p მიისწრაფის ნულისაკენ, ისე რომ $np=\lambda$, სადაც λ პუასონის განაწილების პარამეტრია. პუასონის მოდელის წარმოშობა კარგადაა ცნობილი დისკრეტული ალბათობიდან.

3) **ექსპონენციალური მოდელი** ($X\in\exp(\theta)$ მოდელი): X შემთხვევითი სიდიდე განაწილებულია ექსპონენციალური კანონით, რომლის განაშილების სიმკვრივეს აქვს შემდეგი სახე:

$$f(x;\theta)=\begin{cases}\vartheta\ell^{-\theta x}, & \text{როცა } x\geq 0,\\ 0, & \text{როცა } x<0.\end{cases}$$

აღნიშნული განაწილებისათვის $EX=1/\theta$. მაჩვენებლიანი განაწილების მათემატიკური ლოდინი წარმოადგენს θ პარამეტრის შებრუნებულ სიდიდეს.

4) **ნორმალური მოდელი**: რადგან განაწილებას გააჩნია 2 პარამეტრი, შეიძლება გამოვყოთ შემდეგი სამი შემთხვევა.

a) $X\in N(\mu, \sigma_0^2)$ მოდელი: უცნობია μ საშუალო, ხოლო დისპერსია ცნობილია და უდრის σ_0^2-ს. ამ შემთხვევაში განაწილების ფუნქცია

$$F(x;\mu)= \Phi_{\mu,\sigma_0^2}(x)),\ -\infty<\mu<\infty,\ \mu \text{ უცნობია},\ 0<\sigma_0^2<\infty,\ \sigma_0^2 \text{ ფიქსირებულია}.$$

b) $X\in N(\mu_0,\sigma^2)$ მოდელი: დისპერსია უცნობია, ხოლო საშუალო ცნობილია და უდრის μ_0-ს. . ამ შემთხვევაში

$$F(x;\sigma^2)= \Phi_{\mu_0,\sigma^2}(x),\ \mu_0 \text{ ფიქსირებულია},\ \sigma^2 \text{ უცნობია}.$$

c) $X\in N(\mu,\sigma^2)$ მოდელი: უცნობია როგორც საშუალო, ისევე დისპერსია. ამ შემთხვევაში

$$F(x;\mu,\sigma^2)= \Phi_{\mu,\sigma^2}(x),\ \text{უცნობია როგორც } \mu, \text{ ისე } \sigma^2.$$

აქ ყველგან $\Phi_{\mu,\sigma^2}(x)$ წარმოადგენს μ საშუალოსა და σ^2 დისპერსიის მქონე ნორმალური განაწილების ფუნქციას.

ნორმალურ მოდელს ვიყენებთ მაშინ, როდესაც ექსპერიმენტის შედეგზე გავლენას ახდენს **დამოუკიდებელი შემთხვევითი ფაქტორების** დიდი რაოდენობა და ამასთან, თითოეულის გავლენა საბოლოო შედეგზე უმნიშვნელოა. ამ შემთხვევაში ცენტრალური ზღვარითი თეორემა გვკარნახობს, რომ ნორმალური მოდელი ადეკვატურად ასახავს ექსპერიმენტს. ნორმალურ მოდელებს, როგორც წესი, ვხვდებით **გაზომვათა თეორიაში**.

<u>წერტილოვანი შეფასების ამოცანაა</u> პოპულაციის აღმწერ <u>X შემთხვევით სიდიდეზე</u> დაკვირვებების საფუძველზე <u>**გავაკეთოთ სტატისტიკური დასკვნები $\{F(x;\theta);\theta\in\Theta\}$ სტატისტიკური მოდელის უცნობი θ პარამეტრის ჭეშმარიტი მნიშვნელობის შესახებ.**</u> წერტილოვანი შეფასების აგების ამოცანა მდგომარეობს შემდეგში: უნდა მოიძებნოს ისეთი $T_n(\boldsymbol{X})=T_n(X_1,...,X_n)$ სტატისტიკა (შემფასებელი), რომლის რიცხვითი მნიშვნელობა $t_n(\boldsymbol{x})=T_n(x_1,...,x_n)$, გარკვეული აზრით, **შეიძლება ჩაითვალოს უცნობი θ პარამეტრის ჭეშმარიტი მნიშვნელობის რიცხვით მიახლოებად და გამოყენებული იქნას მის ნაცვლად. ასეთ მიახლოებას ჩვენ ვუწოდებთ წერტილოვან შეფასებას.**

ჩამოვაყალიბოთ $T_n(\boldsymbol{X})=T_n(X_1,...,X_n)$ შემფასებელთა (შემხვევით სიდიდეთა) სასურველი თვისებები.

§ 6. შემფასებელთა თვისებები

ვთქვათ მოცემულია რაიმე **სტატისტიკური მოდელი - $\{F(x;\theta);\theta\in\Theta\}$, სადაც θ უცნობი პარამეტრია.**

1. $T_n(\boldsymbol{X})=T_n(X_1,\ldots,X_n)$ შემფასებელს (სტატისტიკას) θ უცნობი პარამეტრისთვის ეწოდება **ჩაუნაცვლებელი**, თუ ნებისმიერი n-სათვის

$$ET_n(\boldsymbol{X}) = \theta. \tag{5.7}$$

თუ შეფასება არ აკმაყოფილებს (5.7) ტოლობას, მისთვის განიმარტება სიდიდე

$$b(\theta) = ET_n(\boldsymbol{X}) - \theta,$$

რომელსაც **შეფასების ჩანაცვლება** ეწოდება.

ჩაუნაცვლებლობის თვისება ნიშნავს, რომ თუ ჩავატარებთ მრავალ განმეორებით ექსპერიმენტს (ანუ მივიღებთ n მოცულობის შერჩევათა სერიებს), მაშინ აგებულ შეფასებათა საშუალო მნიშვნელობა ახლოს იქნება პარამეტრის ჭეშმარიტ მნიშვნელობასთან. ეს თვისება საშუალებას გვაძლევს დარწმუნებული ვიყოთ, რომ სისტემურ შეცდომებს ადგილი არა აქვს.

2. $T_n= T_n(X_1,\ldots,X_n)$) შემფასებელს θ უცნობი პარამეტრისთვის ეწოდება **ძალმოსილი** (ანუ, როგორც ზოგჯერ ამბობენ, ძალდებული), თუ ნებისმიერი $\varepsilon > 0$–სათვის

$$\boldsymbol{P}\{|T_n-\theta|<\varepsilon\}\to 1, \text{ როცა } n\to\infty. \tag{5.8}$$

თუ შეფასების ჩაუნაცვლებლობის თვისება დაკავშირებულია შერჩევის ფიქსირებულ მოცულობასთან, ძალმოსილება შეფასების **ასიმპტოტური (ზღვარითი)** თვისებაა. ის ახასიათებს შეფასების ქცევას, როდესაც შერჩევის n მოცულობა უსასრულოდ იზრდება.

(5.7) და (5.8) პირობას ერთი და იგივე უცნობი პარამეტრისთვის შეიძლება აკმაყოფილებდეს რამდენიმე სხვადასხვა შემფასებელი. ასეთ შემთხვევაში უპირატესობა უნდა მივანიჭოთ უმცირესი დისპერსიის მქონე შემფასებელს. აქ ჩვენ შემოვიფარგლებით ერთგანზომილებიანი θ პარამეტრის **ჩაუნაცვლებელ სტატისტიკათა** კლასის განხილვით. ბუნებრივია შეფასებათა ხარისხის შედარების მიზნით გამოვიყენოთ მათი დისპერსიები. დავაზუსტოთ, რას ვგულისხმობთ: როგორც უკვე ვიცით, პარამეტრული სტატისტიკური მოდელისათვის შესაძლებელია აიგოს პარამეტრის მრავალი ჩაუნაცვლებელი სტატისტიკა. დავუშვათ, რომ θ პარამეტრის შესაფასებლად ვიყენებთ ორ ჩაუნაცვლებელ T_1 და T_2 სტატისტიკებს. დავსვათ შეკითხვა: ამ ორი სტატისტიკიდან რომელი გვაძლევს უცნობი პარამეტრის უკეთეს შეფასებას? ამ შემთხვევაში, ბუნებრივია შეფასებათა ხარისხის შედარების მიზნით გამოვთვალოთ მათი დისპერსიები.

ამიტომ ჩავთვალოთ, რომ T_1 და T_2 ჩაუნაცვლებელ სტატისტიკებს შორის უკეთესია T_1, თუ

$$DT_1 \le DT_2 \tag{5.9}$$

ყოველი θ-სათვის, $\theta\in\Theta$ და θ-ს ერთი მნიშვნელობისათვის მაინც ადგილი აქვს მკაცრ უტოლობას.

შემფასებელთა შედარების (5.9) კრიტერიუმიდან გამომდინარე შეგვიძლია შემოვიღოთ **ოპტიმალური სტატისტიკის** შემდეგი ცნება:

3. ჩაუნაცვლებელ $T^*=T^*(\boldsymbol{X})$ სტატისტიკას θ უცნობი პარამეტრისთვის ეწოდება **ოპტიმალური**, თუ ნებისმიერი სხვა ჩაუნაცვლებელი $T=T(\boldsymbol{X})$ სტატისტიკისათვის,

$$DT^* \le DT$$

ნებისმიერი θ-სათვის Θ-დან.

წინა და მიმდინარე ლექციათა მასალაზე დაყრდნობით მოვიყვანოთ განხილული პარამეტრული მოდელების უცნობი პარამეტრების შემფასებელთა თვისებები:

- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ - შერჩევითი საშუალო არის უცნობი გენერალური საშუალოს ჩაუნაცვლებელი, ძალმოსილი შემფასებელი.
- ნორმალურად განაწილებული პოპულაციის შემთხვევაში შერჩევითი საშუალო ნორმალურადაა განაწილებული. სხვა შემთხვევაში, ცენტრალური ზღვარითი თეორემის თანახმად, შერჩევითი საშუალო ასიმპტოტურად ნორმალურადაა განაწილებული.
- S_n^2 - შერჩევითი დისპერსია არის უცნობი დისპერსიის ჩანაცვლებული, მაგრამ ძალმოსილი შემფასებელი
- $S_n'^2$ - შესწორებული შერჩევითი დისპერსია არის უცნობი დისპერსიის ჩაუნაცვლებელი, ძალმოსილი შემფასებელი

  შენიშვნა. იმ შემთხვევაში, როცა საშუალო ცნობილია ე.ი. $EX=\mu$, უცნობი დისპერსიის შემფასებლად გამოიყენება სტატისტიკა

$$\overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 .$$

  ადვილი სანახავია, რომ

$$E\,\overline{S}^2 = \sigma^2,$$

  ე.ი., $\overline{S}^2$ წარმოადგენს უცნობი დისპერსიის ჩაუნაცვლებელ შემფასებელი .

დასკვნა. დავუშვათ გენერალური ერთობლიობის პარამეტრების, μ მათემატიკური ლოდინის და σ^2 დისპერსიის ჭეშმარიტი მნიშვნელობები უცნობია და მიღებულია n მოცულობის $x_1,x_2,\ldots x_n$ შერჩევითი მნიშვნელობები აღნიშნული პოპულაციიდან. მაშინ ამ შერჩევითი მნიშვნელობების საფუძველზე გამოთვლილი $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ -შერჩევითი საშუალოსა და $s_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_n)^2$ - შესწორებული შერჩევითი დისპერსიის რიცხვითი მნიშვნელობები შეიძლება ჩაითვალოს პოპულაციის უცნობი საშუალოსა და დისპერსიის ჭეშმარიტი მნიშვნელობების მიახლოებით მნიშვნელობებად (შეფასებად).

ცხრილი . უცნობი პარამეტრის შემფასებლები

მოდელი	უცნობი პარამეტრი	უცნობი პარამეტრის შემფასებელი
$X \in Bi(n,p)$	p	$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
X, $EX=\mu$, $DX=\sigma_0^2$	μ	$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
X, $EX=\mu$, $DX=\sigma^2$	μ	$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
X, $EX=\mu_0$, $DX=\sigma^2$	σ^2	$\overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2,$
X, $EX=\mu$, $DX=\sigma^2$	σ^2	$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,$

მაგალითი. მოცემულია 9 ჩვეულებრივი აქციის წლიური ამონაგების მნიშვნელობები. ეს მონაცემები წარმოადგენს შერჩევას აქციათა ამონაგებების დიდი პოპულაციიდან, რომელიც ნორმალურადაა განაწილებული. ეს მონაცემებია:

11, 13, 8, 9, 12, 8, 11, 10, 8.

ა) რისი ტოლია პოპულაციის საშუალოს ჩაუნაცვლებელი შეფასება?

ა) რისი ტოლია პოპულაციის დისპერსიის ჩაუნაცვლებელი შეფასება?

ამოხსნა. ა) პოპულაციის საშუალოს ჩაუნაცვლებელი შეფასებაა

$$\bar{x} = \frac{11+13+8+9+12+8+11+10+8}{9} = 10,$$

ბ) პოპულაციის დისპერსიის ჩაუნაცვლებელი შეფასებაა

$$s'^2 = \frac{(11-10)^2 + (13-10)^2 + \cdots + (8-10)^2}{8} = 3{,}1. \quad \blacklozenge$$

აღნიშნული შეფასებები, გამოთვლილი კონკრეტული n მოცულობის შერჩევის საფუძველზე, როგორც წესი, განსხვავდება უცნობი პარამეტრის ჭეშმარიტი მნიშვნელობისაგან და ჩვენთვის არ არის ცნობილი, თუ რამდენად დიდია ეს განსხვავება - ამ დროს ჩვენ, რა თქმა უნდა, ვუშვებთ გარკვეულ **შეცდომას**. **წერტილოვანი შეფასების ნაკლი** ისაა, რომ მხოლოდ მისი დაკვირვებული მნიშვნელობის ცოდნა არ იძლევა ამ შეცდომის სიდიდის დადგენის საშუალებას.

ეს მოსაზრება გვაბიძგებს, რომ წერტილოვანი შეფასება – ერთი რიცხვი, შევცვალოთ ინტერვალური შეფასებით – მთელი ინტერვალით. ამასთან, ინტერვალი ისე უნდა იყოს შერჩეული, რომ ჩვენი მტკიცება, რომ პარამეტრის ჭეშმარიტი მნიშვნელობა ეკუთვნის ამ ინტერვალს, მცდარი აღმოჩნდეს მხოლოდ იშვიათ შემთხვევაში (<u>ინტერვალური შეფასების ამოცანებს განვიხილავთ VII -VIII ლექციებში)..</u>

კითხვები თვითშემოწმებისათვის

1. რა არის სტატისტიკური პოპულაცია?
2. რა არის შერჩევა ანუ ამოკრეფა? შერჩევითი მნიშვნელობა?
3. რა არის სტატისტიკური კვლევის საბოლოო მიზანი ზოგადად?
4. რა არის შერჩევის მოცულობა?
5. რა არის პოპულაციის განაწოლება?
6. რა არის პოპულაციის პარამეტრი?
7. განმარტე შერჩევის, როგორც შემთხვევითი ვექტორის წარმოდგენა
8. რა არის სტატისტიკა (შემფასებელი) ?
9. განმარტე სტატისტიკის ორმაგი შინაარსი
10. განმარტე სტატისტიკის განაწილება—შერჩევითი განაწილება
11. განმარტე შერჩევითი განაწილების დადგენის ორი ხერხი, ზუსტი და ასიმპტოტური
12. გამოიყვანე ბინომიალური შერჩევის როგორც ზუსტი ასევე ასიმპტოტური განაწილებანი
13. გააკეთე მოსაზრება-დასკვნა დამოუკიდებელი, ერთნაირად განაწილებული შერჩევისთვის შერჩევითი საშუალოს ასიმპტოტური განაწილების შესახებ
14. გამოიყვანე შერჩევითი საშუალოს ლოდინი და დისპერსია
15. გამოიყვანე შერჩევითი დისპერსიის ლოდინი
16. განმარტე შესწორებული შეჩევითი დისპერსია და მისი შემოტანის მიზეზი
17. რა შერჩევითი განაწილება აქვს ნორმალური პოპულაციის შერჩევით საშუალოს?

18. რა შერჩევითი განაწილება აქვს ნორმალური პოპულაციის შერჩევით დისპერსიას, როდესაც პოპულაციის ლოდინი ცნობილია (გააკეთე კომენტარი χ –კვადრატ განაწილებიდან)
19. რა შერჩევითი განაწილება აქვს ნორმალური პოპულაციის შერჩევით დისპერსიას, როდესაც პოპულაციის ლოდინი უცნობია (გააკეთე კომენტარი χ –კვადრატ განაწილებიდან)
20. ჩამოაყალიბე ფიშერის თეორემა
21. რა განაწილება აქვთ Z და T სტატისტიკებს?
22. ღა არის სტატისტიკური მოდელი?
23. რას ნიშნავს „მოცემულია პუასონის მოდელი“?
24. რას ნიშნავს „მოცემულია ბინომური მოდელი“?
25. რას ნიშნავს „მოცემულია ნორმალური მოდელი“?
26. რას ნიშნავს „მოცემულია ექსპონენციალური მოდელი“?
27. ჩამოაყალიბე წერტილოვანი შეფასების ამოცანა
28. ჩამოაყალიბე შემფასებლის (სტატისტიკის) ჩაუნაცვლებადობის თვისება
29. ჩამოაყალიბე შემფასებლის (სტატისტიკის) ძალმოსილობის თვისება
30. ჩამოაყალიბე შემფასებლის (სტატისტიკის) ოპტიმალურობის თვისება
31. მოიყვანეთ ჩაუნაცვლებელი და ჩანაცვლების მქონე სტატისტიკების მაგალითები.
32. მოიყვანეთ ძალმოსილი სტატისტიკების მაგალითები.

§ 5. EXCEL

აღვნიშნოთ, რომ მონაცემთა ანალიზისა და გამოყენებითი სტატისტიკის ამოცანაა დაკვირვებების საფუძველზე სტატისტიკური დასკვნების გამოტანა შესასწავლი ალბათური მოდელის შესახებ. მომავალში ჩვენ ვიგულისხმებთ, რომ რეალური ფიზიკური პროცესების რეალიზაციების დროს ჩვენ ვაკვირდებით ამა თუ იმ კანონით **განაწილებულ შემთხვევით სიდიდეთა დაკვირვებულ მნიშვნელობებს. დავუშვათ $x_1,x_2,\ldots,x_n$ რაიმე კანონით განაწილებული შემთხვევითი სიდიდის დამოუკიდებელ დაკვირვებათა მიმდევრობაა**. ალბათობის სტატისტიკური განსაზღვრების თანახმად ნებისმიერი A ხდომილობის განხორციელების ფარდობითი სიხშირე იგივე A ხდომილობის განხორციელების ალბათობის ტოლი უნდა იყოს. **მათემატიკური მეთოდები საშუალებას იძლევა "ხელოვნურად" მივიღოთ ასეთი თვისების მქონე რიცხვთა მიმდევრობა, რომლებსაც ფსევდოშემთხვევით რიცხვებს უწოდებენ. ფსევდოშემთხვევითი რიცხვების მიღების (სიმულაციის) მეთოდი სამეცნიერო ლიტერატურაში ხშირად მონტე-კარლოს მეთოდის სახელით მოიხსენიება.** ეს სახელი იმით აიხსნება, რომ ეს მეთოდი იყენებს **შემთხვევითი რიცხვების მიმდევრობებს**, რომელთა როლში შეიძლება ავიღოთ (მაგალითად) მონტე-კარლოს კაზინოში რეგულარულად გამოცხადებული თამაშების შედეგები. **პრაქტიკაში (დღეს) შემთხვევითი რიცხვების გენერირება ხორციელდება კომპიუტერის საშუალებით (ქვემოთ სიტყვებს გენერირება, მოდელირება, სიმულირება სინონიმებად ჩავთვლით).**

Excel-ის ქვეპროგრამა **Tools/Data Analysis/Random Number Generation** საშუალებას გვაძლევს მივიღოთ სხვადასხვა კანონით განაწილებული ფსევდოშემთხვევითი რიცხვების

რეალიზაციები. ქვეპროგრამის გამოყენებისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა მოცემულია ცხრილში.

აღნიშვნა	შინაარსი
Number of Variables	დაკვირვებულ სერიათა რაოდენობა
Number of Random Numbers	მონაცემთა რაოდენობა სერიაში
Distribution	განაწილების კანონი
Discrete	დისკრეტული
Value and Probability Input Range	მითითებულ უნდა იქნას დავთრის ფურცლის ის არე (ვერტიკალური ორიენტაციით), სადაც წინასწარ ჩაწერილია დისკრეტული განაწილების კანონი, რომლის საფუძველზეც უნდა გათამაშდეს პსევდოშემთხვევითი სიდიდეები.
Uniform	თანაბარი
Parameters	შესაბამის უჯრებში მითითებულ უნდა იყოს თანაბარი განაწილების (a,b) შუალედი
Normal	ნორმალური
Mean	საშუალო
Stadard Deviation	სტანდარტული გადახრა
Bernouli	ბერნული
Binomial	ბინომური
***p*-value**	p პარამეტრის რიცხვითი მნიშვნელობა
Number of Trials	n პარამეტრების რიცხვითი მნიშვნელობები ბერნულის სქემაში
Poisson	პუასონი
Lambda	λ პარამეტრი

შენიშვნა.

ვთქვათ, სერიათა რაოდენობაა k, ხოლო მონაცემთა რაოდენობა სერიაში n-ის ტოლია, მაშინ აღნიშნული პროგრამა მოგვცემს kn რაოდენობის ფსევდოშემთხვევითი რიცხვების სიმრავლეს. თითოეულ სვეტში ჩაწერილი იქნება n მოცულობის შერჩევითი მნიშვნელობები, ხოლო k მეზობელ სვეტში გათამაშებული სერიების შედეგები.

Output options ველის გადამრთველები (Output Range, New Worksheet by, New Workbook) განსაზღვრავენ გენერირებული მონაცემების გამოტანის ადგილმდებარეობას.

ლექცია VII, VIII შეფასების ინტერვალური მეთოდი

§ 1. ნდობის ინტერვალის აგება ნორმალური პოპულაციის საშუალოსათვის

1. ***X∈N(μ,σ₀) მოდელი***. ვთქვათ,

$$\boldsymbol{X}=(X_1,\ldots,X_n) \tag{8.1}$$

არის n მოცულობის შერჩევაა. როგორც ვიცით, გენერალური საშუალოს "კარგ" შეფასებას წარმოადგენს შერჩევითი საშუალო

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N(\mu,\sigma/\sqrt{n})\,. \tag{8.2}$$

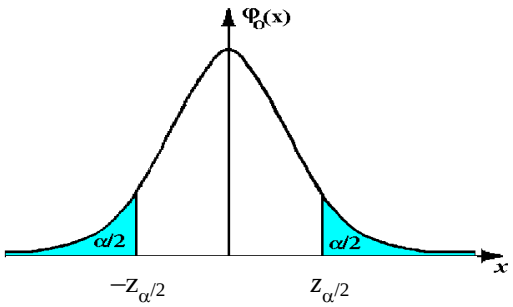
$\overline{X}$ -ის სტანდარტიზაცია გვაძლევს

$$Z_n=\frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1) \tag{8.3}$$

(8.3)-ის თანახმად ნებისმიერი **წინასწარ განსაზღვრული (1-α)-თვის ადვილი მოსაძებნია ნულის მიმართ სიმეტრიული ისეთი ($-z_{\alpha/2}$, $z_{\alpha/2}$) ინტერვალი, რომელშიც Z_n-ის მოხვედრის ალბათობა (1-α)-ს ტოლია**, ე.ი.

$$P\{-z_{\alpha/2}\le Z_n\le z_{\alpha/2}\}=P\left\{-z_{\alpha/2}\le\frac{\overline{X}_n-\mu}{\sigma_0/\sqrt{n}}\le z_{\alpha/2}\right\}=1-\alpha. \tag{8.4}$$

როგორც წესი იღებენ α-ს შემდეგ რიცხვით მნიშვნელობებს α=0,1; 0,05; 0,01. (იხ.ნახ.8.1,სადაც არაგამუქებული ფიგურის ფართობი (1-α)-ს, ანუ მთელი ფართობის (1-α)·100%-ს ტოლია).



ნახ. 8.1.

$z_{\alpha/2}$-ს სტანდარტული ნორმალური განაწილების ზედა α/2-კრიტიკულ წერტილს, ხოლო $-z_{\alpha/2}$-ს ქვედა α/2 კრიტიკულ წერტილს, უწოდებენ. $z_{\alpha/2}$-ის რიცხვითი მნიშვნელობა მარტივად განისაზღვრება განაწილების ცხრილის საშუალებით.

დავუშვათ, რომ შერჩევის მოცულობა n და გენერალური დისპერსია ცნობილია (, $\sigma=\sigma_{0)}$. შერჩევითი მნიშვნელობების საფუძველზე ადვილი გამოსათვლელია $\overline{X}$ - შერჩევითი საშუალოს რიცხვითი მნიშვნელობა. ამრიგად, ერთადერთი უცნობი, რომელსაც შეიცავს (8.4) გამოსახულება არის განერალური ერთობლიობის უცნობი საშუალო-μ. უკანასკნელი გამოსახულების ფიგურულ ფრჩხილებში ჩაწერილი უტოლობა გადავწეროთ μ-ის მიმართ:

$$P\left\{\overline{X}-\frac{z_{\alpha/2}\cdot\sigma_0}{\sqrt{n}}\le\mu\le\overline{X}+\frac{z_{\alpha/2}\cdot\sigma_0}{\sqrt{n}}\right\}=1-\alpha.$$

აქედან

$$\left[\overline{X}-\frac{z_{\alpha/2}\cdot\sigma_0}{\sqrt{n}},\overline{X}+\frac{z_{\alpha/2}\cdot\sigma_0}{\sqrt{n}}\right] \tag{8.5}$$

წარმოადგენს (1-α) ნდობის ალბათობით აგებულ ნდობის ინტერვალს. *n*□□ მოცულობის შერჩევას ვუწოდოთ სერია. განმეორებით სერიებში ამ წესით აგებული ნდობის ინტერვალი შეიცავს უცნობი პარამეტრის ჭეშმარიტ μ_0 მნიშვნელობას ცდათა სერიათა (1-α)100 შემთხვევაში. ამ ინტერვალს (1-α)·100%-იან ნდობის ინტერვალსაც უწოდებენ.

(1-α)-ნდობის ინტერვალი $X\in N(\mu,\sigma_0)$ მოდელის საშუალოსათვის გამოითვლება შემდეგი ფორმულით

$$\left(\overline{x}-\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2},\overline{x}+\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}\right) \tag{8.6}$$

$\overline{x}$ შერჩევითი საშუალოს რიცხვითი მნიშვნელობაა,

$z_{\alpha/2}$- სტანდარტული ნორმალური განაწილების ზედა $\alpha/2$ ზომის კრიტიკული წერტილია.

(1-α)-ნდობის ინტერვალის სწორი ინტერპრეტაცია ეფუძნება ალბათობის შემდეგ სიხშირულ ინტერპრეტაციას:

- თუ რაიმე ექსპერიმენტთან დაკავშირებული A ხდომილობის ალბათობა, $P(A)=(1-\alpha)$, მაშინ ამ ექსპერიმენტის მრავალჯერ განმეორებისას A ხდომილობა განხორციელდება დაახლოებით $(1-\alpha)\cdot100\%$ შემთხვევაში (ანუ A ხდომილობის ფარდობითი სიხშირე მიახლოებით (1–α)-ს ტოლი იქნება).
- თუ (1–α) ერთთან ახლომყოფი რაიმე რიცხვია, მაშინ პრაქტიკულად შეგვიძლია ჩავთვალოთ, რომ აღნიშნული ხდომილობა ერთეულოვან ცდაში თითქმის ყოველთვის განხორციელდება.

პრაქტიკაში ჩვენ ხელთ გვაქვს **მხოლოდ ერთი შერჩევა, გამოითვლება მხოლოდ ერთი შერჩევითი საშუალო და აიგება მხოლოდ ერთი (1-α)-ნდობის ინტერვალი. ყოველ ცალკეულ შემთხვევაში ჩვენ არ ვიცით შეიცავს თუ არა ინტერვალური შეფასება (ნდობის ინტერვალი) პოპულაციის საშუალოს.** უფრო მეტიც, არ არსებობს გზა, რომელიც დაგვარწმუნებდა ჩვენი დასკვნის სისწორეში. (1-α)-ნდობის ინტერვალის ზემომოყვანილი ინტერპრეტაციაის თანახმადაც შანსი იმისა, რომ კონკრეტული შერჩევა არ ეკუთვნის იმ შერჩევათა რიცხვს, რომლებისთვისაც აგებული ნდობის ინტერვალი არ ფარავს პოპულაციის საშუალოს, მეტად მცირეა: 100-დან ~ α·100. მაგალითად, 100 ვარიანტიდან – 10, თუ (1-α)=0.9, 100-დან 1, თუ (1-α)=0.99). აქედან უკვე აშკარა ხდება რამდენად მნიშვნელოვანია ნდობის (1-α)-დონის სწორად შერჩევის პრობლემა.

2. $X\in N(\mu,\sigma)$ მოდელი. განსახილველ შემთხვევაში ორივე პარამეტრი μ და σ^2 უცნობია. როგორც ვიცით, μ და σ^2 პარამეტრების „კარგი" შეფასებებია შერჩევითი საშუალო და შესწორებული შერჩევითი დისპერსია. ამასთან

$$\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\sim N(0,1), \qquad \frac{(n-1)S'^2}{\sigma^2}\sim\chi^2(n-1).$$

ფიშერის თეორემის ძალით აღნიშნული შემთხვევითი სიდიდეები **დამოუკიდებელია**. თუ ახლა გავიხსენებთ **სტიუდენტის *t* განაწილების** განმარტებას, მივიღებთ, რომ

$$T_n=\frac{\dfrac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S'^2}{\sigma^2}\Big/(n-1)}}=\frac{\overline{X}_n-\mu}{S'/\sqrt{n}}\sim t(n\text{-}1).$$

ნებისმიერი წინასწარ განსაზღვრული (1-α)-თვის ადვილი მოსაძებნია **ნულის მიმართ სიმეტრიული ისეთი** $\{t_{n-1,1-\alpha/2},\ t_{n-1,\alpha/2}\}$**ინტერვალი**, რომელშიც T_n სტატისტიკის მოხვედრის ალბათობა (1-α)-ს ტოლია, ე.ი. ადგილი აქვს შემდეგ თანაფარდობას

$$P\left\{t_{n-1,1-\alpha/2}\le\frac{\overline{X}_n-\mu}{S'/\sqrt{n}}\le t_{n-1,\alpha/2}\right\}=1\text{-}\alpha$$

სადაც $t_{n-1,1-\alpha/2}$ და $t_{n-1,\alpha/2}$ სტიუდენტის $t(n\text{-}1)$ განაწილების **ქვედა და, შესაბამისად, ზედა α/2 ზომის კრიტიკული წერტილებია**. ამ ფაქტზე დაყრდნობით (1–α)-ნდობის ინტერვალი უცნობი საშუალოსათვის აიგება ზუსტად იგივე გზით, როგორც ცნობილი დისპერსიის შემთხვევაში. კერძოდ

$$\left(\bar{x}-\frac{s'}{\sqrt{n}}t_{n-1,\alpha/2};\bar{x}+\frac{s'}{\sqrt{n}}t_{n-1,\alpha/2}\right),$$

წარმოადგენს **(1-α) ნდობის ალბათობით აგებულ ნდობის ინტერვალს უცნობი საშუალოსათვის.**

(1–α)-ნდობის ინტერვალი $N(\mu,\sigma)$ მოდელის საშუალოსათვის გამოითვლება შემდეგი ფორმულით

$$\left(\bar{x}-\frac{s'}{\sqrt{n}}t_{n-1,\alpha/2};\bar{x}+\frac{s'}{\sqrt{n}}t_{n-1,\alpha/2}\right), \tag{8.7}$$

სადაც $\bar{x}$ და s' შერჩევითი საშუალოსა და, შესაბამისად, სტანდარტული გადახრის რიცხვითი მნიშვნელობებია, ხოლო $t_{n-1,\alpha/2}$ სტიუდენტის $t(n\text{-}1)$ განაწილების ზედა α/2 ზომის კრიტიკული წერეილია.

მაგალითი განვიხილოთ მონაცემები, რომლებიც წარმოადგენს რომელიმე ფირმის მოსამსახურეთა 25 სატელეფონო საუბრის ხანგრძლივობას (გამოხატულს წუთებში).

11.8	3.6	16.6	13.5	4.8
8.9	9.1	8.6	4.5	11.8
10.2	8.4	10.6	6.8	9.6
15.3	12.3	8.5	15.9	18.7
6.2	11.2	10.4	7.2	5.5

ამ მონაცემების საფუძველზე ავაგოთ 0,95 ნდობის ინტერვალი ფირმის მოსამსახურეთა სატელეფონო საუბრების საშუალო ხანგრძლივობისათვის.

ამოხსნა. უპირველეს ყოვლისა განვსაზღვროთ, ამოცანის პირობის შესაბამისი პარამეტრული მოდელი. ვინაიდან ფირმის მოსამსახურეთა სატელეფონო საუბრის ხანგრძლივობა დამოკიდებულია მრავალ ფაქტორზე, სავსებით ბუნებრივია, ცენტრალური ზღვარითი თეორემის თანახმად ჩავთვალოთ, რომ მონაცემები მიღებულია ნორმალურად განაწილებული პოპულაციიდან. პოპულაციის ორივე პარამეტრი, საშუალო და დისპერსია უცნობია. გამოვთვალოთ შერჩევითი საშუალო და სტანდარტული გადახრა: $\bar{x}=10$, $s'=3.92$, $t_{0.025,24}=2.064$.

ვისარგებლოთ (8.7) ფორმულით. 95%-იანი ნდობის ინტერვალი საშუალოსათვის იქნება

(8.38 ; 11.62). ♦

§2. ნდობის დონის სიზუსტე და შერჩევის მოცულობის მოძებნა.

როგორც ვნახეთ, (1–α)-ნდობის ინტერვალის სიგრძეა $2\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}$. როგორც ვიცით, რაც უფრო დიდია (1–α), მით უფრო დიდია $z_{\alpha/2}$ და მაშასადამე, განიერია ნდობის ინტერვალი და პირიქით. თუ შერჩევის n მოცულობა ფიქსირებულია, **ნდობის ინტერვალის სიგრძის** (ანუ შეფასების სიზუსტის) შემცირება შესაძლებელია მხოლოდ ნდობის დონის, (1–α)-ს შემცირების ხარჯზე, ხოლო ფიქსირებული (1–α) ნდობის დონის დროს ინტერვალის სიგანე **მით უფრო მცირეა, რა უფრო დიდია შერჩევის მოცულობა,** n (ე.ი. იზრდება ინტერვალური შეფასების სიზუსტე).

ბუნებრივია ისმის ამოცანა: **<u>ნდობის ფიქსირებული (1–α)-დონისათვის მოიძებნოს შერჩევის ის მინიმალური n^* მოცულობა, რომელიც უზრუნველყოფს შეფასების წინასწარ ფიქსირებულ სიზუსტეს (L სიგრძეს).</u>**

რადგან, (1–α)-ნდობის ინტერვალის სიგრძეა $2\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}$, ამიტომ სასურველი მოცულობა, n^* მოსაძებნია შემდეგი განტოლებიდან:

$$\frac{2\sigma_0}{\sqrt{n}}z_{\alpha/2}=L,$$

საიდანაც მივიღებთ

$$n^*=\left(\frac{2\sigma_0}{L}z_{\alpha/2}\right)^2.$$

ანუ n^* არის ის უმცირესი მთელი რიცხვი, რომელიც მეტია $\left(2\frac{\sigma_0}{L}z_{\alpha/2}\right)^2$-ზე, ე.ი.

$$n^*=\left[\left(\frac{2\sigma_0}{L}z_{\alpha/2}\right)^2\right]+1. \qquad (8.8)$$

რაც უფრო მცირეა ინტერვალის სიგრძე, მით უფრო დიდია n^*, რაც სავსებით ეთანხმება ჩვენს ინტუიციას, ამავე დროს n^* წარმოადგენს σ^2 დისპერსიისა და 1–α ნდობის დონის ზრდად ფუნქციას.

მაგალითი . ვთქვათ, $\sigma = 3.5$, $(1–\alpha)=0.95$, $L=1$, მაშინ $z_{0,025}=1,96$ და

$$n^* = [7\cdot 1,96]^2+1 = 189. \quad ♦$$

§3. ნდობის ინტერვალი პოპულაციის საშუალოსათვის შერჩევის დიდი მოცულობის შემთხვევაში. ახლა განვიხილოთ შემთხვევა, როდესაც პოპულაცია არაა ნორმალურად განაწილებული. ჯერ ვიგულისხმოთ, რომ პოპულაციის საშუალო უცნობია, ხოლო დისპერსია ცნობილია. ცენტრალური ზღვარითი თეორემის ძალით, როცა შერჩევის n მოცულობა საკმარისად დიდია

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma_0 / \sqrt{n}}$$

სტატისტიკას მიახლოებით სტანდარტული ნორმალური განაწილება აქვს. ამ ფაქტის გამოყენებით, წინა პუნქტებში ჩატარებული სტანდარტული მსჯელობით შეგვიძლია დავასკვნათ, რომ

$$(\bar{x}_n - \sigma_0 \cdot z_{\alpha/2} / \sqrt{n} \; ; \; \bar{x}_n + \sigma_0 \cdot z_{\alpha/2} / \sqrt{n}) \qquad (8.9)$$

ინტერვალს აქვს მიახლოებით $1-\alpha$ ნდობის დონე. ასეთ ინტერვალს, ხშირად, ასიმპტოტურ ნდობის ინტერვალსაც უწოდებენ.

იმ შემთხვევაში, როდესაც პოპულაციის დისპერსიაც უცნობია, განიხილება

$$\frac{\overline{X}_n - \mu}{S_n^{'} / \sqrt{n}}$$

სტატისტიკა, რომელიც n-ის საკმარისად დიდი მნიშვნელობისათვის (იმ დამატებითი ფაქტის გათვალისწინებით, რომ $S_n^{'} \xrightarrow{P} \sigma^2$, როცა $n \to \infty$), მიახლოებით ნორმალურადაა განაწილებული, ნულოვანი საშუალოთი და ერთეულოვანი დისპერსიით. აქედან, სტანდარტული მსჯელობის გამოყენებით, შეიძლება გაკეთდეს დასკვნა: ინტერვალი

$$(\bar{x}_n - s' \cdot z_{\alpha/2} / \sqrt{n} \; ; \; \bar{x}_n + s' \cdot z_{\alpha/2} / \sqrt{n}) \qquad (8.10)$$

წარმოადგენს $1-\alpha$ ნდობის დონის ასიმპტოტურ ინტერვალს (მისი ნდობის დონე მიახლოებით $(1-\alpha)$-ს ტოლია).

როდესაც $n \geq 30$, (8.10) ინტერვალი შეიძლება გამოყენებული იქნეს უცნობი საშუალოს ინტერვალური შეფასებისათვის $\approx (1-\alpha)\cdot 100\%$-იანი ნდობის დონით, ზოგად შემთხვევაშიც.

§ 4. ნდობის ინტერვალი ნორმალური პოპულაციის დისპერსიისათვის

<u>$X \in N(\mu,\sigma)$ მოდელი</u>. განსახილველ შემთხვევაში პოპულაციის უცნობი σ^2 დისპერსიის

$$S_n'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

შეფასებას (სტატისტიკას) აქვს χ^2-განაწილება თავისუფლების ხარისხით $n-1$, ანუ

$$\frac{(n-1) \cdot S_n'^2}{\sigma^2} \sim \chi^2(n-1),$$

რის საფუძველზეც წინასწარ განსაზღვრული $(1-\alpha)$-თვის

$$P\left\{ \chi_{n-1,1-\alpha/2} \leq \frac{(n-1) \cdot S_n'^2}{\sigma^2} \leq \chi_{n-1,\alpha/2} \right\} = 1-\alpha \,, \qquad (8.11)$$

სადაც $\chi^2_{n,p}$ **არის** $\chi^2(n)$ **განაწილების ზედა** p**-კრიტიკული წერტილი**. (8.11)-ის საფუძველზე, ანალოგიურად წინა შემთხვევისა, **აიგება ნდობის ინტერვალი პოპულაციის დისპერსიისათვის.**

(1–α)-ნდობის ინტერვალი $N(\mu,\sigma)$ **მოდელის დისპერსიისათვის გამოითვლება შემდეგი ფორმულით**

$$\left(\frac{(n-1)\cdot s'^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)\cdot s'^2}{\chi^2_{n-1,1-\alpha/2}}\right), \qquad (8.12)$$

სადაც $\chi^2_{n-1,\,p}$ არის $\chi^2(n-1)$ განაწილების ზედა p-კრიტიკული წერტილი.

მაგალითი ნორმალურად განაწილებული გენერალური ერთობლიობიდან მიღებულია n=25-ის ტოლი ამონარჩევი და მის საფუძველზე გამოთვლილია შესწორებული შერჩევითი დისპერსია S'^2=3,5. (1-α)=0.9 ნდობის ალბათობით ააგეთ ნდობის ინტერვალი გენერალური დისპერსიისათვის.

ამოხსნა: ცხრილის საშუალებით ვიპოვით $\chi^2_{\alpha/2;n}=\chi^2_{0.05;24}$=36,41 და $\chi^2_{1-\alpha/2;n}=\chi^2_{0.95;24}$=13,84. 0.9 ნდობის ალბათობით ააგეთ ნდობის ინტერვალია

$$\left[\frac{25\cdot 3{,}5}{36{,}41}; \frac{25\cdot 3{,}5}{13{,}84}\right] = [2{,}40; 6{,}32]. \qquad \blacklozenge$$

§ 5. ბერნულის სქემაში წარმატების უცნობი ალბათობის (პოპულაციის პროპორციის) ინტერვალური შეფასება

ბერნულის სქემაში უცნობი p ალბათობის **საუკეთესო წერტილოვანი შეფასებაა** ფარდობითი სიხშირე,

$$\hat{P}_n = \frac{\nu_n}{n}, \qquad (8.13)$$

სადაც, $\nu_n=\sum_{i=1}^{n} X_i$ წარმატებათა რაოდენობაა n დამოუკიდებელ ცდაში, ამასთან

$$E\hat{P}_n = p, \qquad D\hat{P}_n = \frac{p(1-p)}{n}.$$

უცნობი p **ალბათობისათვის ნდობის ინტერვალის ასაგებად იყენებენ სტანდარტიზაციის შედეგად მიღებულ სტატისტიკას**

$$\frac{\hat{P}_n - p}{\sqrt{\frac{p(1-p)}{n}}},$$

რომელიც, თუ შერჩევის მოცულობა n **საკმარისად დიდია, მიახლოებით ნორმალურადაა განაწილებული, ნულოვანი საშუალოთი და ერთეულოვანი დისპერსიით**, მაგრამ ამ სტატისტიკის გამოსახულებაში მნიშვნელი შეიცავს **შესაფასებელ** p **პარამეტრს**, რაც **აძნელებს სტანდარტული გზით ნდობის ინტერვალის აგებას**.

შეიძლება გამოვიყენოთ გამარტივებული მიდგომა, რომლის თანახმადაც მნიშვნელში უცნობი p უნდა შეიცვალოს მისი $\hat{P}_n$ შეფასებით და აქედან მოყოლებული ნდობის ინტერვალი აიგოს სტანდარტული მსჯელობით.

შერჩევის დიდი მოცულობის შემთხვევაში (1–α)-ნდობის ინტერვალი *Bi*(1, *p*) მოდელის უცნობის ალბათობისათვის გამოითვლება შემდეგი ფორმულით

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{P}_n \hat{Q}_n}{n}}; \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{P}_n \hat{Q}_n}{n}} \right) \tag{8.14}$$

სადაც, $\vec{p}_n = \frac{\nu_n}{n}$, ν_n წარმატებათა რაოდენობაა n ცდაში.

ეს ინტერვალი შეიძლება გამოყენებული იქნეს, როცა $n\hat{P}_n \geq 5$, $n\hat{Q}_n \geq 5$. $z_{\alpha/2}$ სტანდარტული ნორმალური განაწილების ზედა α/2-კრიტიკული წერტილია.

მაგალითი საწარმო უშვებს გარკვეული ტიპის პროდუქციას. მენეჯერს აინტერესებს შეაფასოს დეფექტური პროდუქციის პროპორცია. ამ მიზნით შერჩეული იყო 400 ერთეული, რომელთა შორის დეფექტური აღმოჩნდა 60. 99%-იანი ნდობის დონით ააგეთ ინტერვალური შეფასება საწარმოს მიერ გამოშვებული დეფექტური პროდუქციის უცნობი პროპორციისათვის.

ამოხსნა. $\hat{P}$=60/400=0.15, $\hat{Q}$=1–$\hat{P}$=0.85, α=0.01, $z_{\alpha/2}$=2.58. რადგან n=400, $n\hat{P}$=60, $n\hat{Q}$=340, ამიტომ შეგვიძლია ვისარგებლოთ (8.14) ფორმულით, რომლის თანახმად საძიებელი ინტერვალია

$$0.15 \pm 2.58\sqrt{\frac{0.15 \cdot 0.85}{400}} = 0.15 \pm 0.046.$$

ე.ი., უცნობი პროპორციისათვის 99%-იანი ნდობის ინტერვალია (0.104;0.196), შეფასების სიზუსტეა ε=0.046. ♦

ცხრილი 8.1. (1–α)-ნდობის ალბათობით აგებული ნდობის ინტერვალი

მოდელი	უცნობი პარამეტრი	ნდობის ინტერვალის ასაგებად გამოყენებული სტატისტიკა	(1–α)-ნდობის ალბათობით აგებული ნდობის ინტერვალი
$X \in Bi(1,p)$	p	$\frac{\hat{P}_n - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \overset{AS}{\sim} N(0,1)$	$\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}}; \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}} \right)$
$X \in N(\mu,\sigma_0)$	μ	$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$	$\left(\bar{x} - \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2}, \bar{x} + \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2} \right)$
$X \in N(\mu,\sigma)$	μ	$T = \frac{\bar{X} - \mu}{S'} \cdot \sqrt{n} \sim t(n-1)$	$\left(\bar{x} - \frac{s'}{\sqrt{n}} t_{n-1,\alpha/2}; \bar{x} + \frac{s'}{\sqrt{n}} t_{n-1,\alpha/2} \right)$
$X \in N(\mu_0,\sigma)$	σ^2	$\bar{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$, $\frac{n\bar{S}^2}{\sigma^2} \sim \chi^2(n)$	$\left(\frac{n\bar{s}^2}{\chi^2_{n,\alpha/2}}, \frac{n\bar{s}^2}{\chi^2_{n,1-\alpha/2}} \right)$
$X \in N(\mu,\sigma)$	σ^2	$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	$\left(\frac{(n-1)s'^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s'^2}{\chi^2_{n-1,1-\alpha/2}} \right)$

§ 6. EXCEL

განვიხილოთ **Excel-**ის სტატისტიკური ფუნქცია.

CONFIDENCE,

რომლის ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა მოცემულია ცხრილში

აღნიშვნა	შინაარსი
Alpha	ალფა
Standard dev	სტანდარტული გადახრა
Size	მოცულობა (შერჩევის)

CONFIDENCE ფუნქციის ჩაწერის სინტაქსი ასეთია:

CONFIDENCE (Alpha, Standard dev, Size),

CONFIDENCE - გამოიანგარიშება ნორმალური პოპულაციის საშუალოსათვის (1-α)ნდობის ინტერვალის ნახევარსიგრძე ფორმულით $\frac{z_{\alpha/2}\cdot\sigma_0}{\sqrt{n}}$.

მაგალითი.

რისი ტოლი იქნება ნორმალური პოპულაციის საშუალოსათვის 0.95 ნდობის ალბათობით აგებული ნდობის ინტერვალის ნახევარსიგრძე, თუ პოპულაციის საშუალო კვადრატული გადახრა 8.3-ის, ხოლო შერჩევის მოცულობა 38-ის ტოლია?

ამოხსნა.

მოცემულია: ნდობის ალობათობა1-α=0.95, შერჩევის მოცულობა, n=38, პოპულაციის საშუალო კვადრატული გადახრა, σ=3.7.

ამ შემთხვევაში **CONFIDENCE** ფუნქცია ასე ჩაიწერება:

CONFIDENCE (Alpha, Standard dev, Size)=CONFIDENCE (0.05,8.3,38).

ამოხსნის შედეგად მივიღებთ, რომ ნდობის ინტერვალის ნახევარსიგრძე ტოლი იქნება 2,64-ის.

კითხვები თვითშემოწმებისათვის

- ღა არის ნდობის ინტერვალი?
- ააგეთ ნდობის ინტერვალი უცნობი გენერალური საშუალოსთვის $X\in N(\mu,\sigma_0)$ მოდელში.
- ააგეთ ნდობის ინტერვალი უცნობი გენერალური საშუალოსთვის $X\in N(\mu,\sigma)$ მოდელში.
- ააგე ოპტიმალური შერჩევის მოცულობა $X\in N(\mu,\sigma_0)$ მოდელში.
- ააგეთ ნდობის ინტერვალი უცნობი გენერალური საშუალოსთვის შერჩევის დიდი მოცულობის შემთხვევაში.
- ააგეთ ნდობის ინტერვალი უცნობი გენერალური დისპერსიითვის $X\in N(\mu,\sigma)$ მოდელში.

- ააგეთ ნდობის ინტერვალი უცნობი პროპორციისთვის ბერნულის სქემაში.
- აღწერეთ, როგორ იმოქმედებს უცნობი საშუალოსათვის აგებულ ფიქსირებული $1-\alpha$ დონის ნდობის ინტერვალის სიგრძეზე: a) შერჩევის მოცულობის ზრდა, b) სტანდარტული გადახრის ზრდა.
- აღწერეთ როგორ არის დამოკიდებული ნდობის ინტერვალის სიგრძე ნდობის დონეზე. კერძოდ, რა მოუვა ნდობის ინტერვალის სიგრძეს, თუ: a) ნდობის დონე გაიზრდება 95%-დან 99%-მდე, b) ნდობის დონე შემცირდება 95%-დან 90%-მდე.

სავარჯიშოები

8.1. მოცემულია 10 პასუხისმგებელი მდივნის წლიური ხელფასი ($ 1 000-ში)

35.0 67.5 51.5 53.0 38.0 42.0 29.5 31.5 46.0 37.5

ჩათვალეთ, რომ პასუხისმგებელ მდივანთა ხელფასების პოპულაცია ნორმალურადაა განაწილებული და მოძებნეთ: ა) საშუალო ხელფასის ჩაუნაცვლებელი, ძალმოსილი და ეფექტური შეფასება, ბ) დისპერსიის ჩაუნაცვლებელი და ძალმოსილი შეფასება, გ) $ 40 000-ზე მეტი ხელფასის მქონე პირთა პროპორცია პოპულაციაში.

8.2. სატელეფონო კომუტატორში ყოველ წუთში არასწორ შეერთებათა რაოდენობა პუასონის კანონით განაწილებული შემთხვევითი სიდიდეა უცნობი θ ინტენსივობით. აფიქსირებდნენ არასწორ შეერთებათა რაოდენობას ყოველ წუთში. დაკვირვებების შედეგებია:

3 4 1 2 0 2 4 3 0 5 1 1 2 1 1

რისი ტოლია θ პარამეტრის ჩაუნაცვლებელი და ეფექტური შეფასება?

8.3. აღწერეთ, როგორ იმოქმედებს უცნობი საშუალოსათვის აგებულ ფიქსირებული $1-\alpha$ დონის ნდობის ინტერვალის სიგრძეზე: ა) შერჩევის მოცულობის ზრდა, ბ) სტანდარტული გადახრის ზრდა.

8.4. აღწერეთ როგორ არის დამოკიდებული ნდობის ინტერვალის სიგრძე ნდობის დონეზე. კერძოდ, რა მოუვა ნდობის ინტერვალის სიგრძეს, თუ: ა) ნდობის დონე გაიზრდება 95%-დან 99%-მდე, ბ) ნდობის დონე შემცირდება 95%-დან 90%-მდე.

8.5. ნორმალური პოპულაციიდან, რომლის დისპერსია 100-ის ტოლია, აღებულია შერჩევა, რომლის მონაცემებია

12 8 22 15 30 6 39 48

ააგეთ 90%-იანი ნდობის ინტერვალი უცნობი საშუალოსათვის.

8.6. n=75 მოცულობის შერჩევის მონაცემებმა მოგვცა: $\overline{X}_n$ =27.3, S'=7.8. ააგეთ 99%-იანი ნდობის ინტერვალები უცნობი საშუალოსათვის.

8.7. ერთ-ერთი დიდი უნივერსიტეტის სტუდენტებისაგან აღებული იქნა n=70 მოცულობის შერჩევა. მათი გამოკითხვის საფუძველზე დაადგინეს, რომ მათ მიერ ერთი კვირის მანძილზე საშინაო დავალებებზე დახარჯული საშუალო დრო შეადგენს 14.3 საათს. ჩათვალეთ, რომ σ=4 და ა) რომელ პარამეტრულ მოდელს გამოიყენებთ ამ პირობებში? ბ) ააგეთ 99%-იანი ნდობის ინტერვალი უცნობი საშუალოსათვის.

8.8. სადაზღვევო კომპანია ატარებს ფასების (პრემიების) კატალოგის რევიზიას. რიგით აქტუარს სურს შეაფასოს იმ მოთხოვნათა საშუალო სიდიდე, რომლებიც გამოწვეულია სახანძრო შემთხვევებით გარკვეული ტიპის საცხოვრებელ სახლში. შერჩევის როლში აღებული იქნა ბოლო წლის ანგარიშსწორების მონაცემები. სულ წარმოდგენილი იყო 19 სახანძრო შემთხვევა, რომელთა მოთხოვნების საშუალო სიდიდე აღმოჩნდა $ 73 249, შესწორებული სტანდარტული

გადახრა კი – $ 37 246. ა) რომელ პარამეტრულ მოდელს გამოიყენებთ ამ პირობებში? ბ) ააგეთ 99%-იანი ნდობის ინტერვალი უცნობი საშუალოსათვის.

8.9. შემთხვევითი სიდიდე ემორჩილება ნორმალური განაწილების კანონს. ააგეთ 95%-იანი ნდობის ინტერვალი უცნობი დისპერსიისათვის, თუ შერჩევის მოცულობაა n=22, ხოლო შესწორებული შერჩევითი დისპერსია 17,5.

8.10. ბანკის მენეჯერს აინტერესებს შეაფასოს დეპოზიტების საშუალო მოცულობა კლიენტების გარკვეული კლასისათვის. 400 ასეთი დეპოზიტისაგან აღებული შემთხვევითი შერჩევის საშუალომ შეადგინა $\overline{X}_n$ =6.2 (1000$), ხოლო შესწორებული სტანდარტული გადახრაა – S '=2.2 ($1000). ა) რომელ პარამეტრულ მოდელს გამოიყენებთ ამ პირობებში? ბ) ააგეთ 95%-იანი ნდობის ინტერვალი დეპოზიტების საშუალო მოცულობისათვის.

8.11. ფარმაცევტული წარმოების პროდუქციის, ხარისხის კონტროლის მენეჯერმა, წინა ისტორიიდან გამომდინარე იცის, რომ ყოველი ცალკეული კაფსულა შეიცავს მინარევთა გარკვეულ დოზას, ამასთან დოზები მერყეობს. კონტროლის მექანიზმი არეგულირებს მინარევთა დოზის საშუალო დონეს. პერიოდულად კონტროლის მექანიზმი მწყობრიდან გამოდის და შესაძლოა, მინარევთა დოზის საშუალო დონე აღარ აღმოჩნდეს დასაშვებ ფარგლებში. კონტროლის მექანიზმის შემოწმების მიზნით მზა პროდუქციიდან აიღება სხვადასხვა მოცულობის შემთხვევითი შერჩევა და აიგება ნდობის ინტერვალი პოპულაციის საშუალოსათვის. ითვლება, რომ მინარევთა დონის ვარიაბელობა დაახლოებით მუდმივია და მოიცემა სტანდარტული გადახრით, σ=2.5 (დოზის ერთეული).

შერჩევების მიხედვით მიღებული შედეგები არითმეტიკული საშუალოსათვის შემდეგია: 1. n= 25, $\overline{X}_n$ =153.4; 2. n=100, $\overline{X}_n$ =125.6; 3. n=200, $\overline{X}_n$ =148.9; 4. n=500, $\overline{X}_n$ =206.4.

ა) რომელ პარამეტრულ მოდელს გამოიყენებთ ამ პირობებში?

ბ) ააგეთ 95%-იანი ნდობის ინტერვალი უცნობი საშუალოსათვის და აღწერეთ, თუ როგორ იცვლება ნდობის ინტერვალების სიგანე დაკვირვებათა რიცხვის ზრდასთან ერთად.

8.12. ერთ-ერთი დიდი ბანკის კლიენტთაგან შემთხვევით შეირჩა 25 კლიენტი და შეგროვებული იქნა შემდეგი მონაცემები: 1. თითოეული მათგანის საბალანსო ანგარიშზე მყოფი თანხების სიდიდე; 2. იმ კლიენტთა პროპორცია შერჩევაში, რომლებსაც გააჩნიათ სადებიტორო ბარათები ($\hat{p}$ =0.433). ამ მონაცემების საფუძველზე შესაფასებელია ბანკის კლიენტთა საბალანსო ანგარიშის საშუალო სიდიდე (ბანკის ყველა კლიენტის საბალანსო ანგარიშისაგან შედგენილი პოპულაციის საშუალო) და სადებიტორო ბარათის მქონე კლიენტთა პროპორცია.

საბალანსო ანგარიშისათვის გამოთვლილი იქნა შერჩევითი საშუალო $\overline{X}_n$ =$1499.9 და სტანდარტული გადახრა S '=596.9.

ა) რომელ პარამეტრულ მოდელს გამოიყენებთ ამ პირობებში?

ბ) ააგეთ 99%-იანი ნდობის ინტერვალი უცნობი საშუალოსა და დისპერსიისათვის.

გ) ააგეთ 95%-იანი ნდობის ინტერვალი სადებიტორო ბარათის მქონე კლიენტთა პროპორციისათვის.

8.13. შესაფასებელ პარამეტრს წარმოადგენს იმ მყიდველთა პროპორცია პოპულაციაში, რომლებიც ამჯობინებს X მარკის კონსერვირებულ პროდუქციას Y მარკისას. ამ მიზნით შემთხვევითად შერჩეულ 100 ადამიანს დაურიგდა X და Y მარკის კონსერვები ეტიკეტების გარეშე და იმ ფაქტორთა ჩამონათვალი, რომელთა მიხედვითა უნდა შეფასებულიყო კონსერვირებული პროდუქციის ხარისხი. ის მარკა, რომელიც მიიღებდა ყველაზე მაღალ აგრეგირებულ შეფასებას ფაქტორების მიხედვით, ჩაითვლებოდა უკეთესად. მიღებული იქნა შედეგები:

იმ პირთა რაოდენობა, რომლებიც ამჯობინებენ X მარკის კონსერვს – 59,

იმ პირთა რაოდენობა, რომლებიც ამჯობინებენ Y მარკის კონსერვს – 37,

დანარჩენები – 4.

ა) შეაფასეთ პოპულაციის უცნობი პროპორცია.

ბ) ააგეთ 99%-იანი ნდობის ინტერვალი.

8.14. ამომრჩეველთა სიიდან აღებული იქნა n=1000 მოცულობის შემთხვევითი შერჩევა. თითოეულ ამომრჩეველს უნდა შეეფასებინა პრეზიდენტის პოპულარობა მიმდინარე მომენტში. ამ მიზნით ამომრჩეველს დადებითი პასუხი უნდა გაეცა მხოლოდ ერთზე შემდეგი სამი წინადადებიდან:

1. პრეზიდენტი კარგად ასრულებს მოვალეობას;
2. პრეზიდენტი ვერ ასრულებს მოვალეობას;
3. აზრი არ გამაჩნია.

პირველ წინადადებაზე დადებითი პასუხი გასცა 59%-მა. ამ მონაცემების საფუძველზე:

ა) შეაფასეთ იმ პირთა პროპორცია პოპულაციაში, რომლებიც დადებითად აფასებენ პრეზიდენტის მოღვაწეობას;.

ბ) ააგეთ 95%-იანი ნდობის ინტერვალი ამ პროპორციისათვის.

ლექცია IX,X,XI პარამეტრულ ჰიპოთეზათა შემოწმება

ამ ლექციაში ჩვენ გავეცნობით ე.წ. **პარამეტრულ ჰიპოთეზათა შემოწმების მეთოდებს.** საზოგადოდ, რაიმე **მოსაზრებას პოპულაციის განაწილების კანონის ან მისი რიცხვითი მახასიათებლების შესახებ სტატისტიკური ჰიპოთეზა ეწოდება.** სტატისტიკურ ჰიპოთეზებს H სიმბოლოთი აღნიშნავენ (პირველი ასო ინგლისური სიტყვისა **Hypothesis** – ჰიპოთეზა). ჰიპოთეზას ეწოდება **მარტივი**, თუ ის ცალსახად განსაზღვრავს გენერალური ერთობლიობის განაწილების კანონს. წინააღმდეგ შემთხვევაში ჰიპოთეზას ეწოდება **რთული**. ჩვენ დავუშვებთ, რომ **პოპულაციის განაწილების კანონის ანალიზური სახე ცნობილია**, მაგრამ განაწილება **დამოკიდებულია რაიმე θ უცნობ პარამეტრზე**. ამ შემთხვევაში ჰიპოთეზა ფაქტიურად არის **მოსაზრება უცნობი პარამეტრის რიცხვითი მნიშვნელობის შესახებ**, რის გამოც ასეთ ჰიპოთეზებს **პარამეტრულ ჰიპოთეზებს** უწოდებენ.

პარამეტრულ ჰიპოთეზათა შემოწმებისას **ძირითად** (ნულოვან) H_0 ჰიპოთეზასთან ერთად განიხილება **ალტერნატიული** (საწინააღმდეგო) H_1 ჰიპოთეზა. ვთქვათ, გენერალური ერთობლიობის **უცნობი θ პარამეტრის შესახებ შესამოწმებელია შემდეგი სახის ჰიპოთეზები:**

$$\begin{array}{lll} \text{ა) } \boldsymbol{H_0: \theta=\theta_0,} & \text{ბ) } \boldsymbol{H_0: \theta=\theta_0,} & \text{გ) } \boldsymbol{H_0: \theta=\theta_0,} \\ \boldsymbol{H_1: \theta>\theta_0,} & \boldsymbol{H_1: \theta<\theta_0,} & \boldsymbol{H_1: \theta \neq \theta_0,} \end{array} \qquad (9.1)$$

ყველგან θ **წარმოადგენს რაიმე ცნობილ რიცხვს**. ა), ბ) და გ) ჰიპოთეზებში რთულ H_1 არტერნატივებს, შესაბამისად, ეწოდება **მარჯვენა ცალმხრივი, მარცხენა ცალმხრივი და ორმხრივი ალტერნატივა.**

ჰიპოთეზათა შემოწმების ამოცანა მდგომარეობს **ისეთი წესის – სტატისტიკური კრიტერიუმის (ან მოკლედ კრიტერიუმის) ჩამოყალიბებაში**, რომელიც საშუალებას მოგვცემს პოპულაციიდან **მიღებული $x_1,x_2,...,x_n$ შერჩევის საფუძველზე მივიღოთ გადაწყვეტილება**: ან **უარვყოთ ნულოვანი ჰიპოთეზა**, ან ჩავთვალოთ, რომ მონაცემები **ნულოვანი ჰიპოთეზის უარყოფის საშუალებას არ გვაძლევს.** ამისათვის განიხილავენ სპეციალურად შერჩეულ **$T_n=T(X_1,X_2,...,X_n)$ სტატისტიკას**, რომლის შესაძლო მნიშვნელობათა სიმრავლეს ყოფენ ორ **თანაუკვეთ U_0 და U_1 არედ**. $x=(x_1,x_2,...,x_n)$ დაკვირვებებისათვის**, თუ შეფასება- $t_n=T_n(x)\in U_1$, მაშინ H_0 ჰიპოთეზას უარვყოფთ, ანუ გადაწყვეტილებად მივიჩნევთ H_1 ჰიპოთეზას, ხოლო თუ $t_n=T_n(x)\in U_0$, მაშინ ვიტყვით, რომ H_0 ჰიპოთეზის უარყოფის საფუძველი არ არსებობს.** გადაწყვეტილების მიღების ასეთი წესის გამო, U_0 სიმრავლეს H_0 ჰიპოთეზის **მიღების არეს**, ხოლო U_1 სიმრავლეს **კრიტიკულ არეს** უწოდებენ. $T_n=T(X_1,X_2,...,X_n)$ სტატისტიკას **კრიტერიუმის სტატისტიკა** (ანუ ტესტის სტატისტიკა) ეწოდება. სასურველია, რომ კრიტერიუმის სტატისტიკა იყოს მარტივად გამოსათვლელი და რაც უფრო მნიშვნელოვანია, უნდა ვიცოდეთ **მისი ზუსტი ან ზღვარითი განაწილება.**

ჰიპოთეზათა გარჩევის ზემოთ ჩამოყალიბებული წესის გამოყენებისას ჩვენ შეიძლება დავუშვათ შემდეგი **ორი სახის შეცდომა**:

1) უარვყოთ H_0 ჰიპოთეზა მაშინ, როდესაც ის სამართლიანია. ასეთ შეცდომას **პირველი გვარის შეცდომა** ეწოდება.

2) არ უარვყოთ H_0 ჰიპოთეზა მაშინ, როდესაც ის სამართლიანი არ არის. ასეთ შეცდომას **მეორე გვარის შეცდომა** ეწოდება.

9.1 ცხრილში ნაჩვენებია გადაწვეტილების მიღების შესაძლო ვარიანტები და შესაბამისი შეცდომები.

ცხრილი **9.1**

	გადაწყვეტილება	
	არ უარვყოფთ H_0-ს $(T_n(x)\in U_0)$	უარვყოფთ H_0-ს $(T_n(x)\in U_1)$
სამართლიანია H_0	მიღებულია სწორი გადაწყვეტილება	მიღებულია არასწორი გადაწყვეტილება, პირველი გვარის შეცდომა
სამართლიანია H_1	მიღებულია არა სწორი გადაწყვეტილება, მეორე გვარის შეცდომა	მიღებულია სწორი გადაწყვეტილება

პირველი გვარის შეცდომის ალბათობას α ასოთი აღნიშნავენ

$$\alpha\equiv P\{T_n\in U_1|H_0\}, \tag{9.2}$$

სადაც ჩანაწერი $P\{T_n\in U_1|H_0\}$ აღნიშნავს $\{T_n\in U_1\}$ ხდომილობის ალბათობას H_0 ჰიპოთეზის სამართლიანობის პირობებში.

სასურველია კრიტიკული არე ისე შეირჩეს, რომ პირველი გვარის **შეცდომის ალბათობა იყოს მცირე.** როგორც წესი, წინასწარ აფიქსირებენ პირველი გვარის შეცდომის ალბათობის რაიმე მცირე α დონეს, რის შემდეგ განსაზღვრავენ კრიტიკულ არეს. **α რიცხვს კრიტერიუმის მნიშვნელოვნობის დონე ეწოდება.**

α დონე ამოცანის შინაარსიდან აირჩევა; ხშირად α=0.05 (ან 0.01-ს), რაც იმას ნიშნავს, რომ ასიდან საშუალოდ ხუთ (ერთ) შემთხვევაში შესაძლოა უარვყოთ სამართლიანი H_0 ჰიპოთეზა.

H_0 ჰიპოთეზის სამართლიანობისას $\{T_n(X)\in U_1\}$ ხდომილობის ალბათობა არ აღემატება α-ს, α კი იმდენად მცირეა, რომ α-ს ტოლი ალბათობის მქონე ხდომილობის განხორციელება ერთეულოვან ცდაში პრაქტიკულად შეუძლებელად შეგვიძლია ჩავთვალოთ, ამიტომ ის ფაქტი, რომ T_n სტატისტიკის დაკვირვებული მნიშვნელობა აღმოჩნდა U_1 არეში ანუ $\{T_n(X)\in U_1\}$ ხდომილობა მაინც განხორციელდა, **მეტყველებს იმის სასარგებლოდ, რომ H_0 ჰიპოთეზა მცდარია.**

ამრიგად, H_0 ჰიპოთეზის შემოწმების წესი შემდეგნაირად შეგვიძლია ჩამოვაყალიბოთ: **წინასწარ დასახელებული α-სათვის, თუ მოცემული $x=(x_1,x_2,...,x_n)$ რეალიზაცია ისეთია, რომ $t_n=T_n(x)\in U_1$, მაშინ H_0 ჰიპოთეზას უარვყოფთ. წინააღმდეგ შემთხვევაში, $(t_n\notin U_1)$ H_0 ჰიპოთეზის უარყოფის საფუძველი არ არსებობს**. შევნიშნოთ, რომ ფაქტი $t_n\notin U_1$ არ ამტკიცებს, რომ ჰიპოთეზა სამართლიანია; ის მხოლოდ მიუთითებს იმაზე, რომ მონაცემებსა და თეორიულ დაშვებებს შორის თანხმობა საკმარისად კარგია (α დონეზე). საბოლოო შედეგი მნიშვნელოვნად ითვლება მაშინ, როცა H_0 ჰიპოთეზა უარყოფილია.

ჩვენ განვიხილავთ შემთხვევებს, როდესაც უცნობი პარამეტრი, რომლისთვისაც მოწმდება (9.1) სახის ჰიპოთეზა, არის

- **ნორმალური პოპულაციის მათემატიკური ლოდინი μ, როდესაც საშუალო კვადრატული გადახრა ცნობილია $(X\in N(\mu,\sigma_0)$ მოდელი)**

- **ნორმალური პოპულაციის მათემატიკური ლოდინი μ, როდესაც საშუალო კვადრატული გადახრა უცნობია ($X\in N(\mu,\sigma)$ მოდელი)**
- **ნორმალური პოპულაციის σ^2 დისპერსია ($X\in N(\mu,\sigma)$ მოდელი)**
- **წარმატების უცნობი ალბათობა p ბერნულის სქემაში ($X\in Bi(1, p)$ მოდელი)**

§1. ჰიპოთეზათა შემოწმების ამოცანა ნორმალური პოპულაციის საშუალოსათვის

დავუშვათ **ნორმალური პოპულაციის** მათემატიკური ლოდინის შესახებ გამოთქმულია **შემდეგი ჰიპოთეზები**

ა) $H_0: \mu=\mu_0$, $H_1: \mu>\mu_0$, ბ) $H_0: \mu=\mu_0$, $H_1: \mu<\mu_0$, გ) $H_0: \mu=\mu_0$, $H_1: \mu\neq\mu_0$, (9.3)

სადაც ყველგან μ_0 წარმოადგენს რაიმე ცნობილ რიცხვს. უნდა გავარჩიოთ ორი შემთხვევა:

1) პოპულაციის საშუალო კვადრატული გადახრა σ ცნობილია და ის σ_0-ის ტოლია ($X\in N(\mu,\sigma_0)$ მოდელი),
2) პოპულაციის საშუალო კვადრატული გადახრა σ უცნობია ($X\in N(\mu,\sigma)$ მოდელი).

$N(\mu,\sigma_0)$ მოდელის შემთხვევაში, H_0 ჰიპოთეზა მარტივია, ხოლო ალტერნატივა რთულია. $N(\mu,\sigma)$ მოდელის შემთხვევაში, როგორც ნულოვანი, ისე ალტერნატიული ჰიპოთეზები რთულია.

§ 1.1. ჰიპოთეზათა შემოწმება $X\in N(\mu,\sigma_0)$ მოდელისათვის

ა) **ალტერნატივა** მიუთითებს მხოლოდ იმ ფაქტზე, რომ $\mu>\mu_0$. კრიტერიუმის სტატისტიკად ავირჩიოთ $\overline{X}_n$ შერჩევითი საშუალო, რომელიც გენერალური საშუალოს ჩაუნაცვლებელი შეფასებაა. ამის გამო $x_1,x_2,\ldots,x_n$ შერჩევის საფუძველზე გა-მოთვლილი $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ -ის მნიშვნელობა, როგორც წესი, ახლოს უნდა იყოს უცნობი μ პარამეტრის ჭეშმარიტ მნიშვნელობასთან: $\overline{X}_n$-მა დიდი ალბათობით უნდა მიიღოს რიცხვითი მნიშვნელობა μ_0–ის „მიდამოდან" ნულოვანი ჰიპოთეზის სამართლიანობის შემთხვევაში და μ_0-ზე „დიდი" რიცხვითი მნიშვნელობა მარჯვენა ცალმხრივი ($\mu>\mu_0$) ალტერნატივის სამართლიანობის შემთხვევაში. ე.ი. $\overline{X}_n$-ის მიერ „დიდი" მნიშვნელობის მიღება მეტყველებს H_1 ჰიპოთეზის სასარგებლოდ. ამიტომ კრიტიკული არე განვსაზღვროთ შემდეგნაირად $U_1=(x_{კრ},+\infty)$.

ნულოვანი ჰიპოთეზის სამართლიანობის შემთხვევაში ტესტის $\overline{X}_n$ სტატისტიკას აქვს $N(\mu_0, \sigma_0/\sqrt{n})$ განაწილება. თუ დავაფიქსირებთ პირველი გვარის შეცდომის α ალბათობის რიცხვითი მნიშვნელობის, მაშინ მოსაძებნია ისეთი $x_{კრ}$, რომლისთვისაც

$$P\{\overline{X}_n\geq x_{კრ}|H_0\}=P\left\{\frac{\overline{X}_n-\mu_0}{\sigma_0/\sqrt{n}}\geq\frac{x_{კრ}-\mu_0}{\sigma_0/\sqrt{n}}\right\}=1-\Phi\left(\frac{x_{კრ}-\mu_0}{\sigma_0/\sqrt{n}}\right)=\alpha, \quad (9.4)$$

საიდანაც კრიტიკული წერტილი $x_{კრ}$ განისაზღვრება შემდეგი ფორმულით

$$x_{კრ} = \mu_0 + x_{1-\alpha} \cdot \frac{\sigma_0}{\sqrt{n}}\ , \qquad (9.5)$$

სადაც $x_{1-\alpha}$ არის სტანდარტული ნორმალური განაწილების **(1–α)-კვანტილი**. $x_{1-\alpha}$ კვანტილს ასევე უწოდებენ სტანდარტული ნორმალური განაწილების **ზედა α-კრიტიკულ წერტილს** და z_α სიმბოლოთი აღნიშნავენ: $x_{1-\alpha} = z_\alpha$.

განხილულ შემთხვევაში, ასევე ბუნებრივია, **ტესტის სტატისტიკად** ავირჩიოთ სტანდარტიზებული შერჩევითი საშუალო, რომლისთვისაც კრიტიკული არეა სტანდარტული ნორმალური განაწილების ზედა (მარჯვენა) α (იხ. (9.4)) ზომის არე

$$U_1 = \{z_\alpha, +\infty\}. \qquad (9.6)$$

მართლაც, **დავუშვათ კრიტიკული არეა** $U_1 = \{c_\alpha, +\infty\}$. თუ დავაფიქსირებთ პირველი გვარის შეცდომის α ალბათობას, მაშინ

$$P\{Z_n \geq c_\alpha \mid H_0\} = 1 - \Phi(c_\alpha) = \alpha \qquad (9.7)$$

და

$$c_\alpha = x_{1-\alpha} = z_\alpha$$

ამრიგად $\overline{X}_n$ და Z_n სტატისტიკებისათვის აგებული კრიტიკული არეების შესაბამის $x_{კრ}$ და z_α კრიტიკულ წერტილებს შორის კავშირი მოიცემა (9.5) ფორმულით.

კვლავ ჩამოვაყალიბოთ ჰიპოთეზათა გარჩევის კრიტერიუმი:

თუ მოცემული $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)$ შერჩევითი მნიშვნელობისათვის და α მნიშვნელოვნობის დონისათვის

$$\overline{x}_n > \mu_0 + z_\alpha \cdot \frac{\sigma_0}{\sqrt{n}} \qquad (\text{ან} \quad z_n \geq z_\alpha), \qquad (9.8)$$

მაშინ $\boldsymbol{H_0}$ **ჰიპოთეზას უარვყოფთ**, წინააღმდეგ შემთხვევაში H_0 ჰიპოთეზის **უარყოფის საფუძველი არ გვაქვს.**

საბოლოოდ პოთეზათა შემოწმების ამოცანის გადაწყვეტა შეიძლება დაიყოს შემდეგ ეტაპებად:

I. **სტატისტიკური მოდელის განსაზღვრა**

II. $\boldsymbol{H_0}$ **და** $\boldsymbol{H_1}$ **ჰიპოთეზების ჩამოყალიბება**

III. **კრიტერიუმის** $\boldsymbol{T_n}$ **სტატისტიკის შერჩევა**

IV. **მნიშვნელოვნობის** α **დონის მითითება და** $\boldsymbol{U_1}$ **კრიტიკული არის დადგენა**

V. $\boldsymbol{t_n = T_n(x)}$**-ის რიცხვითი მნიშვნელობის გამოთვლა,**

VI. **გადაწყვეტილების მიღება**

ჰიპოთეზათა შემოწმების მეორე ეკვივალენტური გზა მდგომარეობს შემდეგში: $x_1,x_2,\ldots,x_n$ შერჩევითი მნიშვნელობების საფუძველზე გამოვთვალოთ ტესტის Z_n სტატისტიკის $z_n = \dfrac{\overline{x}_n - \mu_0}{\sigma_0 / \sqrt{n}}$ რიცხვითი მნიშვნელობა და შემდეგი ალბათობა, **რომელსაც** $\boldsymbol{P}$**-მნიშვნელობას უწოდებენ**

$$P\text{-მნიშვნელობას} = P\{Z_n \geq z_n \mid H_0\} = 1 - \Phi(z_n). \qquad (9.9)$$

ცხადია, რომ თუ P-მნიშვნელობა$\le\alpha$, მაშინ $z_n \ge z_\alpha$, რაც იწვევს H_0 ჰიპოთეზის უარყოფას. ეს გვაძლევს შემდეგი დასკვნის გაკეთების შესაძლებლობას: $x_1,x_2,\ldots,x_n$ შერჩევითი მნიშვნელობების საფუძველზე ნებისმიერი α, $\alpha \ge P$ მნიშვნელოვნობის დონით ხდება H_0 ჰიპოთეზის უარყოფა. მაშასადამე P-მნიშვნელობა არის ის მინიმალური მნიშვნელოვნობის დონე, რომლითაც $x_1,x_2,\ldots,x_n$ მონაცემებით H_0 ჰიპოთეზას უარყყოფთ. საბოლოოდ, შეიძლება ჩამოვაყალიბოთ **ჰიპოთეზათა გარჩევის შემდეგი კრიტერიუმი**:

მოცემული α მნიშვნელოვნობის დონისათვის, მონაცემების საფუძველზე H_0 ჰიპოთეზას უარვყოფთ, თუ P-მნიშვნელობა $\le \alpha$.

ახლა განვიხილოთ ბ) და გ) ალტერნატივები:

ბ) ალტერნატივა. ანალოგიური მსჯელობის საფუძველზე მარცხენა ცალმხრივი ალტერნატივისათვის და ტესტის Z_n სტატისტიკისთვის კრიტიკული არეა სტანდარტული ნორმალური განაწილების ქვედა α ზომის არე: $U_1=(-\infty,-z_\alpha]$, ხოლო

$$P\text{-მნიშვნელობა} = P\{Z_n \le z_n \mid H_0\} = \Phi(z_n).$$

გ) ალტერნატივა. ამ შემთხვევაში ალტერნატივა ორმხრივია. $z_n=\left|\frac{\bar{x}_n-\mu_0}{\sigma_0/\sqrt{n}}\right|$-ის დიდი მნიშვნელობები მეტყველებს H_1 ალტერნატივის სასარგებლოდ. ტესტის Z_n სტატისტიკისთვის α მნიშვნელოვნობის დონის კრიტიკული არე შედგება ორი ნაწილისაგან: სტანდარტული ნორმალური განაწილების $\alpha/2$ ზომის ქვედა და $\alpha/2$ ზომის ზედა კრიტიკული არეების გაერთიანებისაგან: $U_1=(-\infty\,,-z_{\alpha/2}]\cup[z_{\alpha/2}\,,+\infty)$.

$$P\text{-მნიშვნელობა} = 2\cdot P\{Z_n \ge |z_n| \mid H_0\} = 2\cdot(1-\Phi(z_n)).$$

***P*-მნიშვნელობა არის მნიშვნელოვნობის დონის ის მინიმალური მნიშვნელობა, რომლითაც $x_1,x_2,\ldots,x_n$ მონაცემების საფუძველზე H_0 ჰიპოთეზას უარყყოფთ.**

ჰიპოთეზათა შემოწმების ალგორითმი $X\in N(\mu,\sigma_0)$ მოდელის საშუალოსათვის

ჰიპოთეზა: $H_0 : \mu=\mu_0$

მნიშვნელოვნობის დონე: α

ტესტის სტატისტიკა: $Z_n = \dfrac{\bar{X}_n-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

კრიტერიუმის სტატისტიკის მნიშვნელობა: $z_n = \dfrac{\bar{x}_n-\mu_0}{\sigma/\sqrt{n}}$

ალტერნატივა	კრიტიკული არე U_1 (H_0-ის უარყოფის არე)
$H_1 : \mu=\mu_1>\mu_0$	$[z_\alpha,+\infty)$
$H_1 : \mu=\mu_1<\mu_0$	$(-\infty,-z_\alpha]$
$H_1 : \mu\ne\mu_1$	$(-\infty\,,-z_{\alpha/2}]\cup[z_{\alpha/2}\,,+\infty)$

სადაც z_α არის $N(0,1)$ განაწილების ზედა α-კრიტიკული წერტილი.

გადაწყვეტილება: თუ $z_n = \frac{\bar{x}_n-\mu_0}{\sigma_0/\sqrt{n}} \in U_1$, H_0 ჰიპოთეზას უარვყოფთ, წინააღმდეგ შემთხვევაში H_0-ის უარყოფის საფუძველი არ გვაქვს.

შენიშვნა.

როდესაც შერჩევა მიღებულია **არა ნორმალური** პოპულაციიდან, მაგრამ შერჩევის **მოცულობა *n* საკმაოდ დიდია (*n*>30),** ცენტრალური ზღვარითი თეორემის ძალით, $\overline{X}_n$ შემთხვევითი სიდიდე მიახლოებით ნორმალურადაა განაწილებული, რის გამოც **სტატისტიკური ჰიპოთეზების შემოწმების ზემოთ მოყვანილი კრიტერიუმები ამ შემთხვევაშიც გამოიყენება.**

მაგალითი 1. სამრეცხაოს მფლობელს სურს გაყიდოს თავისი საწარმო. იგი პოტენციურ მყიდველს არწმუნებს, რომ საწარმოს დღიური შემოსავალი უკანასკნელი ხუთი წლის მანძილზე შეადგენდა $675-ს, სტანდარტული გადახრით $75. შემთხვევით არჩეული 25 დღის საშუალო შემოსავალმა შეადგინა $638. α=0.01 მნიშვნელოვნობის დონით აქვს თუ არა საფუძველი მყიდველს ამტკიცოს, რომ საწარმოს მფლობელი არ არის სწორი? განსაზღვრეთ *P*-მნიშვნელობა.

ამოხსნა: ამოხსნისას ვიხელმძღვანელოდ ამოცანის გადაწყვეტის ზემოთ მოცემული თანმიმდევრობით

I. სტატისტიკური მოდელის განსაზღვრა

ვინაიდან საწარმოს ყოველდღიური შემოსავალი მრავალ ფაქტორზეა დამოკიდებული (კლიენტების რაოდენობა, შეკვეთების მოცულობა, შესრულების ვადა და სხვა), ცენტრალური ზღვარითი თეორემის თანახმად ჩავთვალოთ, რომ აღნიშნულ რეალურ სიტუაციას კარგად აღწერს $N(\mu,\sigma_o)$, σ_0=75, მოდელი.

II. H_0 და H_1 ჰიპოთეზების ჩამოყალიბება

მყიდველისათვის მეტად მნიშვნელოვანია შეამოწმოს, რომ საწარმოს ყოველდღიური შემოსავალი არ იყოს 675-ზე ნაკლები. ამიტომ ჰიპოთეზები ჩამოვაყალიბოდ შემდეგნაირად

H_0 : μ=675,

H_1: μ<675□.

III. კრიტერიუმის სტატისტიკის შერჩევა.

ტესტის სტატისტიკა იყოს Z_n სტატისტიკა.

IV. მნიშვნელოვნობის α დონის მითითება და U_1 კრიტიკული არის დადგენა

პირობის თანახმად α=0.01, ხოლო კრიტიკული არია $(-\infty, -z_{0,01})= (-\infty, -2,33)$

V. z_n-ის რიცხვითი მნიშვნელობის გამოთვლა.

$z_n= (638-675)\cdot 5/75 = -2,47$

VI. გადაწყვეტილების მიღება.

ვინაიდან $z_n=-2,47<-2,33=-z_{0,01}$, ნულოვან ჰიპოთეზას ვუარყოფთ, ე.ი. ჩავთვლით, რომ საწარმოს მფლობელი არ არის სწორი.

VII. *P*-მნიშვნელობის განსაზღვრა.

P-მნიშვნელობა=$P(Z_n<-2,47)$=0,07.

§ 1.2. ჰიპოთეზათა შემოწმება $X \in N(\mu,\sigma)$ მოდელისათვის

$N(\mu,\sigma)$ მოდელის შემთხვევაში, როგორც ნულოვანი, ისე ალტერნატიული ჰიპოთეზა რთულია. უცნობ μ და σ^2 პარამეტრების „კარგი" შემფასებლებია **შერჩევითი საშუალო და შესწორებული შერჩევითი დისპერსია.** ამასთან

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{და} \quad \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$$

დამოუკიდებელი შემთხვევითი სიდიდეებია. თუ ახლა გავიხსენებთ სტიუდენტის t განაწილების განმარტებას, მივიღებთ, რომ

$$T_n = \frac{\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S'^2}{\sigma^2} \Big/ (n-1)}} = \frac{\overline{X}_n - \mu}{S'/\sqrt{n}} \sim t(n\text{-}1).$$

$N(\mu,\sigma)$ მოდელის შემთხვევაში პოპულაციის მათემატიკური ლოდინის შესახებ (9.3) ჰიპოთეზების შემოწმებისას ტესტის სტატისტიკაა

$$T_n = \frac{\overline{X} - \mu_0}{S'/\sqrt{n}} \sim t(n\text{-}1), \tag{9.10}$$

რომელიც არ შეიცავს მოდელის σ– უცნობ პარამეტრს და რომლის საფუძველზეც, სტანდარტული გზით, აიგება ალტერნატივის შესაბამისი კრიტიკული არე.

აღნიშნული კრიტერიუმი გამოიყენება, როცა $n<30$. როდესაც n იზრდება, როგორც ვიცით, სტიუდენტის განაწილება **უახლოვდება ნორმალურ განაწილებას**, რის გამოც განაწილების **კრიტიკული წერტილების მოსაძებნად სარგებლობენ სტანდარტული ნორმალური განაწილებით.**

ჰიპოთეზათა შემოწმების ალგორითმი $N(\mu,\sigma)$ მოდელის საშუალოსათვის
(პატარა მოცულობის შერჩევის შემთხვევაში, $n<30$)

ჰიპოთეზა: $H_0 : \mu=\mu_0$
მნიშვნელოვნობის დონე: α

კრიტერიუმის სტატისტიკის მნიშვნელობა: $T_n = \dfrac{\overline{X}_n - \mu_0}{S'/\sqrt{n}} \sim t(n\text{-}1)$

კრიტერიუმის სტატისტიკის მნიშვნელობა: $t_n = \dfrac{\overline{x}_n - \mu_0}{s'/\sqrt{n}}$

ალტერნატივა	კრიტიკული არე U_1 (H_0-ის უარყოფის არე)
$H_1 : \mu=\mu_1>\mu_0$	$[\, t_{n-1,\,\alpha}, +\infty)$
$H_1 : \mu=\mu_1<\mu_0$	$(-\infty, -t_{n-1,\,\alpha}\,]$
$H_1 : \mu\neq\mu_1$	$(-\infty, -t_{n-1,\,\alpha/2}\,] \cup [\, t_{n-1,\,\alpha/2}, +\infty)$

გადაწყვეტილება: თუ $t_n = \dfrac{\overline{x}_n-\mu_0}{s'/\sqrt{n}} \in U_1$, H_0 ჰიპოთეზას უარვყოფთ, წინააღმდეგ შემთხვევაში H_0-ის უარყოფის საფუძველი არ გვაქვს.

მაგალითი 2. ახალი ელექტრონული ხელსაწყოს შემოწმების მიზნით, კომპანიამ გადაწყვიტა გამოუშვას მხოლოდ ექვსი ასეთი ხელსაწყო. ყოველი მათგანისათვის გაიზომა დრო ხელსაწყოს მწყობრიდან გამოსვლამდე, რის შედეგადაც მიღებული იქნა შემდეგი ამონარჩევი: 59.2, 68.3, 57.8, 56.5, 63.7 და 57.3 საათი. გვაძლევს თუ არა ეს შერჩევა საკმარის საფუძველს α=0.05 მნიშვნელოვნობის დონით შემდეგი დასკვნის გასაკეთებლად: ახალი ხელსაწყოს სიცოცხლის საშუალო ხანგრძლიობა აღემატება 55 საათს.

ამოხსნა. ვიხელმძღვანელოდ ამოცანის გადაწყვეტის ზევით მოცემული თანმიმდევრობით

I. სტატისტიკური მოდელის განსაზღვრა

ვინაიდან ელექტრონული ხელსაწყოს სიცოცხლის ხანგრძლიობა მრავალ ფაქტორზეა დამოკიდებული, ცენტრალური ზღვარითი თეორემის თანახმად ჩავთვალოთ, რომ ამოცანის პირობები აღიწერება $N(\mu,\sigma)$, მოდელით.

II. H_0 და H_1 ჰიპოთეზების ჩამოყალიბება

კომპანიისათვის მეტად მნიშვნელოვანია, რომ ხელსაწყოს სიცოცხლის საშუალო ხანგრძლიობა აღემატებოდეს 55 საათს. ამიტომ ჰიპოთეზები ჩამოვაყალიბოდ შემდეგნაირად

H_0 : μ=55;

H_1 : μ>55.

III . კრიტერიუმის სტატისტიკის შერჩევა.

ტესტის სტატისტიკაა T_n სტატისტიკა.

IV . მნიშვნელოვნობის α დონის მითითება და U_1 კრიტიკული არის დადგენა

პირობის თანახმად α=0.05, n=6, U_1= $(t_{n-1,\,\alpha},+\infty,)$= (2,015, $+\infty$)

V. t_n-ის რიცხვითი მნიშვნელობის გამოთვლა.

$$\bar{x}_n = 60.4667,\quad s'^2 = = 21.2587.,\qquad t_n = \frac{\bar{x}_n - 55}{s'/\sqrt{n}} = 2.9$$

VI. გადაწყვეტილების მიღება.

რადგან t სტატისტიკის დაკვირვებული მნიშვნელობა მეტია 2.015-ზე ჩვენ უკუვაგდებთ H_0 ჰიპოთეზას, ანუ არის საკმარისი საფუძველი მტკიცების, რომ ახალი ხელსაწყოს საშუალო სიცოცხლის ხანგრძლივობა აღემატება 55 საათს.

VII. P-მნიშვნელობის განსაზღვრა.

P-მნიშვნელობა გამოითვლება EXCEL-ის შესაბამისი სტატისტიკური ფუნქციის გამოყენებით.

P-მნიშვნელობა=0.017. ♦

მაგალითი 3. ნორმალურად განაწილებული გენერალური ერთობლიობისათვის მიღებულია 50 მოცულობის შერჩევა. შერჩევის საფუძველზე გამოთვლილი შერჩევითი საშუალო და შესწორებული შერჩევითი დისპერსია შესაბამისად 10,7 და 7,4-ის ტოლია. 0.05 მნიშვნელოვნების დონით შეამოწმეთ H_0:μ=10 ნულოვანი ჰიპოთეზა H_1:$\mu \neq 10$ ალტერნატივის მიმართ:

ა) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?

ბ) შეიცვლება თუ არა პასუხი, თუ ალტერნატიულ ჰიპოთეზად განვიხილავთ H_1:μ>10

ამოხსნა. ამოცანის პირობები აღიწერება $N(\mu,\sigma)$ მოდელით, α=0,05, $\bar{x}_n$ =10,7, s'^2 =7,4, n=50. ტესტის სტატისტიკაა T_n სტატისტიკა, რომლის დაკვირვებული მნიშვნელობაა

$$t_n = \frac{\bar{x}_n - \mu_o}{s'/\sqrt{n}} = 1{,}81$$

ვინაიდან n>30, ორმხრივი კრიტიკული $U_1=(-\infty, -x_{კრ})\cup(x_{კრ},+\infty)$ არის კრიტიკული $x_{კრ}$ წერტილის მოსაძებნად უნდა ვისარგებლოდ სტანდარტული ნორმალური განაწილებით. $x_{კრ}=z_{0,025}=1{,}96$, $H_1:\mu\neq 10$ ალტერნატივის შემთხვევაში და $x_{კრ}=z_{0,05}=1{,}64$, $H_1:\mu>10$ ალტერნატივის შემთხვევაში.

ამრიგად, $H_1:\mu\neq 10$ ალტერნატივის შემთხვევაში: α=0,05 მნიშვნელოვნობის დონით, ნულოვანი ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს, ხოლო, $H_1:\mu>10$ ალტერნატივის შემთხვევაში, უარვყოფთ ნულოვან ჰიპოთეზას ალტერნატივის სასარგებლოდ. ♦

§ 2. ჰიპოთეზათა შემოწმება $N(\mu,\sigma)$ მოდელის დისპერსიისათვის

ახლა განვიხილოთ შემთხვევა, როდესაც ნორმალური პოპულაციიდან მიღებული შერჩევის საფუძველზე შესამოწმებელია შემდეგი სახის ჰიპოთეზები პოპულაციის დისპერსიის შესახებ

ა) $H_0: \sigma^2=\sigma_0^2$, $H_1: \sigma^2>\sigma_0^2$, (მარჯვენა ცალმხრივი ალტერნატივა)

ბ) $H_0: \sigma^2=\sigma_0^2$, $H_1: \sigma^2<\sigma_0^2$, (მარცხენა ცალმხრივი ალტერნატივა)

გ) $H_0: \sigma^2=\sigma_0^2$, $H_1: \sigma^2\neq\sigma_0^2$. (ორხრივი ალტერნატივა)

ყველგან σ_0^2 წარმოადგენს რაიმე ცნობილ რიცხვს. ვინაიდან **μ უცნობია, როგორც ნულოვანი**, ისე **ალტერნატიული ჰიპოთეზა რთულია**.

რადგან

$$S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 . \qquad (9.11)$$

უცნობი დისპერსიის ჩაუნაცვლებელ შეფასებაა, ამიტომ კრიტერიუმის სტატისტიკაა

$$\chi^2 = \frac{(n-1)S_n'^2}{\sigma_0^2} \sim \chi^2(n-1). \qquad (9.12)$$

ვიცით რა **კრიტერიუმის სტატისტიკის** (ტესტის სტატისტიკის) განაწილების კანონი სტანდარტული გზით აიგება ალტერნატივის შესაბამისი კრიტიკული არე.

ჰიპოთეზათა შემოწმების ალგორითმი ნორმალური პოპულაციის დისპერსიისათვის

(პატარა მოცულობის შერჩევის შემთხვევაში, $n<30$)

ჰიპოთეზა $H_0 : \sigma^2 = \sigma_0^2$

მნიშვნელოვნობის დონე α

კრიტერიუმის სტატისტიკის მნიშვნელობა $\dfrac{(n-1)s'^2}{\sigma_0^2}$

ალტერნატივა	**კრიტიკული არე**
$H_1 : \sigma^2 > \sigma_0^2$,	$U_1 = [\chi^2_{n-1,\alpha}, +\infty)$.
$H_1 : \sigma^2 < \sigma_0^2$,	$U_1 = (0, \chi^2_{n-1,1-\alpha}]$
$H_1 : \sigma^2 \neq \sigma_0^2$,	$U_1 = (0, \chi^2_{n-1,1-\alpha/2}] \cup [\chi^2_{n-1,\alpha/2}, +\infty)$

სადაც $\chi^2_{n,p}$ **არის** $\chi^2(n)$ **განაწილების ზედა** α**-კრიტიკული წერტილი**

მაგალითი 4. ნორმალურად განაწილებული გენერალური ერთობლიობიდან ამოკრეფილია n=24-ის ტოლი მოცულობის შერჩევა და მის საფუძველზე გამოთვლილია შესწორებული შერჩევითი დისპერსია, s'^2=17.3, α=0.05 მნიშვნელოვნობის დონისათვის შევამოწმოთ ჰიპოთეზები

$H_0 : \sigma^2 = 15$,

$H_1 : \sigma^2 > 15$.

ამოხსნა. ამოცანის პირობიდან გამომდინარე საქმე გვაქვს ცალმხრივ მარჯვენა კრიტერიუმთან. χ^2 განაწილების ცხრილიდან α=0.05-სათვის ვპოულობთ ზედა კრიტიკული წერტილის მნიშვნელობას $\chi^2_{23,0.05}$=35.17, ვითვლით კრიტერიუმის სტატისტიკის რიცხვით მნიშვნელობას და რადგან χ^2_{24}=23·17.3/15=26,53<35,17=$\chi^2_{23,0.05}$, ვასკვნით: α=0.05 მნიშვნელოვნობის დონით ნულოვანი ჰიპოთეზის უარყოფის საფუძველი არა გვაქვს.

ამით ჩვენ დავასრულეთ ჰიპოთეზათა **შემოწმების ამოცანების მიმოხილვა ნორმალურად განაწილებული გენერალური ერთობლიობის პარამეტრებისათვის**.

§ 3. უცნობი ალბათობის შესახებ ჰიპოთეზათა შემოწმების ამოცანა ბერნულის სქემაში

ვთქვათ, $X=(X_1,\ldots,X_n)$ n მოცულობის შერჩევაა **ბერნულის კანონით განაწილებული** გენერალური ერთობლიობიდან **წარმატების უცნობი** p **ალბათობით**. ამ შერჩევის საფუძველზე, α მნიშვნელოვნობის დონით, შესამოწმებელია ჰიპოთეზები

ა) $H_0 : p = p_0$, $H_1 : p > p_0$, ბ) $H_0 : p = p_0$, $H_1 : p < p_0$, გ) $H_0 : p = p_0$, $H_1 : p \neq p_0$.

როგორც ვიცით, $\hat{P}_n = S_n/n$ წარმოადგენს უცნობი p ალბათობის ჩაუნაცვლებელ შეფასებას და ნულოვანი ჰიპოთეზის სამართლიანობისას $E\hat{P}_n = p_0$, $D\hat{P}_n = \frac{p_0(1-p_0)}{n}$ ჩვენ შევისწავლით მხოლოდ იმ შემთხვევას, როდესაც შერჩევის ***n* მოცულობა იმდენად დიდია, რომ *n* და *p* პარამეტრების მქონე ბინომური განაწილებისათვის შესაძლებელია ნორმალური აპროქსიმაციის გამოყენება.** ამ შემთხვევაში **კრიტერიუმის სტატისტიკად** გამოიყენება

$$Z_n = \frac{\hat{P}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1) \qquad (9.13)$$

სტატისტიკა.

ჰიპოთეზათა შემოწმების ალგორითმი უცნობი *p* ალბათობის შესახებ ბერნულის სქემაში
(დიდი მოცულობის შემთხვევაში)

ჰიპოთეზა $H_0 : p = p_0$
მნიშვნელოვნობის დონე α

კრიტერიუმის სტატისტიკის დაკვირვებული მნიშვნელობა $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

ალტერნატივა	კრიტიკული არე U_1 (H_0-ის უარყოფის არე)
$H_1 : p>p_0$,	$[z_{\alpha}, +\infty)$
$H_1 : p<p_0$,	$(-\infty, -z_{\alpha}]$
$H_1 : p\neq p_0$,	$(-\infty, -z_{\alpha/2}]\cup[z_{\alpha/2}, +\infty)$

სადაც z_α არის $N(0,1)$ განაწილების ზედა α-კრიტიკული წერტილი.

გადაწყვეტილება: თუ $z \in U_1$, H_0 ჰიპოთეზას უარვყოფთ, წინააღმდეგ შემთხვევაში H_0-ის უარყოფის საფუძველი არა გვაქვს

მაგალითი 5. ვთქვათ, ხარისხის კონტროლის მენეჯერს სურს გაიგოს, აღემატება თუ არა ახალი დანადგარით დამზადებული დეფექტური პაკეტების წილი 10%–ს. ამოცანის გადასაწყვეტად შერჩეული 200 პაკეტიდან 26 დეფექტური აღმოჩნდა. აქვს თუ არა საფუძველი მენეჯერს ამტკიცოს, რომ ახალი დანადგარი არ აკმაყოფილებს მის მოთხოვნებს, თუ ის თანახმაა, რომ მის მიერ მიღებული გადაწყვეტილების რისკი არ აღემატებოდეს 0.05-ს.

ამოხსნა. ახალი დანადგარი არ დააკმაყოფილებს მენეჯერის პირობებს, თუ დეფექტური პაკეტების წილი პოპულაციაში იქნება 10%-ზე მეტი (p>0.1). ამიტომ ის ასე აყალიბებს ჰიპოთეზებს:

$H_0 : p_0$=0.1,
$H_1 : p_1$>0.1.

ნორმალური განაწილების ცხრილებიდან ვპოულობთ, რომ $z_{0.05}$=1.645 და მაშასადამე, კრიტიკული არე იქნება U_1=[1.645;+∞). კრიტერიუმის სტატისტიკის შერჩევითი მნიშვნელობაა

$$z_n = \frac{26 - 200 \cdot 0{,}1}{\sqrt{200 \cdot 0{,}1 \cdot 0.9}} \approx 1.415,$$

რადგან $1.415<1.645$ ანუ $z \notin U_1$, H_0 ჰიპოთეზის უარყოფის საფუძველი არა გვაქვს.

§ 4. კავშირი ჰიპოთეზათა შემოწმებისა და ინტერვალური შეფასების ამოცანებს შორის

უცნობი θ პარამეტრის შესახებ **მარტივი ჰიპოთეზის შემოწმების ამოცანასა** და ამავე პარამეტრისათვის ნდობის **ინტერვალის აგების ამოცანებს შორის არსებობს მჭიდრო კავშირი.** საილუსტრაციოდ დავადგინოთ ურთიერთკავშირი ჰიპოთეზის შემოწმებასა და ნდობის ინტერვალის აგების ამოცანებს შორის ***N*(μ,σ$_0$) მოდელისათვის**.

როგორც ვიცით, ტესტის $\overline{X}_n$ სტატისტიკისთვის, H_0 : μ=μ$_0$ მარტივი ჰიპოთეზის შემოწმებისას H_1 : μ≠μ$_0$ რთული ალტერნატივის წინააღმდეგ α-მნიშვნელოვნობის დონით H_0 ჰიპოთეზის მიღების არეა

$$U_0=(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}} \ , \ \mu_0 + z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}}),$$

ხოლო

$$(\overline{x}_n - z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}} \ , \ \overline{x}_n + z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}})$$

შემთხვევითი ინტერვალი წარმოადგენს (1–α)-ნდობის ინტერვალს μ–სათვის. ორივე ეს გამოსახულება შედეგია იმ ფაქტის, რომ თუ $\overline{X}_n \sim N(\mu, \sigma_0^2 / n)$, მაშინ

$$P\left(- z_{\alpha/2} < \frac{\overline{X}_n - \mu}{\sigma_0 / n} < z_{\alpha/2} \right) = 1 - \alpha \ .$$

ამრიგად, თუ შესამოწმებელია H_0 : μ=μ$_0$ ჰიპოთეზა H_1 : μ≠μ$_0$ ალტერნატივის მიმართ, მაშინ H_0 ჰიპოთეზა უნდა უარვყოთ α მნიშვნელოვნობის დონით, თუ μ$_0$ არ ეკუთვნის (1–α)-ნდობის ალბათობით აგებულ ნდობის ინტერვალს.

§ 5. მარტივი ჰიპოთეზების შემოწმება $X \in N(\mu,\sigma_0)$ მოდელისათვის

ახლა განვიხილოთ მარტივი ჰიპოთეზები ნორმალური მოდელის საშუალოსთვის, როდესაც ალტერნატიული ნაწილი მარტივია.

ვთქვათ, n მოცულობის $X_1,X_2,...,X_n$. შერჩევის საფუძველზე, წინასწარ განსაზღვრული α-მნიშვნელოვნობის დონისათვის, **შესამოწმებელია შემდეგი მარტივი ჰიპოთეზები**

$$\boldsymbol{H_0 : \mu=\mu_0,}$$

$$\boldsymbol{H_1 : \mu=\mu_1, \ \mu_1>\mu_0,} \qquad (9.14)$$

სადაც μ$_0$ და μ$_1$ რაიმე **ცნობილი რიცხვებია.**

იმ შემთხვევაში, როდესაც მოწმდება H_0 : μ=μ_0 ჰიპოთეზა რთული H_1 : μ>μ_1 ალტერნატივის წინააღმდეგ, კრიტიკული არე დამოკიდებული იყო იმ ფაქტზე, რომ μ>μ_0, ამიტომ ბუნებრივია, გამოვიყენოთ **იგივე კრიტერიუმი** მარტივი H_1: μ=μ_1, μ_1>μ_0, ალტერნატივის შემთხვევაშიც. კრიტერიუმის $\bar{X}_n$ სტატისტიკისათვის კრიტიკული არეა განაწილების ზედა α ზომის არე:

$$U_1=[\mu_0+z_{\alpha}\cdot\frac{\sigma_0}{\sqrt{n}},+\infty).$$

მარტივ ჰიპოთეზათა შემოწმების ალგორითმი $X\in N(\mu,\sigma_0)$ მოდელისათვის

ჰიპოთეზა: H_0 **:** μ=μ_0

მნიშვნელოვნობის დონე: α

კრიტერიუმის სტატისტიკა (ტესტის სტატისტიკა): $Z=\frac{\bar{X}_n-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$

კრიტერიუმის სტატისტიკის დაკვირვებული მნიშვნელობა: $z=\frac{\bar{x}_n-\mu_0}{\sigma/\sqrt{n}}$

ალტერნატივა	კრიტიკული არე U_1 (H_0-ის უარყოფის არე)
H_1 : μ=μ_1 > μ_0	$z > z_{\alpha}$,
H_1 : μ=μ_1 < μ_0	$z < -z_{\alpha}$

სადაც z_{α} არის $N(0,1)$ განაწილების ზედა α-კრიტიკული წერტილი.

გადაწყვეტილება: თუ $z=\frac{\bar{x}_n-\mu_0}{\sigma/\sqrt{n}}\in U_1$, H_0 **ჰიპოთეზას უარვყოფთ, წინააღმდეგ შემთხვევაში H_0-ის უარყოფის საფუძველი არ გვაქვს**

მაგალითი 6. ამოღებულია n=16 მოცულობის შერჩევა ნორმალურად განაწილებული გენერალური ერთობლიობიდან და შერჩევის საფუძველზე გამოთვლილია შერჩევითი საშუალო $\bar{x}=17.8$. საშუალო კვადრატული გადახრა ცნობილია, σ=7. α=0,05 მნიშვნელოვნობის დონით უნდა შემოწმდეს ჰიპოთეზები

H_0 : μ=19,

H_1 : μ=15.

ამოხსნა.

I . კრიტერიუმის სტატისტიკის შერჩევა

კრიტერიუმის სტატისტიკად ავირჩიოთ ნორმირებული შერჩევითი საშუალო

$$Z_n=\frac{\bar{X}_n-\mu_0}{\sigma_0/\sqrt{n}},$$

II . მნიშვნელოვნობის α დონის მითითება და U_1 კრიტიკული არის დადგენა

ალტერნატივის ძალით საქმე გვაქვს მარცხენა ცალმხრივ კრიტერიუმთან, α=0.05, $z_{0.05}$ = 1.65, U_1=(-∞, -1.65)

III. კრიტერიუმის სტატისტიკის დაკვირვებული მნიშვნელობის გამოთვლა.

$$z_n = \frac{\bar{x}_n - \mu_0}{\sigma_0/\sqrt{n}} = \frac{17.8-19}{7/4} = -0.68.$$

IV. გადაწყვეტილების მიღება
ვინაიდან

$$z_n = -0.68 > -1.65$$

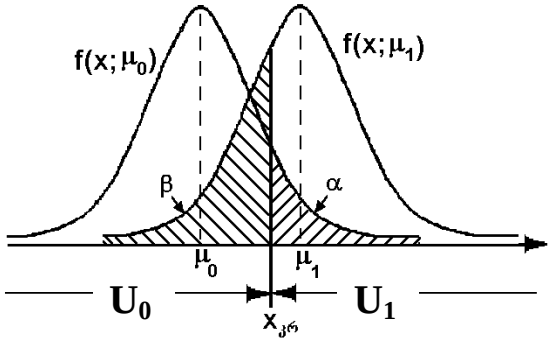
0.05 მნიშვნელოვნობის დონით H_0 ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.
V. გამოვთვალოთ P-მნიშვნელობა:

$$P = P\{Z \le -0.68\} = \Phi(-0.68) = 0.25.$$

ვინაიდან P=0.25>0.05=α, ამიტომ 0.05 მნიშვნელოვნობის დონით H_0 ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს. ♦

§ 5. 1. სიმძლავრე.

ის რომ **ჰიპოთეზები მარტივია, საშუალებას გვაძლევს გამოვითვალოთ მეორე გვარის შეცდომის ალბათობა, როდესაც პირველი გვარის შეცდომის ალბათობა ცნობილია:**



ნახ. 9.1

9.1. ნახაზზე მოცემულია ტესტის სტატისტიკის სიმკვრივის გრაფიკები μ პარამეტრის ჰიპოტეტური მნიშვნელობებისათვის. როგორც ადრე აღვნიშნეთ, პირველი გვარის შეცდომის ალბათობას α ასოთი აღნიშნავენ. რაც შეეხება **მეორე გვარის შეცდომის ალბათობას, მას β ასოთი აღნიშნავენ** და

$$\alpha \equiv P\{\bar{X}_n \in U_1 | \text{სამართლიანია } H_0\} \equiv P\{\bar{X}_n \in U_1 | H_0\}, \quad (9.15)$$

$$\beta \equiv P\{\bar{X}_n \notin U_1 | \text{სამართლიანია } H_1\} \equiv P\{\bar{X}_n \notin U_1 | H_1\}. \quad (9.16)$$

H_0 ჰიპოთეზის უარყოფის ალბათობას, როცა სამართლიანია H_1 ალტერნატივა, ანუ

$$P\{\bar{X}_n \in U_1 | H_1\} \quad (9.17)$$

სიდიდეს კრიტერიუმის სიმძლავრეს უწოდებენ.

შევნიშნოთ, რომ (9.15) და (9.17) გამოსახულებებში ლაპარაკია ერთი და იმავე ხდომილობის, $\bar{X}_n$ შემთხვევითი სიდიდის კრიტიკულ არეში მოხვედრის ალბათობაზე, მხოლოდ პირველ შემთხვევაში ნულოვანი, ხოლო მეორე შემთხვევაში – ალტერნატიული ჰიპოთეზის

სამართლიანობის დროს; პირველი გვარის შეცდომის ალბათობა განსაზღვრავს არასწორი გადაწყვეტილების მიღების ალბათობას იმ შემთხვევაში, როცა სამართლიანია ნულოვანი ჰიპოთეზა, ხოლო კრიტერიუმის სიმძლავრე განსაზღვრავს სწორი გადაწყვეტილების მიღების ალბათობას იმ შემთხვევაში, როცა სამართლიანია H_1 ჰიპოთეზა .

დავუშვათ, მარტივი H_0 და H_1 ჰიპოთეზების გარჩევისას მნიშვნელოვნობის α=0.05 დონისათვის აიგო კრიტიკული არე და გამოთვლილ იქნა კრიტერიუმის სიმძლავრე, რომელიც, ვთქვათ, 0.65-ის ტოლია. ეს იმას ნიშნავს, რომ H_1 ალტერნატივის სამართლიანობის დროს ასიდან საშუალოდ 65 შემთხვევაში მივიღებთ სწორ გადაწყვეტილებას (უარვყოფთ H_0-ს), ხოლო 100-დან საშუალოდ 35 შემთხვევაში ვიღებთ არასწორ გადაწყვეტილებას (არ უარვყოფთ H_0-ს).

<u>მარჯვენა ცალმხრივი კრიტერიუმისათვის და ტესტის $\overline{X}_n$ სტატისტიკისათვის, კრიტერიუმის სიმძლავრეა</u>

$$P\{\overline{X}_n > x_{\text{კრ}} | H_1\} = 1 - \Phi(z_\alpha - \sqrt{n}\,(\mu_1 - \mu_0)/\sigma_0) = 1 - \Phi\left(z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right) \qquad (9.18)$$

შენიშვნა. სიმძლავრის იგივე მნიშვნელობას მივიღებთ, თუ კრიტერიუმის სტატისტიკად ავირჩევთ Z_n-ს.

$$P\{Z_n > z_\alpha \mid H_1\} = P\left\{\frac{\overline{X}_n - \mu_0}{\sigma_0/\sqrt{n}} > z_\alpha \mid H_1\right\} = P\left\{\frac{\overline{X}_n - \mu_0 - (\mu_1 - \mu_0)}{\sigma_0/\sqrt{n}} > z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right\} =$$

$$= 1 - \Phi\left(z_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right). \qquad (9.19)$$

როგორც მიღებული გამოსახულებებიდან ჩანს, **კრიტერიუმის სიმძლავრე არსებითადაა დამოკიდებული ალტერნატიულ ჰიპოთეზაზე**. თუ გავიხსენებთ, რომ $\Phi(x)$ ფუნქცია ზრდადია, ადვილად დავრწმუნდებით, რომ **კრიტერიუმის სიმძლავრე μ_1 არგუმენტის ზრდადი ფუნქციაა:** რაც უფრო დიდია სხვაობა μ_1-სა და μ_0-ს შორის, მით უფრო დიდია კრიტერიუმის სიმძლავრე, ანუ H_1-ის მიღების ალბათობა H_1-ის სამართლიანობის პირობებში იზრდება μ_1-ის ზრდასთან ერთად, რაც იმას ნიშნავს, რომ H_1-ის სამართლიანობის დროს $x_1,x_2,...,x_n$ მონაცემების საფუძველზე **არასწორი გადაწყვეტილების მიღების შანსი მცირდება.**

მაგალითი 7. გამოვთვალოთ კრიტერიუმის სიმძლავრე **მაგალითი 6**-სათვის.

ამოხსნა. მარცხენა ცალმხრივი კრიტერიუმისათვის (H_1: $\mu_1 < \mu_0$) კრიტერიუმის სიმძლავრე გამოითვლება შემდეგი ფორმულით

$$P\{Z_n < -z_\alpha | H_1\} = \Phi(-z_\alpha - \sqrt{n}\,(\mu_1 - \mu_0)/\sigma_0) = \Phi(-1{,}64 - 16/7) = \Phi(0{,}64) = 0{,}73. \quad \blacklozenge$$

ექსპერიმენტის დაგეგმვისას ექსპერიმენტატორისათვის მნიშვნელოვანია ერთის მხრივ, **პირველი და მეორე გვარის შეცდომების ალბათობათა სიმცირის მიღწევა** და მეორეს მხრივ, დანახარჯების გამო, **ცდების მინიმალური რაოდენობის ჩატარება**. განვიხილოთ <u>შემდეგი ამოცანა</u>: **განვსაზღვროთ, რისი ტოლი უნდა იყოს ცდების ის მინიმალური რაოდენობა, *n**, რომლისთვისაც პირველი გვარის შეცდომის ალბათობა α-ს ტოლია, ხოლო მეორე გვარის შეცდომის ალბათობა ნაკლებია β-ზე, სადაც α და β წინასწარ მოცემული სიდიდეებია.**

ამოცანა ამოვხსნათ მარჯვენა ცალმხრივი კრიტერიუმისათვის. დავაფიქსიროთ α, $0<\alpha<1$. შერჩევის ფიქსირებული მოცულობისათვის **მოვძებნოთ $x_{კრ}=x_{კრ}(\alpha, n)$** ისე, რომ პირველი გვარის შეცდომა α-ს ტოლი გამოვიდეს. გვექნება:

$$x_{კრ}= \mu_0 + z_\alpha \sigma / \sqrt{n} . \qquad (9.20)$$

მართლაც

$$P\left\{\overline{X}_n \geq x_{კრ}\right\}= P\left\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq z_\alpha\right\} = \alpha .$$

ახლა ***n* ისე შევარჩიოთ, რომ მეორე გვარის შეცდომის $P\{\overline{X}_n < x_{კრ}|H_1\}$ ალბათობა არ აღემატებოდეს β–ს** ანუ ადგილი ჰქონდეს შემდეგ უტოლობას

$$\beta \geq P\{\overline{X}_n < x_{კრ}|H_1\}= \Phi\left(\frac{x_{კრ}(\alpha,n)-\mu_1}{\sigma/\sqrt{n}}\right).$$

$\Phi(x)$ ფუნქციის ზრდადობის გამო

$$\frac{x_{კრ}-\mu_1}{\sigma/\sqrt{n}} \leq x_\beta .$$

სადაც **x_β სტანდარტული ნორმალური განაწილების β-კვანტილია**, ე.ი. $x_\beta = -z_\beta$. უკანასკნელ უტოლობაში შევიტანოთ $x_{კრ}$-ს მნიშვნელობა (9.16)-იდან და ამოვხსნათ ეს უტოლობა n-ის მიმართ. გვექნება

$$n^* \geq \sigma^2(z_\beta+z_\alpha)^2/(\mu_1-\mu_0)^2 \qquad . \qquad (9.21)$$

ამგვარად, შერჩევის მინიმალური რაოდენობა n^*, რომლისთვისაც პირველი გვარის შეცდომის ალბათობა α-ს ტოლია, ხოლო მეორე გვარის შეცდომის ალბათობა ნაკლებია β-ზე მოიცემა შემდეგი **ფორმულით**

$$n^* =[\sigma^2(z_\beta+z_\alpha)^2/(\mu_1-\mu_0)^2]+1 \quad , \qquad (9.22)$$

სადაც $[a]$ აღნიშნავს a რიცხვის მთელ ნაწილს. მაშასადამე, შერჩევის მინიმალური მოცულობა არის ის მინიმალური ნატურალური რიცხვი, რომლისათვისაც სრულდება (9.17). ვინაიდან $z_{1-\alpha} = -z_\alpha$, (9.18) განსაზღვრავს შერჩევის მინიმალურ მოცულობას როგორც მარჯვენა, ისე მარცხენა ცალმხრივი კრიტერიუმისათვის.

მაგალითი 8. შევაფასოთ ცდების რა მინიმალური რაოდენობაა საჭირო იმისათვის, რომ მიღწეული იქნას $\alpha=0.05$, $\beta=0.1$ მნიშვნელობები ნორმალური განაწილების საშუალოს შესახებ შემდეგი ჰიპოთეზების შემოწმებისას: $H_0 : \mu_0=8$, $H_1 : \mu_1=11$, σ^2 ცნობილია, $\sigma=7$.

ამოხსნა. ნორმალური განაწილების ცხრილებიდან ვიპოვოთ $z_{0.1}$ და $z_{0.05}$. გვექნება $z_\beta= 1.29$, $z_\alpha= 1.64$. (9.18)-ის ძალით, მივიღებთ $n>47$, მაშასადამე, $n^*=48$.

ამით ჩვენ დავასრულეთ ნორმალური პოპულაციის საშუალოსათვის მარტივი ჰიპოთეზების გარჩევის ამოცანის მიმოხილვა. ♦

მიღებულ ზოგიერთ შედეგს თავი მოვუყაროთ შემდეგ ცხრილში

მოდელი	ჰიპოთეზები	კრიტერიუმის სტატისტიკა	α მნიშვნელოვნობის დონით უარყოფთ H_0-ს, თუ
$X \in Bi(n,p)$	$H_0 : p = p_0$ $H_1 : p \neq p_0$,	$\frac{\hat{P}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{AS}{\sim} N(0,1)$	$\lvert z \rvert = \left\lvert \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right\rvert \geq z_{\alpha/2}$
$X \in N(\mu,\sigma_0)$	$H_0 : \mu=\mu_0$ $H_1 :$ $\mu=\mu_1 > \mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma_0/\sqrt{n}} \sim N(0,1)$	$z = \frac{\bar{x}_n - \mu_0}{\sigma_0/\sqrt{n}} \geq z_{\alpha}$
$X \in N(\mu,\sigma_0)$	$H_0 : \mu=\mu_0$ $H_1 : \mu\neq\mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma_o/\sqrt{n}} \sim N(0,1)$	$\lvert z \rvert = \left\lvert \frac{\bar{x}_n - \mu_0}{\sigma_0/\sqrt{n}} \right\rvert \geq z_{\alpha/2}$
$X \in N(\mu,\sigma)$	$H_0 : \mu=\mu_0$ $H_1 : \mu\neq\mu_0$	$T = \frac{\bar{X} - \mu}{S'} \cdot \sqrt{n} \sim t(n-1)$	$\lvert t \rvert = \left\lvert \frac{\bar{x}_n - \mu_0}{S'/\sqrt{n}} \right\rvert \geq t_{n-1,\,\alpha/2}$
$X \in N(\mu,\sigma)$	$H_0 : \sigma^2=\sigma_0^2$ $H_1 : \sigma^2\neq\sigma_0^2$	$\frac{(n-1)S'^2}{\sigma_0^2} \sim \chi^2(n-1)$	$\frac{(n-1)s'^2}{\sigma_0^2} \geq \chi^2_{n-1,\alpha/2}$, ან $\frac{(n-1)s'^2}{\sigma_0^2} \leq \chi^2_{n-1,1-\alpha/2}$

კითხვები თვითშემოწმებისათვის

1. რას ეწოდება სტატისტიკური ჰიპოთეზა?
2. პარამეტრული ჰიპოთეზის შემოწმების ნულოვანი და ალტერნატიული ნაწილები
3. მარჯვენა ცალმხრივი, მარცხენა ცალმხრივი და ორმხრივი ალტერნატივები
4 ჰიპოთეზის შემოწმების წესი—სტატისტიკური კრიტერიუმი
5 ტესტის (კრიტერიუმის) სტატისტიკა
6 კრიტიკული არე და ნულოვანი ჰიპოთეზის მიღების დასაშვები არე
7 პირველი და მეორე გვარის შეცდომები
8 პირველი გვარის შეცდომის ალბათობა და მნიშვნელოვნების დონე
9 ჰიპოთეზის შემოწმება $X \in N(\mu,\sigma_0)$ მოდელის პოპულაციის საშუალოსთვის
10. ჩამოაყალიბე $X \in N(\mu,\sigma_0)$ მოდელში ჰიპოთეზის გარჩევის ალგორითმი პოპულაციის საშუალოსთვის
11. ჩამოაყალიბე $X \in N(\mu,\sigma_0)$ მოდელში ჰიპოთეზის გარჩევის კრიტერიუმი პოპულაციის საშუალოსთვის
12. მარჯვენა ცალმხრივი, მარჯვენა ცალმხრივი და ორმხრივი კრიტიკული არეები $X \in N(\mu,\sigma_0)$ მოდელში პოპულაციის საშუალოსთვის

13. ჰიპოთეზის შემოწმება $X \in N(\mu,\sigma)$ მოდელის პოპულაციის საშუალოსთვის
14. ჩამოაყალიბე $X \in N(\mu,\sigma)$ მოდელში ჰიპოთეზის გარჩევის ალგორითმი პოპულაციის საშუალოსთვის
15. ჩამოაყალიბე $X \in N(\mu,\sigma)$ მოდელში ჰიპოთეზის გარჩევის კრიტერიუმი პოპულაციის საშუალოსთვის
16. მარჯვენა ცალმხრივი, მარცხენა ცალმხრივი და ორმხრივი კრიტიკული არეები $X \in N(\mu,\sigma)$ მოდელში პოპულაციის საშუალოსთვის
17. ჰიპოთეზის შემოწმება $X \in N(\mu,\sigma)$ მოდელის პოპულაციის დისპერსიისთვის
18. ჩამოაყალიბე $X \in N(\mu,\sigma)$ მოდელში ჰიპოთეზის გარჩევის ალგორითმი პოპულაციის დისპერსიისთვის
19. ჩამოაყალიბე $X \in N(\mu,\sigma)$ მოდელში ჰიპოთეზის გარჩევის კრიტერიუმი პოპულაციის დისპერსიისთვის
20. მარჯვენა ცალმხრივი, მარცხენა ცალმხრივი და ორმხრივი კრიტიკული არეები $X \in N(\mu,\sigma)$ მოდელში პოპულაციის დისპერსიისთვის
21. ჰიპოთეზის შემოწმება ბერნულის სქემაში პოპულაციის უცნობი წარმატების ალბათობისთვის
22. ჩამოაყალიბე ბერნულის სქემაში ჰიპოთეზის გარჩევის ალგორითმი პოპულაციის უცნობი წარმატების ალბათობისთვის
23. ჩამოაყალიბე ბერნულის სქემაში ჰიპოთეზის გარჩევის კრიტერიუმი პოპულაციის უცნობი წარმატების ალბათობისთვის
24. მარჯვენა ცალმხრივი, მარცხენა ცალმხრივი და ორმხრივი კრიტიკული არეები ბერნულის სქემაში პოპულაციის უცნობი წარმატების ალბათობისთვის
25. კავშირი ჰიპოთეზათა შემოწმებისა და ინტერვალური შეფასების ამოცანებს შორის
26. ჩამოაყალიბე მარტივი ჰიპოთეზის შემოწმება $X \in N(\mu,\sigma_0)$ მოდელის პოპულაციის საშუალოსთვის
27. ჩამოაყალიბე $X \in N(\mu,\sigma_0)$ მოდელში მარტივი ჰიპოთეზის გარჩევის ალგორითმი პოპულაციის საშუალოსთვის
28. სიმძლავრის გამოთვლა $X \in N(\mu,\sigma_0)$ მოდელში მარტივი ჰიპოთეზის გარჩევის ამოცანაში
29. გამოთვალე შერჩევის მინიმალური რაოდენობა n^*, რომლისთვისაც პირველი გვარის შეცდომის ალბათობა α-ს ტოლია, ხოლო მეორე გვარის შეცდომის ალბათობა ნაკლებია β-ზე

§ 6. EXCEL

განვიხილოთ **Excel**-ის სტატისტიკური ფუნქცია

ZTEST,

რომლის ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა მოცემულია ცხრილში

აღნიშვნა	შინაარსი

Array	არე (დაკვირვებულ მონაცემთა)
x	საშუალო (პოპულაციის)
Sigma	სტანდარტული გადახრა (პოპულაციის)

ZTEST ფუნქციის ჩაწერის სინტაქსი ასეთია:

ZTEST (Array, x, Sigma).

ZTEST - გამოიანგარიშება Z_n სტატისტიკის დაკვირვებული $z_n = \frac{\bar{x}_n - \mu_0}{\sigma_o / \sqrt{n}}$ მნიშვნელობისათვის

P-მნიშვნელობა, ანუ $P\left\{Z \geq \frac{\bar{x}_n - \mu_0}{\sigma_o / \sqrt{n}}\right\}$ ალბათობა.

მაგალითი.

ნორმალური პოპულაციიდან, რომლის საშუალო კვადრატული გადახრა, σ=5, მიღებული

8.5 3.61 11.22 16.38 15.99 18.67 -0.92 8.83 15.48 4.57 6.55 1.55

შერჩევის საფუძველზე შესამოწმებელია შემდეგი ჰიპოთეზები:

H_0 : μ=10,

H_1 : μ>10

რისი ტოლია P-მნიშვნელობა?

ამოხსნა.

ამ შემთხვევაში **ZTEST** ფუნქცია ასე ჩაიწერება:

ZTEST (Array, x, Sigma)=ZTEST (D13:O13,10,5).

ამოხსნის შედეგად მივიღებთ, რომ P-მნიშვნელობა = 0,035.

სავარჯიშოები

9.1. სამრეცხაოს მფლობელს სურს გაყიდოს თავისი საწარმო. იგი პოტენციურ მყიდველს არწმუნებს, რომ საწარმოს დღიური შემოსავალი უკანასკნელი ხუთი წლის მანძილზე შეადგენდა \$675-ს, სტანდარტული გადახრით \$75. შემთხვევით არჩეული 30 დღის შემოსავალთა საშუალო სიდიდემ შეადგინა \$625. α=0.01 მნიშვნელოვნობის დონით აქვს თუ არა საფუძველი მყიდველს ამტკიცოს, რომ საწარმოს მფლობელი არ არის სწორი?
ა) შეადგინეთ მოდელი, რომელიც შეესაბამება ამოცანის პირობას და ჩამოაყალიბეთ ჰიპოთეზები.
ბ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი .
გ) რა გადაწყვეტილებას მიიღებთ?.
დ) განსაზღვრეთ P-მნიშვნელობა.

9.2. საბურავების ქარხნის მონაცემებით საბურავმა უნდა გაიაროს საშუალოდ 60 000 მილი, სტანდარტული გადახრით 5 000 მილი. სატრანსპორტო კომპანიამ გამოსცადა 48 საბურავი. აღმოჩნდა, რომ საბურავებმა გაიარეს საშუალოდ 58 380 მილი. გამოიყენეთ მნიშვნელოვნობის დონე 0.02 და შეასრულეთ შემდეგი დავალება:
ა) შეადგინეთ მოდელი, რომელიც შეესაბამება ამოცანის პირობას და ჩამოაყალიბეთ ჰიპოთეზები.
ბ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი .
გ) რა გადაწყვეტილებას მიიღებთ?.
დ) განსაზღვრეთ P-მნიშვნელობა.

9.3. რესტორანის მმართველი აცხადებს, რომ კლიენტთა მომსახურების საშუალო დროა 3 წუთი, სტანდარტული გადახრით 1 წუთი. შეამოწმეს 50 კლიენტის მომსახურების დრო. იგი აღმოჩნდა, საშუალოდ 2.75 წუთის ტოლი. მნიშვნელოვნობის დონით 0.05 შეგიძლიათ თუ არა დაასკვნათ, რომ მომსახურების დრო 3 წუთზე ნაკლებია?
ა) შეადგინეთ მოდელი, რომელიც შეესაბამება ამოცანის პირობას და ჩამოაყალიბეთ ჰიპოთეზები.
ბ) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?.
გ) განსაზღვრეთ P-მნიშვნელობა.
დ) ააგეთ 95%-იანი ნდობის ინტერვალი კლიენტთა მომსახურების საშუალო დროისთვის.

9.4. ნორმალურად განაწილებული გენერალური ერთობლიობისათვის ამოკრეფილია n=16-ის ტოლი მოცულობის შერჩევა და მის საფუძველზე გამოთვლილი შერჩევითი საშუალო და შესწორებული შერჩევითი დისპერსია, შესაბამისად, 13.9-ისა და 3.6-ის ტოლია.
ა) α=0.05 მნიშვნელოვნობის დონისათვის შეამოწმეთ H_0 : μ=13 ნულოვანი ჰიპოთეზა H_1 : $\mu\neq13$ ალტერნატივის მიმართ,
ბ) შეიცვლება თუ არა პასუხი, თუ ალტერნატიული ჰიპოთეზა იქნება H_1 : $\mu>13$.

9.5. ახალი ელექტრონული ხელსაწყოს შემოწმების მიზნით, კომპანიამ გადაწყვიტა გამოუშვას მხოლოდ ექვსი ასეთი ხელსაწყო. ყოველი მათგანისათვის გაიზომა დრო ხელსაწყოს მწყობრიდან გამოსვლამდე, რის შედეგადაც მიღებული იქნა შემდეგი შერჩევა: 59.2, 68.3, 57.8, 56.5, 63.7 და 57.3 საათი. გვაძლევს თუ არა ეს შერჩევა საკმარის საფუძველს α=0.05 ნდობის დონით შემდეგი დასკვნის გასაკეთებლად: ახალი ხელსაწყოს სიცოცხლის საშუალო ხანგრძლივობა აღემატება 55 საათს?
ა) შეადგინეთ მოდელი, რომელიც შეესაბამება ამოცანის პირობას და ჩამოაყალიბეთ ჰიპოთეზები.
ბ) გამოვთვალოთ შერჩევითი საშუალო და საშუალო კვადრატული გადახრა.
გ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი
დ) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ.
ე) ააგეთ 95%-იანი ნდობის ინტერვალი ხელსაწყოს სიცოცხლის საშუალო ხანგრძლივობისათვის.

9.6. „სწრაფი მომსახურება“ ამტკიცებს, რომ მისი პერსონალი ავტომობილის ზეთის გამოცვლას, ზეთის ფილტრის გამოცვლას და დაზეთვას ასწრებს 15 წუთში. შეამოწმეს 21 მანქანის მომსახურების დრო. მომსახურების დრო საშუალოდ აღმოჩნდა 17 წუთი, შერჩევის სტანდარტული გადახრა 5 წუთის ტოლია. მნიშვნელოვნობის დონით α=0.05, შეამოწმეთ „სწრაფი მომსახურების“ მტკიცების სამართლიანობა.
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და ჩამოაყალიბეთ ჰიპოთეზები..
ბ) ალტერნატივიდან გამომდინარე აირჩიეთ კრიტერიუმი.
გ) გამოთვალეთ კრიტერიუმის მნიშვნელობა და განსაზღვრეთ კრიტიკული წერტილი
დ) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?

9.7. ნორმალურად განაწილებული გენერალური ერთობლიობისათვის მიღებულია 26 მოცულობის შერჩევა. შერჩევის საფუძველზე გამოთვლილი შერჩევითი საშუალო და შესწორებული შერჩევითი დისპერსია 10,1 და 7,5-ის ტოლია. 0.05 მნიშვნელოვნების დონით შეამოწმეთ H_0:μ=9 ნულოვანი ჰიპოთეზა H_1:$\mu\neq9$ ალტერნატივის მიმართ:
ა) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?
ბ) შეიცვლება თუ არა პასუხი, თუ ალტერნატიულ ჰიპოთეზად განვიხილავთ H_1:$\mu>9$

9.8. მოტოციკლების მწარმოებელი აცხადებს, რომ გრძელ გზაზე მოტოციკლი დახარჯავს სამ ლიტრ საწვავს 78 კილომეტრზე. ჩაატარეს ექსპერიმენტი. მოტოციკლმა სამი ლიტრა საწვავით გაიარა 88, 82, 81, 87, 75, 79, 84, 76, 80, 78 კილომეტრი. გამოიყენეთ

მნიშვნელოვნობის დონე 0.01 და გააკეთეთ დასკვნა, არის, თუ არა სამი ლიტრა საწვავით გავლილი საშუალო მანძილი უფრო მეტი ვიდრე გამოცხადებული 78 კილომეტრი.
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და განსაზღვრეთ ჰიპოთეზები.
ბ) გამოთვალეთ შერჩევითი საშუალო და საშუალო კვადრატული გადახრა.
გ) გამოთვალეთ კრიტერიუმის დაკვირვებული მნიშვნელობა და განსაზღვრეთ კრიტიკული წერტილი.
დ) რა გადაწყვეტილებას მიიღებთ?
ე) 0.9 ნდობის ალბათობით ააგეთ ნდობის ინტერვალი პოპულაციის საშუალოსათვის.

9.9. გაზეთის ცნობით წლიური საპროცენტო განაკვეთი დიდ ბანკებში აღემატება 9%-ს. 8 მცირე მანკიდან აღებულმა შერჩევამ აჩვენა შემდეგი საპროცენტო განაკვეთები 10.1, 9.3, 9.2, 10.2, 9.3, 9.6, 9.4, 8.8. შეგიძლიათ თუ არა, 0.01 მნიშვნელოვნობის დონით, დაასკვნათ, რომ მცირე ბანკების საპროცენტო განაკვეთი აღემატება 9%-ს?
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და განსაზღვრეთ ჰიპოთეზები.
ბ) ჩაატარეთ აუცილებელი გამოთვლები. როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?.
გ) 0.9 ნდობის ალბათობით ააგეთ ნდობის ინტერვალი ის საპროცენტო განაკვეთისათვის.

9.10. ნორმალურად განაწილებული გენერალური ერთობლიობისათვის მიღებულია n=16-ის ტოლი შერჩევა და მის საფუძველზე გამოთვლილია შესწორებული შერჩევითი დისპერსია, რომლებიც 1.5-ის ტოლია. 0.1 მნიშვნელოვნების დონისათვის შეამოწმეთ H_0:$\sigma^2=10$ ნულოვანი ჰიპოთეზა H_1:$\sigma^2 \neq 10$ ალტერნატივის მიმართ:
ა) ჩაატარეთ აუცილებელი გამოთვლები. როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?.
ბ) შეიცვლება თუ არა პასუხი, თუ ალტერნატიულ ჰიპოთეზად განვიხილავთ H_1:$\sigma^2>10$.

9.11. ნორმალურად განაწილებული გენერალური ერთობლიობიდან ამოკრეფილია n=23-ის ტოლი მოცულობის შერჩევა და მის საფუძველზე გამოთვლილია შესწორებული შერჩევითი დისპერსია, $s'^2=5{,}22$. α=0.05 მნიშვნელოვნობის დონისათვის შეამოწმეთ H_0 : $\sigma^2=10$ ნულოვანი ჰიპოთეზა H_1 : $\sigma^2<10$ ალტერნატივის მიმარ.

9.12. ნორმალურად განაწილებული გენერალური ერთობლიობიდან ამოკრეფილია n=16-ის ტოლი მოცულობის შერჩევა და მის საფუძველზე გამოთვლილია შესწორებული შერჩევითი დისპერსია, $s'^2 = 26{,}2$.
ა) α=0.05 მნიშვნელოვნობის დონისათვის შეამოწმეთ H_0 : $\sigma^2=15$ ნულოვანი ჰიპოთეზა H_1 : $\sigma^2\neq15$ ალტერნატივის მიმარ.
ბ) შეიცვლება თუ არა პასუხი, თუ ალტერნატიული ჰიპოთეზა იქნება H_1 : $\sigma^2>15$.
გ) 0.95-ის ნდობის ალბათობით ააგეთ ნდობის ინტერვალი პოპულაციის დისპერსიისათვის .

9.13. გაზეთის ცნობით კოლეჯის დამთავრებულთაგან 1/3-ს ელის სამუშაო. ერთი სკოლის 200 დამთავრებულისაგან 80-მა მიიღო სამუშაო. მნიშვნელოვნობის დონით 0.02, შეიძლება თუ არა დაასკვნათ, რომ ამ კოლეჯის დამთავრებულები ღებულობენ სამუშაოს უფრო მეტი პროპორციით ვიდრე 1/3-ია?
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და ჩამოაყალიბეთ ჰიპოთეზები
ბ) რა სტატისტიკა გამოიყენება ამ შემთხვევაში, როგორია მისი განაწილების კანონი.
გ) გამოთვალეთ კრიტერიუმის დაკვირვებული მნიშვნელობა და განსაზღვრეთ კრიტიკული წერტილი.
დ)როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?

9.14. ჟურნალში გამოქვეყნებული ცნობის მიხედვით ამერიკელ მამაკაცს კვირაში აქვს 40 თავისუფალი საათი. განხორციელდა ამ ცნობის დამოუკიდებელი შემოწმება. აღმოჩნდა, რომ შემთხვევით აღებულ 60 მამაკაცს საშუალოდ კვირაში 37.8 თავისუფალი საათი ჰქონდა. შერჩევის სტანდარტული გადახრა იყო 12.2 საათი. შეგიძლიათ თუ არა დაასკვნათ, რომ ჟურნალის ცნობა არ არის სწორი? გამოიყენეთ მნიშვნელოვნობის დონე α=0.05.
ა) რა სტატისტიკა გამოიყენება ამ შემთხვევაში, როგორია მისი განაწილების კანონი.
ბ) ჩაატარეთ აუცილებელი გამოთვლები. როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?.

9.15. არის ცნობა, რომ ავტომანქანის ძრავის ნაპერწკლიანი სანთლის მუშაობის ხანგრძლივობა არის 22100 მილი. მწარმოებელი აცხადებს, რომ სანთლის მუშაობის საშუალო ხანგრძლივობა მეტია 22100 მილზე. 18 ეგზემპლარის შესწავლამ აჩვენა, რომ მუშაობის ხანგრძლივობა არის საშუალოდ 23400 მილი, შერჩევითი სტანდარტული გადახრით 3000 მილი. მნიშვნელოვნობის დონით α=0.05 შეამოწმეთ მწარმოებელის აზრის სისწორე.
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და განსაზღვრეთ ჰიპოთეზები.
ბ)როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?
გ) ააგეთ 95%-იანი ნდობის ინტერვალი ნაპერწკლიანი სანთლის მუშაობის ხანგრძლივობისათვის.

9.16. ავტომობილის მწარმოებელი ფირმა აცხადებს, რომ ქალაბში ავტომობილი 10 ლიტრი საწვავით გაივლის 125 კილომეტრს. ჩაატარეს ექსპერიმენტი. ავტომობილმა 10 ლიტრა საწვავით გაიარა 125, 117, 130, 131, 126, 113, 115, 119, 116, 128 კილომეტრი. გამოიყენეთ მნიშვნელოვნობის დონე 0.05 და გააკეთეთ დასკვნა, გაივლის, თუ არა ავტომობილი 10 ლიტრა საწვავით საშუალოდ 125 კმ-ზე ნაკლებ მანძილს.
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და განსაზღვრეთ ჰიპოთეზები.
ბ) გამოთვალეთ შერჩევითი საშუალო და საშუალო კვადრატული გადახრა.
გ) გამოთვალეთ კრიტერიუმის დაკვირვებული მნიშვნელობა და განსაზღვრეთ კრიტიკული წერტილი. რა გადაწყვეტილებას მიიღებთ?
დ) 0.9 ნდობის ალბათობით ააგეთ ნდობის ინტერვალი პოპულაციის საშუალოსათვის.

9.17. ხარისხის კონტროლის მენეჯერს სურს გაიგოს, აღემატება თუ არა ახალი დანადგარით დამზადებული დეფექტური პაკეტების წილი 10%–ს. ამოცანის გადასაწყვეტად შერჩეული 200 პაკეტიდან 26 დეფექტური აღმოჩნდა. მნიშვნელოვნობის დონით 0.01, შეიძლება თუ არა დაასკვნათ, რომ ახალი დანადგარით დამზადებული დეფექტური პაკეტების წილი 10% - ზე მეტია ?
ა) დაახასიათეთ ამოცანის პირობით განსაზღვრული შემთხვევითი სიდიდე და ჩამოაყალიბეთ ჰიპოთეზები
ბ) ჩაატარეთ აუცილებელი გამოთვლები. როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?.

9.18. 120 დამოუკიდებელ ცდაში საინტერესო ხდომილობა 25-ჯერ განხორციელდა. 0.05 მნიშვნელოვნების დონისათვის შეამოწმეთ ნულოვანი ჰიპოთეზა H_0: p=0.25 ალტერნატიული ჰიპოთეზის H_1: $p \neq 0.25$ მიმართ.
ა) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?
ბ) ამოხსენით იგივე ამოცანა, როდესაც ალტერნატიულ ჰიპოთეზას აქვს შემდეგისახე: H_1:p>0,25.

9.19. $N(\mu,49)$ კანონით განაწილებული გენერალური ერთობლიობის საშუალოს შესახებ გამოთქმულია ორი ჰიპოთეზა

H_0 : μ_0=25,

H_1 : μ_1=19.

α=0.01 მნიშვნელოვნობის დონით ჰიპოთეზების შესამოწმებლად მიღებულია n=64 მოცულობის შერჩევა, რომლის საფუძველზე გამოთვლილი შერჩევითი საშუალოა, $\bar{x}$ =23,8.
ა) მარტივია თუ რთულია ჰიროთეზები?
ბ) ალტერნატივიდან გამომდინარე აირჩიეთ კრიტერიუმი.
გ) როგორ განისაზღვრება კრიტიკული არე?
დ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი.
ე) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ.
ვ) განსაზღვრეთ P-მნიშვნელობა.

9.20. $N(\mu,36)$ კანონით განაწილებული გენერალური ერთობლიობის საშუალოს შესახებ გამოთქმულია ორი ჰიპოთეზა

H_0 : μ_0=6,5,

H_1 : μ_1=9.

ჰიპოთეზის შესამოწმებლად მიღებულია n=49 მოცულობის შერჩევა, რომლის საფუძველზე გამოთვლილია შერჩევითი საშუალო, $\bar{x}$ =8.1. α=0.01 მნიშვნელოვნობის დონისათვის განსაზღვრეთ, სამართლიანია თუ არა ნულოვანი ჰიპოთეზა.
ა) ალტერნატივიდან გამომდინარე აირჩიეთ კრიტერიუმი.
ბ) როგორ განისაზღვრება კრიტიკული არე?
გ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი.
დ) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ.
ე) განსაზღვრეთ P-მნიშვნელობა.

9.21. ნორმალურად განაწილებული პოპულაციის საშუალოსათვის, რომელიც ხასიათდება 15-ის ტოლი სტანდარტული გადახრით, α=0.05 მნიშვნელოვნობის დონით შესამოწმებელია ჰიპოთეზები:

H_0 : μ_0=50,

H_1 : μ_1=60.

n=64 მოცულობის შერჩევის საფუძველზე გამოთვლილი შერჩევითი საშუალოა, $\bar{x}$ =54. განსაზღვრეთ სამართლიანია თუ არა ნულოვანი ჰიპოთეზა.
ა) როგორ განისაზღვრება კრიტიკული არე?
ბ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი .
გ) როგორია თქვენი გადაწყვეტილება H_0-ის შესახებ?
დ) განსაზღვრეთ P-მნიშვნელობა.

9.22. ბანკის მენეჯერს სურს α=0.05 მნიშვნელოვნობის დონით, n=169 მოცულობის შერჩევითი მნიშვნელობების საფუძველზე შეამოწმოს ჰიპოთეზა იმის შესახებ, არის თუ არა დეპოზიტების საშუალო მნიშვნელობა 105-ის ტოლი. წარსული გამოცდილების თანახმად ის თვლის, რომ დეპოზიტები ნორმალურად განაწილებული შემთხვევითი სიდიდეებია 45-ის ტოლი საშუალო კვადრატული გადახრით.
ა) რა სტატისტიკა გამოიყენება ჰიპოთეზის შემოწმებისათვის?
ბ) განსაზღვრეთ კრიტიკული არე კრიტერიუმის სტატისტიკისათვის.

9.23. ნორმალურად განაწილებული გენერალური ერთობლიობიდან, რომლის დისპერსია ცნობილია, σ^2=100 ამოღებულია n=49 მოცულობის შერჩევა და მის საფუძველზე გამოთვლილია შერჩევითი საშუალო, $\bar{x}$ =42. α=0.05 მნიშვნელოვნობის დონით შესამოწმებელია შემდეგი ჰიპოთეზები:

H_0 : μ=37,

H_1 : μ=41.

ა) რომელ სტატისტიკას გამოიყენებთ ამ შემთხვევაში და როგორია მისი განაწილება?
ბ) როგორ განისაზღვრება კრიტიკული არე?
გ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი .
დ) რა გადაწყვეტილებას მიიღებთ?
ე) რისი ტოლია მეორე გვარის შეცდომის ალბათობა.

ვ)ცდების რა მინიმალური რაოდენობაა საჭირო იმისათვის, რომ α=0.05 მნიშვნელოვნობის დონით აღნიშნული ჰიპოთეზების შემოწმებისას მეორე გვარის შეცდომის ალბათობა ნაკლები იყოს 0.1-ზე?

9.24. ნორმალურად განაწილებული გენერალური ერთობლიობიდან, რომლის დისპერსია ცნობილია, σ^2=121 ამოღებულია n=36 მოცულობის შერჩევა და მის საფუძველზე გამოთვლილია შერჩევითი საშუალო, $\bar{x}=52$. α=0.01 მნიშვნელოვნობის დონით შესამოწმებელია შემდეგი ჰიპოთეზები:

H_0 : μ=56,

H_1 : μ=52.

ა) რომელ სტატისტიკას გამოიყენება ამ შემთხვევაში და როგორია მისი განაწილება?
ბ) როგორ განისაზღვრება კრიტიკული არე?
გ) გამოთვალეთ კრიტერიუმის სტატისტიკა და განსაზღვრეთ კრიტიკული წერტილი.
დ) რა გადაწყვეტილებას მიიღებთ?
ე) რისი ტოლია კრიტერიუმის სიძლავრე?

9.25. $N(\mu,100)$ კანონით განაწილებული პოპულაციისათვის, n=49 მოცულობის შერჩევის საფუძველზე შესამოწმებელია შემდეგი ჰიპოთეზები

H_0:$\mu=\mu_0$=10,

H_1:$\mu=\mu_1$=5.

0.01 მნიშვნელოვნების დონისათვის რისი ტოლია კრიტერიუმის სიმძლავრე.

9.26. $N(\mu,25)$კანონით განაწილებული პოპულაციისათვის, n=36 მოცულობის შერჩევის საფუძველზე შესამოწმებელია შემდეგი ჰიპოთეზები

H_0:$\mu=\mu_0$=10,

H_1:$\mu=\mu_1$=12.

ა) 0,05 მნიშვნელოვნების დონისათვის რისი ტოლია კრიტერიუმის სიმძლავრე.
ბ) შერჩევის რა მინიმალური მოცულობაა საჭირო იმისათვის, რომ პირველი და მეორე გვარის შეცდომები არ აღემატებოდეს შესაბამისად 0,05-სა და 0,1-ს.

9.27. შერჩევის რა მინიმალური მოცულობაა საჭირო იმისათვის, რომ პირველი და მეორე გვარის შეცდომები არ აღემატებოდეს შესაბამისად 0.05-სა და 0.1-ს შემდეგი ჰიპოთეზის: ფაკულტეტის სტუდენტების საშუალო რეიტინგი 50-ია, შემოწმების დროს ალტერნატივის წინააღმდეგ, რომ საშუალო რეიტინგი 47-ის ტოლია. ცნობილია რომ სტანდარტული გადახრა 10-ის ტოლია.

9.28. ნორმალური პოპულაციის დისპერსია 7-ის ტოლია. საშუალოს შესახებ არსებობს შემდეგი ჰიპოთეზები

H_0:$\mu=\mu_0$=8,3

H_1:$\mu=\mu_1$=5.

ჰიპოთეზების შემოწმება უნდა განვახორციელოდ n=25 მოცულობის შერჩევის საფუძველზე. α=0.01 მნიშვნელოვნების დონისათვის განსაზღვრეთ რისი ტოლი იქნება კრიტერიუმის სიმძლავრე

9.29. მზა პროდუქციის მახასიათებელი სიდიდე ნორმალურადაა განაწილებული. მენეჯერს სურს n=9 მოცულობის შერჩევის საფუძველზე 0.05 მნიშვნელოვნობის დონით შეამოწმოს შემდეგი ჰიპოთეზები:

H_0:EX=10,

H_1:EX=6.

წარსული გამოცდილების თანახმად, მენეჯერი თვლის, რომ საშუალო კვადრატული გადახრა 8-ის ტოლია. ამ პირობებში რისი ტოლი იქნება მეორე გვარის შეცდომის ალბათობა.

9.30. ნორმალურად განაწილებული პოპულაციისათვის, რომელიც ხასიათდება 4-ის ტოლი სტანდარტული გადახრით, α=0.05 მნიშვნელოვნობის დონისათვის შესამოწმებელია ჰიპოთეზები:

H_0 : μ=2,

H_1 : μ=3.

იპოვეთ შერჩევითი მოცულობის მნიშვნელობა *n*, რომლისთვისაც კრიტერიუმის სიმძლავრეა 0,6.

9.31. ნორმალურად განაწილებული პოპულაციისათვის, რომელიც ხასიათდება 4-ის ტოლი სტანდარტული გადახრით, α=0.05 მნიშვნელოვნობის დონისათვის შესამოწმებელია ჰიპოთეზები:

H_0 : μ=17,

H_1 : μ=15.

იპოვეთ შერჩევითი მოცულობის მნიშვნელობა *n*, რომლისთვისაც კრიტერიუმის სიმძლავრეა 0,8.

9.32. ნორმალური გენერალური ერთობლიობის მათემატიკური ლოდინის მიმართ არსებობს შემდეგი ჰიპოთეზები

$H_{0:}$ $\mu=\mu_0$

$H_{0:}$ $\mu=\mu_1>\mu_0$ ($\mu=\mu_1<\mu_0$).

n მოცულობის შერჩევისათვის, α მნიშვნელოვნობის დონით, გამოთვლილ იქნა კრიტერიუმის სიმძლავრე, რომელიც 0.5□-ის ტოლი აღმოჩნდა. კრიტერიუმის სიმძლავრე მიიჩნიეს არასაკმარისად. ამ მიზეზით ოთხჯერ ზრდიან შერჩევის მოცულობას. რისი ტოლი იქნება კრიტერიუმის სიმძლავრე იგივე მნიშვნელოვნობის დონისათვის?

ლექცია XII,XIII* არაპარამეტრული ჰიპოთეზათა შემოწმება

§ 1. χ^2-კრიტერიუმი დისკრეტული პოპულაციისთვის

ვთქვათ პოპულაცია აღიწერება **დისკრეტული** X შემთხვევითი სიდიდით, რომლის განაწილების კანონია

X	x_1	x_2	. . .	X_k
	p_1	p_2	. . .	p_k

(12.1)

სადაც

$$p_i=P\{X=x_i\} \quad , i=1,2,...,k,$$

პოპულაცია უცნობი დისკრეტული განაწილებაა.

ახლა დავუშვათ, რომ პოპულაციიდან აღებული n (n გაცილებით მეტია k-ზე) მოცულობის შერჩევის შედეგად რიცხვი x_i შეგვხდა n_i-ჯერ, i=1,2,...,k, სადაც $0\le n_i\le n$ ყოველი i-თვის და $\sum_{i=1}^{k} n_i = n$. **ამ შერჩევის საფუძველზე გვინდა შევამოწმოთ ჰიპოთეზა**

$$H_0 : p_1=p_{01}, p_2=p_{02}, \dots , p_k=p_{0k}$$

$$H_a : p_i\neq p_{0i} \text{ ერთი მაინც } i\text{-სთვის}$$

სადაც $p_{01}, p_{02}, \dots , p_{0k}$ **- კონკრეტული ჰიპოტეტური ალბათური განაწილებაა** (ვგულისხმობთ, რომ სასრული პოპულაციის შემთხვევაში შერჩევა პოპულაციიდან აიღება დაბრუნებით);.

გადაწყვეტილების მიღების წესი მოითხოვს n_i და np_{0i} სიდიდეებს შორის გადახრის ისეთი ზომის შემოღებას, რომლის მიხედვით H_0 ჰიპოთეზის უარყოფა უნდა მოხდეს მაშინ, როდესაც გადახრა საკმარისად დიდია.

H_0 ჰიპოთეზის შემოწმების მიზნით ბუნებრივია ერთმანეთს შევადაროთ დაკვირვებული და მოსალოდნელი სიხშირეები. თუ H_0 ჰიპოთეზა მართებულია, მაშინ np_{0i} მოსალოდნელი და n_i დაკვირვებული სიხშირეები ახლოს უნდა იყოს ერთმანეთთან და პირიქით, როდესაც p_i-ს ჭეშმარიტი მნიშვნელობები საგრძნობლად განსხვავდება ნულოვანი ჰიპოთეზით განსაზღვრული ალბათობებისაგან, მოსალოდნელი და დაკვირვებული სიხშირეებიც მკვეთრად უნდა განსხვავდებოდნენ.

ცხრილი 1. **დაკვირვებული და მოსალოდნელი სიხშირეები**

	i=1	i=2	. . .	$i=k$	ჯამი
დაკვირვებული	n_1	n_2	. . .	n_k	n
მოსალოდნელი	nP_{01}	nP_{02}	. . .	nP_{0k}	n

(12.2)

ბუნებრივია, მოსალოდნელ და დაკვირვებულ სიხშირეებს შორის **განსხვავების საზომად** $(n_1-np_{01})^2, \dots , (n_k-np_{0k})^2$ **კვადრატული გადახრები, ხოლო საერთო ზომის როლში**

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(n_i - np_{0i})^2}{np_{0i}} \qquad (12.3)$$

სტატისტიკა გამოვიყენოთ, რომელიც **მიახლოებით განაწილებული იქნება χ^2 განაწილების კანონით (k–1)-ის ტოლი თავისუფლების ხარისხით.**

მნიშვნელოვანია, რომ (12.3) ფორმულით განსაზღვრული სტატისტიკის ზღვრული განაწილება არ არის დამოკიდებული ჰიპოთეტური ალბათობების კონკრეტულ მნიშვნელობებზე. ჰიპოთეზის დროს (12.3) ყოველთვის (k–1)-ის ტოლი თავისუფლების ხარისხის მქონე χ^2 განაწილებას ემთხვევა.

ეს გარემოება საშუალებას გვაძლევს **ავაგოთ კრიტიკული არე χ^2 განაწილების ცხრილის მიხედვით. განხილული H_0 ჰიპოთეზის შემოწმება ხდება შემდეგნაირად: დავაფიქსიროთ რაიმე მნიშვნელოვნობის დონე α** (მაგალითად ავიღოთ α=0.1; 0.05 ან 0.01) და **მოვნახოთ (k–1) თავისუფლების ხარისხის χ^2 განაწილების შესაბამისი ზედა კრიტიკული მნიშვნელობა (ცხრილის მიხედვით), რომელსაც აღვნიშნავთ $\chi^2_{k-1,\alpha}$ -თი. თუ ადგილი აქვს უტოლობას**

$$\hat{\chi}^2 > \chi^2_{k-1,\alpha}$$

მაშინ H_0 ჰიპოთეზას უარვყოფთ, ხოლო როდესაც სრულდება

$$\hat{\chi}^2 \le \chi^2_{k-1,\alpha}$$

საწინააღმდეგო უტოლობა, ვასკვნით, რომ დაკვირვებული მონაცემები არ ეწინააღმდეგება H_0 ჰიპოთეზას, ანუ არსებული მონაცემები არ იძლევა ნულოვანი ჰიპოთეზის უარყოფის საფუძველს.

მაგალითი 1. დისკრეტული X პოპულაციიდან მიღებული 1000 მოცულობის შერჩევით მნიშვნელობებს შორის შემთხვევითი სიდიდის მნიშვნელობა 0 შეგვხდა 343-ჯერ, მნიშვნელობა 1 - 372-ჯერ, მნიშვნელობა 2- 201-ჯერ, მნიშვნელობა 3 - 68-ჯერ, ხოლო სამზე დიდი მნიშვნელობა 16-ჯერ. α=0,05 მნიშვნელოვნობის დონით შევამოწმოთ H_0 ნულოვანი ჰიპოთეზა იმის შესახებ, რომ შერჩევა მიღებულია λ=1-ის ტოლი ინტენსივობის მქონე პუასონის კანონით განაწილებული გენერალური ერთობლიობიდან .

ნულოვანი ჰიპოთეზის სამართლიანობის შემთხვევაში

$p_{01}=P\{X=0\}=0{,}368$, $p_{02}=P\{X=1\}=0{,}368$, $p_{03}=P\{X=2\}=0{,}184$, $p_{04}=P\{X=3\}=0{,}061$.

$p_{05}=P\{X>3\}=0{,}019$.

სტატისტიკის დაკვირვებული მნიშვნელობის გამოსათვლელად ვისარგებლოდ 2 ცხრილით

ცხრილი 2

i	n_i	np_{0i}	n_i-np_{0i}	$(n_i-np_{0i})^2/np_{0i}$
1	343	368	-25	1.68
2	372	368	4	0.05
3	201	184	17	1.56
4	68	61	7	0.73
5	16	19	-3	0.47
			$\sum$	4.59

ამრიგად, (11.3) ფორმულის საფუძველზე მივიღებთ

$$\hat{\chi}^2 = 4{,}51.$$

χ^2 განაწილების ცხრილიდან ვპოულობთ, რომ ზედა კრიტიკული მნიშვნელობა $\chi^2_{4,0,05} \approx 9{,}49$. ვინაიდან ადგილი აქვს უტოლობას $\hat{\chi}^2 = 4{,}51 < 9{,}49 = \chi^2_{4,0,05}$, H_0 ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს. ♦

მაგალითი 2. XVIII საუკუნეში ბიუფონმა ჩაატარა n=4040 დამოუკიდებელი ცდა მონეტის აგდებაზე, რის შედეგადაც გერბი გამოჩნდა N_1=2048 ცდაში, ხოლო დანარჩენ $N_2 = n - N_1$ შემთხვევაში გამოჩნდა საფასური. ეთანხმება თუ არა ექსპერიმენტის შედეგი ჰიპოთეზას, რომ არსებობს გერბის გამოჩენის მუდმივი P ალბათობა, რომელიც 1/2-ის ტოლია (ანუ, მართებულია თუ არა ჰიპოთეზა მონეტის სიმეტრიულობის შესახებ და გერბისა და საფასურის გამოჩენის ალბათობა ყოველ ცდაში ერთი და იგივეა; α=0.05).

ვიანგარიშოთ $\hat{\chi}^2$ სტატისტიკა (12.3) ფორმულის მიხედვით. (12.3) გამოსახულებაში P_{01}=P_{02}= 1/2, n=4040, N_1=2048 მნიშვნელობების ჩასმით, მივიღებთ

i	n_i	np_{0i}	$n_i - np_{0i}$	$(n_i - np_{0i})^2/np_{0i}$
1	2048	2020	28	0.388
2	1992	2020	-28	0.388
			$\sum$	0.776

ამრიგად,

$$\hat{\chi}^2 = 0.776.$$

რადგან ერთის ტოლი თავისუფლების ხარისხის χ^2 განაწილების α=0.05 მნიშვნელოვნობის დონის კრიტიკული მნიშვნელობა $\chi^2_{1,0.05}$=3.841 აღემატება $\hat{\chi}^2$-სტატისტიკის რეალიზებულ მნიშვნელობას, შეგვიძლია ჩავთვალოთ, რომ ექსპერიმენტის შედეგი არ ეწინააღმდეგება ჰიპოთეზას. შევნიშნოთ, რომ ალბათობა იმისა, რომ აღნიშნული ექსპერიმენტის ჩატარებისას $\hat{\chi}^2$ სტატისტიკა მიიღებს 0.776-ზე მეტ მნიშვნელობას (იმ დაშვებით, რომ P=1/2) დაახლოებით 0.38-ის ტოლია და ასეთი ხდომილობის განხორციელება (ანუ $\hat{\chi}^2$ სტატისტიკის ასეთი გადახრა) სავსებით ბუნებრივად შეიძლება ჩაითვალოს. ♦

§ 2. χ^2-კრიტერიუმი უწყვეტი პოპულაციისთვის

ჩვენ ახლა განვიხილავთ **ჰიპოთეზის შემოწმების ამოცანის ალგორითმს უწყვეტი ერთგანზომილებიანი ჰიპოთეტური $F_0(x)$ განაწილების ფუნქციისათვის.**

ასეთი ჰიპოთეზის შემოწმების მიზნით უწყვეტი X **შემთხვევითი სიდიდის შესაძლო მნიშვნელობათა არე დავყოთ ერთმანეთის გამომრიცხავ $\Delta_1, \Delta_2, \ldots, \Delta_k$ სიმრავლეებად** (ინტერვალებად). Δ_i სიმრავლეებს, როგორც წესი, ირჩევენ მიმდევრობით ინტერვალებად დაკვირვებადი შემთხვევითი სიდიდის მნიშვნელობათა სივრციდან (თუმცა ეს არ არის აუცილებელი). **სიმარტივისათვის, ავიღოთ რიცხვთა ღერძის**

$$\Delta_1 = (-\infty, a_1], \Delta_2 = (a_1, a_2], \ldots, \Delta_{k-1} = (a_{k-2}, a_{k-1}], \Delta_k = (a_{k-1}, \infty)$$

დანაწილება.

რადგან $F_0(x)$ განაწილების ფუნქცია ცნობილია, ჩვენ შეგვიძლია დაკვირვებადი X შემთხვევითი სიდიდის ყოველ Δ_i ინტერვალში მოხვედრის ალბათობის გამოთვლა (ცხადია, იმ დაშვებით, რომ H_0 ჰიპოთეზა მართებულია). თუ p_{0i}-თი აღვნიშნავთ X შემთხვევითი სიდიდის Δ_i ინტერვალში მოხვედრის ალბათობას H_0 ჰიპოთეზის დროს, მივიღებთ, რომ

$$p_{01}=F_0(a_1),\ p_{02}=F_0(a_2)-F_0(a_1),\ \ldots,\ p_{0k}=1-F_0(a_{k-1}). \qquad (12.4)$$

n_i-ით აღვნიშნოთ Δ_i ინტერვალში მოხვედრილ შერჩევის ელემენტთა რაოდენობა. ნულოვანი ჰიპოთეზის სამართლიანობის შემთხვევაში იგივე ინტერვალში მოხვედრილ შერჩევის ელემენტთა რაოდენობა np_{0i} -ის ტოლია.

ამრიგად, ჰიპოთეზის შესამოწმებლად **ვაგებთ ისევ** (12.3) **ფორმულით განსაზღვრულ** $\hat{\chi}^2$ **სტატისტიკას, რომელიც ასიმპტოტურად ასევე განაწილებული იქნება χ^2 კანონით ($k-1$)-ის ტოლი თავისუფლების ხარისხით. ამიტომ ამ ჰიპოთეზის შემოწმება ხდება ჩვენთვის უკვე ცნობილი გზით, წინა პუნქტში განხილული შემთხვევის მსგავსად**.

ისმის კითხვა, როგორ შევარჩიოთ k-ს მნიშვნელობა და დაყოფის a_i საზღვრები? ამ კითხვაზე მკაცრი, ცალსახა პასუხი არ არსებობს. არსებობს თეორიულ გამოკვლევებზე და პრაქტიკაზე დაყრდნობილი რეკომენდაციები.

ჯერ მოვიყვანთ რეკომენდაციებს k-ს შესარჩევად. როცა შერჩევის მოცულობა მცირეა, ეს კრიტერიუმი არ გამოიყენება, რადგან ის არ მოგვცემს სანდო გადაწყვეტილებას. თუ შერჩევის მოცულობა $n>50$, k ისე უნდა შევარჩიოთ, რომ სიხშირის მოსალოდნელი რაოდენობა np_i არ იყოს ნაკლები 5-ზე (უმჯობესია, თუ $np_i \geq 10$) ყოველი ინტერვალისათვის. თუ შერჩევის მოცულობა 100-დან 200-მდეა პრაქტიკაში ხშირად სარგებლობენ k-ს მნიშვნელობებით 12-დან 16-მდე. როდესაც $200 \leq n \leq 400$, k-ს ირჩევენ [16;20] ინტერვალიდან. თუ $n>1000$, k-ს იღებენ 25-ის ტოლად. ხშირად საჭირო ხდება k-ს რამდენიმე მნიშვნელობის მოსინჯვა.

მას შემდეგ რაც k განსაზღვრულია, a_i რიცხვების დასადგენად გამოიყენება, უპირატესად, ორი ხერხი.

1. **ტოლ ალბათობათა ხერხი**. a_i განისაზღვრება ისე, რომ ყოველ ინტერვალში შერჩევითი მნიშვნელობის მოხვედრის ალბათობა $1/k$-ს ტოლი იყოს.

2. **ტოლი ინტერვალების ხერხი**. შესაძლო მნიშვნელობათა სიმრავლე დაიყოფა k რაოდენობა ტოლი სიგრძის მქონე ნაწილად.

ინტერვალის საზღვრების დადგენის **შემდეგ განისაზღვრება ინტერვალში შერჩევითი მნიშვნელობების მოხვედრის ალბათობები** (12.4) **ფორმულებით.**

თუ აღმოჩნდა, რომ რომელიმე ინტერვალში სიხშირის მოსალოდნელი რაოდენობა ნაკლებია 5-ზე, მაშინ ერთი ან რამდენიმე მეზობელი ინტერვალი უნდა გაერთიანდეს ერთ ინტერვალად

მაგალითი 3. ა.ეიტკენმა შეისწავლა საათების შემკეთებელ სახელოსნოებში გამოფენილი საათების ჩვენებები. მოგვყავს მიღებულ დაკვირვებათა ცხრილი, დაყოფილი 12 ჯგუფად, საათის ისრის ციფერბლატზე მდგომარეობის მიხედვით

შუალედები ციფერბლატზე	დაკვირვებული სიხშირეები	შუალედები ციფერბლატზე	დაკვირვებული სიხშირეები
0–1	41	6–7	41
1–2	34	7–8	33
2–3	54	8–9	37
3–4	39	9–10	41
4–5	49	10–11	47
5–6	45	11–12	39

შევამოწმოთ მარტივი H_0 ჰიპოთეზა, რომ საათის ისრის მდგომარეობა თანაბრად არის განაწილებული (0,12) ინტერვალში, რთული H_a ალტერნატივის წინააღმდეგ (H_a: H_0 არ არის მართებული). ამ მაგალითში n=500, P_{0i}=1/12, i=1,2,...,12 და ყოველ ინტერვალში საათის ისრის მოხვედრის მოსალოდნელი რაოდენობა nP_{oi}≈41.67. ამიტომ (12.3) ტოლობით განსაზღვრული $\hat{\chi}^2$ სტატისტიკა დაახლოებით მიიღებს 10-ის ტოლ მნიშვნელობას. რადგან ეს სტატისტიკა დაახლოებით 11-ის ტოლი თავისუფლების ხარისხის მქონე χ^2 კანონით არის განაწილებული, ალბათობა იმისა, რომ $\hat{\chi}^2$ სტატისტიკამ მიიღოს 10-ზე დიდი მნიშვნელობა, დაახლოებით 0.47-ის ტოლია (იხ. χ^2 განაწილების ცხრილი). ამის გამო, ვასკვნით, რომ ექსპერიმენტის მონაცემები ეთანხმება H_0 ჰიპოთეზას ნებისმიერი მნიშვნელოვნობის დონისათვის, რომელიც მოთავსებულია 0-სა და 0.47-ს შორის. ანუ მიღებული მონაცემები არ გვაძლევენ H_0 ჰიპოთეზის უარყოფის არანაირ საფუძველს, თუ α<0.47. ♦

χ^2-კრიტერიუმის გამოყენების სფერო არ ამოიწურება მხოლოდ მარტივი ჰიპოთეზების შემოწმებით. ამ კრიტერიუმის გამოყენება შესაძლებელია იმ შემთხვევაშიც, როდესაც **ჰიპოთეტური განაწილების ფუნქცია უცნობ პარამეტრებს შეიცავს**. მოვიყვანოთ შესაბამისი მაგალითები.

მაგალითი 4. დისკრეტული X პოპულაციიდან მიღებული 200 მოცულობის შერჩევით მნიშვნელობებს შორის შემთხვევითი სიდიდის მნიშვნელობა 0 შეგვხდა 14-ჯერ, მნიშვნელობა 1 - 31-ჯერ, მნიშვნელობა 2- 47-ჯერ, მნიშვნელობა 3 - 41-ჯერ, მნიშვნელობა 4 - 29-ჯერ, მნიშვნელობა 5 - 21-ჯერ, მნიშვნელობა 6 - 10-ჯერ, მნიშვნელობა 7 - 6-ჯერ, მნიშვნელობა 8 - 2-ჯერ. α=0,05 მნიშვნელოვნობის დონით შევამოწმოთ H_0 ნულოვანი ჰიპოთეზა იმის შესახებ, რომ შერჩევა მიღებულია უცნობი ინტენსივობის მქონე პუასონის კანონით განაწილებული გენერალური ერთობლიობიდან. ვინაიდან ჰიპოთეზა რთულია, უპირველეს ყოვლისა შევაფასოთ ინტენსივობა. მონაცემების საფუძველზე

$$\hat{\lambda} = \frac{1}{200}(1\cdot 31 + 2\cdot 47 + 3\cdot 41 + \ldots + 8\cdot 2) = 2.9\perp$$

ნულოვანი ჰიპოთეზის სამართლიანობის პირობებში, $\lambda=2.9$ ინტენსივობის მქონე პუასონის განაწილე-ბის სავარაუდო დაკვირვებული $n_e=n\cdot P\{X=k\}$ მნიშვნელობები მოცემულია 3 ცხრილში

ცხრილი 3

დაკვირვებები	სიხშირე		მოსალოდნელი სიხშირე
k	n_0	$P\{X=k\}$	n_e
0	14	0.0550	11.00
1	31	0.1596	31.91
2	47	0.2314	46.27
3	41	0.2237	44.73
4	29	0.1622	32.43
5	21	0.0940	18.81
6	10	0.0455	9.09
7	5	0.0188	3.77
8	2	0.0068	1.37
>8	0	0.0031	0.61

χ^2 სტატისტიკის რიცხვითი მნიშვნელობის გამოსათვლელად ვისარგებლოდ 4 ცხრილით

ცხრილი 4

k	n_0	n_e	n_0-n_e	$(n_0$-$n_e)^2$	$(n_0$-$n_e)^2/n_e$
0	14	11.00	3.00	8.9722	0.81530
1	31	31.91	-0.91	0.8344	0.02614
2	47	46.27	0.73	0.5263	0.01137
3	41	44.73	-3.73	13.9282	0.31136
4	29	32.43	-3.43	11.7699	0.36292
5	21	18.81	2.19	4.7969	0.25501
6	10	9.09	0.91	0.8255	0.09080
7	5	3.77	1.23	1.5217	0.40400
>8	2	1.98	0.02	0.0004	0.00020
					2.27714

ვინაიდან $N=200$ მნიშვნელობებს შორის არის k განსხვავებული მნიშვნელობა, χ^2 კრიტერიუმის სტატისტიკის თავისუფლების ხარისხი გამოითვლება ფორმულით.

$$k\text{-}p\text{-}1,$$

სადაც p წინასწარ შეფასებულ პარამეტრთა რაოდენობაა. ამრიგად, ვინაიდან $k-p-1=9-1-1=7$, χ^2 განაწილების ზედა კრიტიკული მნიშვნელობა $\chi^2_{7,0,05}=14.067.$, ხოლო სტატისტიკის დაკვირვებული მნიშვნელობა 2.77-ის ტოლია, H_0 ჰიპოთეზის უარყოფის საფუძველი არ გვაქვს.
♦

§ 3.* ორი შერჩევის დამოუკიდებლობის ჰიპოთეზის შემოწმება

წარმოვიდგინოთ, რომ პოპულაციიდან აღებულია n მოცულობის შერჩევა და ამ **შერჩევის ელემენტები კლასიფიცირებულია ორი რაიმე A და B ნიშნის მიხედვით**. დავუშვათ, რომ დაკვირვებათა ყველა შესაძლო შედეგი (ან პოპულაციის ყველა ელემენტი) **დაყოფილია A ნიშნით $A_1,\dots,A_k$, ხოლო B ნიშნით $B_1,\dots,B_r$ კატეგორიებად (ან კლასებად).** პოპულაციის ყოველი ელემენტი ეკუთვნის ზუსტად ერთ კატეგორიას A ნიშნის შესაბამისი რომელიმე კლასიდან და ასევე ზუსტად ერთ რომელიმე კატეგორიას B ნიშნის მიხედვით. ამიტომ **დაკვირვებული მონაცემები იყოფა $r\times k$ რაოდენობის A_iB_j არათავსებად ჯგუფად**.

თუ n_{ij} -ით აღვნიშნავთ იმ მონაცემთა რაოდენობას, რომლებიც ერთდროულად ეკუთვნიან A ნიშნის i-ურ და B ნიშნის j-ურ კატეგორიას და ჩავწერთ ამ სიდიდეს ცხრილის i-ური სვეტისა და j-ური სტრიქონის გადაკვეთაზე, მივიღებთ ორგანზომილებიან ($k\times r$-ზე) ნიშანთა შეუღლების ქვემოთ მოყვანილ ცხრილს, რომელსაც იყენებენ A და B ნიშნების დამოუკიდებლობის ჰიპოთეზის შესამოწმებლად.

ცხრილი **5** ნიშანთა შეუღლების სიხშირული ცხრილი

A \ B	B_1	B_2	…	B_j	…	B_r	Σ
A_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1r}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2r}	$n_{2\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{ir}	$n_{i\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_k	n_{k1}	n_{k2}	…	n_{kj}	…	n_{kr}	$n_{k\bullet}$
Σ	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet r}$	n

(12.5)

ამ ცხრილში **($n_{i\bullet}$,$1\le i\le k$) და ($n_{\bullet j}$,$1\le j\le r$) სიდიდეები აღნიშნავს A და B ნიშნების შესაბამის მარგინალურ სიხშირეებს.**

$n_{i\bullet}$ წარმოადგენს შერჩევის იმ ელემენტთა სიხშირეს, რომლებიც მოხვდნენ i-ურ კლასში A ნიშნის მიხედვით, ხოლო $n_{\bullet j}$ არის შერჩევის ელემენტთა რაოდენობა, რომლებიც მოხვდნენ j-ურ კლასში B ნიშნით.

შევნიშნოთ, რომ $n_{i\bullet}$ (შესაბამისად $n_{\cdot j}$) i-ურ სტრიქონში (შესაბამისად j-ურ სვეტში) მოცემულ სიხშირეთა ჯამის ტოლია

$$n_{i\bullet}=\sum_{j=1}^{r} n_{ij}\ ,\ n_{\bullet j}=\sum_{i=1}^{k} n_{ij}\ .$$

გარდა ამისა, ცხრილის ყველა n_{ij} სიხშირეთა ჯამი დაკვირვებათა საერთო რიცხვის, ანუ შერჩევის მოცულობის ტოლია და

$$\sum_{i=1}^{k} n_{i\bullet} = \sum_{j=1}^{r} n_{\bullet j} = \sum_{i=1}^{k}\sum_{j=1}^{r} n_{ij} = n.$$

შემოვიღოთ შემდეგი აღნიშვნები

P_{ij}-ით აღვნიშნოთ ალბათობა იმისა, რომ პოპულაციიდან შემთხვევით არჩეულ ელემენტი აღმოჩნდება ერთდროულად A ნიშნის i-ური და B ნიშნის j-ური კატეგორიაში. მაშინ

$P_{i\bullet} = \sum_{j=1}^{r} P_{ij}$ იქნება ალბათობა იმისა, რომ პოპულაციის ელემენტი აღმოჩნდება i-ური კატეგორიაში A ნიშნის მიხედვით, ხოლო

$P_{\bullet j} = \sum_{i=1}^{k} P_{ij}$ – ალბათობა იმისა, რომ პოპულაციის ელემენტი მოხვდება j-ურ კატეგორიაში B ნიშნით.

ამრიგად, ჩვენ შეგვიძლია ჩავთვალოთ, რომ (12.5) წარმოადგენს n დამოუკიდებელი დაკვირვების შედეგს ალბათურ მოდელზე, რომლის ერთობლივი განაწილების კანონია

ცხრილი 6 ნიშანთა ერთობლივი განაწილების ცხრილი

X \ Y	B_1	B_2	…	B_j	…	B_r	Σ
A_1	P_{11}	P_{12}	…	P_{1j}	…	P_{1r}	$P_{1\bullet}$
A_2	P_{21}	P_{22}	…	P_{2j}	…	P_{2r}	$P_{2\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_i	P_{i1}	P_{i2}	…	P_{ij}	…	P_{ir}	$P_{i\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_k	P_{k1}	P_{k2}	…	P_{kj}	…	P_{kr}	$P_{k\bullet}$
Σ	$P_{\bullet 1}$	$P_{\bullet 2}$	…	$P_{\bullet j}$	…	$P_{\bullet r}$	*1*

(12.6)

სადაც $A_1,A_2,...,A_k$ არის A ნიშნის შესაძლო შედეგი, ხოლო $B_1,B_2,...,B_r$ – არის B ნიშნის შესაძლო შედეგია.

თანხმობის ჰიპოთეზის შემოწმება.

როდესაც **P_{ij} ალბათობები მოცემულია**, ამ ალბათობათა (12.5) დაკვირვებულ მონაცემებთან **თანხმობის ჰიპოთეზის შესამოწმებლად შეიძლება გამოვიყენოთ**

$$\hat{\chi}^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - n P_{ij}\right)^2}{n P_{ij}}. \qquad (12.7)$$

სტატისტიკა, რომელიც მიახლოებით χ^2 კანონით არის განაწილებული, რომლის თავისუფლების ხარისხი $f = (rk-1)$-ის ტოლია, ერთადერთი, $\sum_{i=1}^{k}\sum_{j=1}^{r} n_{ij} = n$ შეზღუდვის გამო.

დამოუკიდებლობის ჰიპოთეზის შემოწმება.

როდესაც **A და B ნიშნები დამოუკიდებელია, მაშინ ყოველი i-სა და j-სათვის უნდა შესრულდეს ტოლობა**

$$P_{ij} = P_{i\bullet} P_{\bullet j}, \quad 1 \le i \le k, \ 1 \le j \le r, \qquad (12.8)$$

სადაც

$$\sum_{i=1}^{k} P_{i\bullet} = 1, \qquad \sum_{j=1}^{r} P_{\bullet j} = 1. \tag{12.9}$$

<u>როგორ უნდა შევამოწმოთ *A* და *B* ნიშნების დამოუკიდებლობის ჰიპოთეზა არსებული დაკვირვებების საფუძველზე?</u>

ა) <u>განვიხილოთ შემთხვევა, როდესაც $P_{i\bullet}$ და $P_{\bullet j}$ მარგინალური განაწილებები ცნობილია.</u>

ამ შემთხვევაში (12.7) **ფორმულით მოცემული $\hat{\chi}^2$ სტატისტიკა, P_{ij} ალბათობების მარგინალური განაწილებებით გამოსახვის შემდეგ იღებს შემდეგ სახეს:**

$$\hat{\chi}^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - n P_{i\bullet} P_{\bullet j}\right)^2}{n P_{i\bullet} P_{\bullet j}} \tag{12.10}$$

(12.7) გვაძლევს ჯამურ გადახრას მოსალოდნელ და დაკვირვებულ სიხშირეებს შორის. ამ სტატისტიკამ შეიძლება კრიტიკულ მნიშვნელობაზე დიდი მნიშვნელობა მიიღოს ჰიპოთეტურ მარგინალურ განაწილებებსა და მარგინალურ სიხშირეებს შორის განსხვავების და არა *A* და *B* ნიშნებს შორის დამოკიდებულების არსებობის გამო. ამიტომ $\hat{\chi}^2$ **სტატისტიკიდან გამოყოფენ ორ კომპონენტს**

$$\hat{\chi}_1^2 = \sum_{i=1}^{k} \frac{\left(n_{i\bullet} - n P_{i\bullet}\right)^2}{n P_{i\bullet}}, \qquad \hat{\chi}_2^2 = \sum_{j=1}^{r} \frac{\left(n_{\bullet j} - n P_{\bullet j}\right)^2}{n P_{\bullet j}} \qquad , \tag{12.11}$$

(რომლებიც გამოსახავენ ჰიპოთეტურ მარგინალურ განაწილებებსა და მარგინალურ სიხშირეებს შორის გადახრის სიდიდეს) და **დამოუკიდებლობის ჰიპოთეზის შესამოწმებლად იყენებენ $\hat{\chi}_3^2 = \hat{\chi}^2 - \hat{\chi}_1^2 - \hat{\chi}_2^2$ სხვაობას.** (12.10) და (12.11) ფორმულებიდან, მარტივი **გარდაქმნების შემდეგ ვიღებთ, რომ**

$$\hat{\chi}_3^2 = \hat{\chi}^2 - \hat{\chi}_1^2 - \hat{\chi}_2^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - n_{i\bullet}P_{\bullet j} - n_{\bullet j}P_{i\bullet} + n P_{i\bullet}P_{\bullet j}\right)^2}{n P_{i\bullet} P_{\bullet j}} \ . \tag{12.12}$$

რადგან $\hat{\chi}^2$, $\hat{\chi}_1^2$ და $\hat{\chi}_2^2$ სტატისტიკებს დაახლოებით აქვთ χ^2 განაწილება და მათი თავისუფლების ხარისხები შესაბამისად (*rk*–1)-ის, (*k*–1)-ისა და (*r*–1)-ის ტოლია (შევნიშნოთ, რომ $\hat{\chi}_3^2$ სტატისტიკა დადებითია), $\hat{\chi}_3^2$ სტატისტიკაც მიახლოებით განაწილებული იქნება χ^2 კანონით და მისი თავისუფლების ხარისხი ტოლი იქნება კომპონენტების თავისუფლების ხარისხთა სხვაობის, ანუ $f = (rk-1) - (r-1) - (k-1) = (r-1)(k-1)$. ამრიგად, $\hat{\chi}_3^2$ სტატისტიკა შეიძლება გამოყენებული იქნეს ნიშანთა დამოუკიდებლობის ჰიპოთეზის შესამოწმებლად, ცნობილი მარგინალური განაწილებების შემთხვევაში.

ორი შერჩევის დამოუკიდებლობის ჰიპოთეზის შემოწმება.

ორგანზომილებიან (*k*×*r*-ზე) ნიშანთა შეუღლების (12.5) **ცხრილის სახის მონაცემები არის *n* დამოუკიდებელი დაკვირვების შედეგი ალბათურ მოდელზე, რომლის ერთობლივი**

განაწილების კანონია (12.6) და $P_{i\bullet}$, $P_{\bullet j}$ მარგინალური განაწილებები ცნობილია.

ნულოვანი ჰიპოთეზა: $H_0 : P_{1j} = P_{2j} = \cdots = P_{kj}$, $j=1,2,\ldots,r$

ალტერნატიული ჰიპოთეზა: H_a : H_0 არ არის მართებული

სტატისტიკის მნიშვნელობა $\hat{\chi}^2_3 = \hat{\chi}^2 - \hat{\chi}^2_1 - \hat{\chi}^2_2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - n_{i\bullet}P_{\bullet j} - n_{\bullet j}P_{i\bullet} + nP_{i\bullet}P_{\bullet j}\right)^2}{nP_{i\bullet}P_{\bullet j}}$

უარყოფის არე $\chi^2 \geq \chi^2_{\alpha,(k-1)(r-1)}$

მაგალითი 5. საილუსტრაციოდ მოვიყვანოთ ბეიტსონის ცნობილი მაგალითი ტკბილი მუხუდოს ორი ნიშნის გენის დამოკიდებულობის შესახებ. ერთი ნიშანი წარმოადგენს ყვავილის შეფერილობას (იისფერი ან წითელი), ხოლო მეორე შეესაბამება ყვავილის მტვრის ფორმას (მოგრძო ან მრგვალი). (იხილე მონაცემები ქვემოთ მოყვანილ ნიშანთა შეუღლების (7 ცხრილში)

ცხრილი 7. ტკბილი მუხუდოს განაწილება მისი ყვავილის ფერისა და მტვრის ფორმის მიხედვით

	ყვავილის შეფერილობა		
მტვრის ფორმა	იისფერი	წითელი	სულ
მოგრძო	296	27	323
მრგვალი	19	85	104
სულ	315	112	427

ჩავთვალოთ, რომ 7 ცხრილი წარმოადგენს *n=427* დამოუკიდებელი დაკვირვების შედეგს ალბათურ მოდელზე, რომლის ერთობლივი განაწილების კანონია

ცხრილი 8.

	ყვავილის შეფერილობა		
მტვრის ფორმა	იისფერი	წითელი	სულ
მოგრძო	9/16	3/16	3/4
მრგვალი	3/16	1/16	1/4
სულ	3/4	1/4	1

განაწილების კანონის თანახმად დაშვებულია, რომ მარგინალური განაწილებები იყოფიან 1:3 პროპორციით, ანუ ალბათობა იმისა, რომ შემთხვევით აღებულ მუხუდოს ყვავილს აღმოაჩნდება წითელი ფერი (შესაბამისად იისფერი) ტოლია 1/4-ის (შესაბამისად 3/4-ის), ხოლო ალბათობა იმისა, რომ მისი ყვავილის მტვრის ფორმა მგვრალი ან მოგრძოა ასევე შესაბამისად 1/4-ისა და 3/4-ის ტოლია. თუ ეს ორი მახასიათებელი მემკვიდრეობით ერთმანეთისაგან დამოუკიდებლად გადაეცემა, მაშინ სიხშირეები, ნიშანთა შეუღლების 7 ცხრილში, დაახლოებით უნდა აკმაყოფილებდნენ 9:3:3:1 ფარდობას.

ვიცით რა მარგინალური განაწილებები, ნულოვანი ჰიპოთეზის დროს ვიანგარიშოთ $\hat{\chi}^2$ სტატისტიკის მნიშვნელობა (12.10) ფორმულის მიხედვით შემდეგნაირად

	n_{ij}	$P_{ij} = P_{i\bullet}P_{\bullet j}$	$n \cdot P_{i\bullet}P_{\bullet j}$	$n_{ij} - n \cdot P_{i\bullet}P_{j\bullet}$	$\left(n_{ij} - n \cdot P_{i\bullet}P_{j\bullet}\right)^2$	$\left(n_{ij} - n \cdot P_{i\bullet}P_{j\bullet}\right)^2 / nP_{i\bullet}P_{\bullet j}$
მოგრძო/ იისფერი	296	0.75	240.1875	55.8125	3115.035	12.96918
მრგვალი/ იისფერი	19	0.25	80.0625	61.0625	3728.629	46.57148
მოგრძო/ წითელი	27	0.75	80.0625	53.0625	2815.629	35.16789
მრგვალი/ წითელი	85	0.25	26.6875	58.3125	3400.348	127.4135
n=	427				$\hat{\chi}^2$ =	222.122

ანალოგიურად გამოითვლება $\hat{\chi}_1^2$ და $\hat{\chi}_2^2$ სტატისტიკების რიცხვითი მნიშვნელობები:

	$n_{i\cdot}$	$P_{i\cdot}$	$n\cdot P_{i\cdot}$	$n_{i\cdot} - n\cdot P_{i\cdot}$	$(n_{i\cdot} - n\cdot P_{i\cdot})^2$	$(n_{i\cdot} - n\cdot P_{i\cdot})^2/n\cdot P_{i\cdot}$
მოგრძო	323	0.75	320.25	2.75	7.5625	0.0236
მრგვალი	104	0.25	106.75	-2.75	7.5625	0.0708
n=	427				$\hat{\chi}_3^2$ =	0.0944

	$n_{\cdot j}$	$P_{\cdot j}$	$n\cdot P_{\cdot j}$	$n_{\cdot j} - n\cdot P_{\cdot j}$	$(n_{\cdot j} - n\cdot P_{\cdot j})^2$	$(n_{\cdot j} - n\cdot P_{\cdot j})^2 / n\cdot P_{\cdot j}$
იისფერი	315	0.75	320.25	-5.25	27.5625	0.086066
წითელი	112	0.25	106.75	5.25	27.5625	0.258197
n=	427				$\hat{\chi}_3^2$ =	0.3443

ამრიგად

$$\hat{\chi}^2 = 222.122,\ \hat{\chi}_1^2 = 0.0945 \text{ და } \hat{\chi}_2^2 = 0.3443-$$

და ალბათობა იმისა, რომ 3-ის ტოლი თავისუფლების ხარისხის χ^2 განაწილების მქონე შემთხვევითმა სიდიდემ მიიღოს ამ რიცხვზე დიდი მნიშვნელობა, ძალიან მცირეა (მაგალითად, ამ განაწილების 0.01 ალბათობის შესაბამისი კრიტიკული მნიშვნელობა 11.345-ის ტოლია) ეს მიუთითებს იმაზე, რომ საერთო გადახრა ექსპერიმენტის რეალურ და მოსალოდნელ შედეგებს შორის დიდია. ვნახოთ, ხომ არ არის ეს გადახრა გამოწვეული იმით, რომ ცხრილში მოცემული მარგინალური სიხშირეები არ ეთანხმებიან ჩვენს დაშვებას ცალკეულ ნიშანთა 1:3 პროპორციით განაწილების შესახებ. (იხ. ცხრილი 8) $\hat{\chi}_1^2$ და $\hat{\chi}_2^2$ სტატისტიკების რიცხვითი მნიშვნელობები ძალიან მცირეა 1-ის ტოლი თავისუფლების ხარისხის მქონე χ^2 განაწილებისათვის. ეს ნიშნავს, რომ ცალკე აღებული ვექტორი (ყვავილის ფერი ან მისი მტვრის ფორმა) ისეთი სიხშირით ჩნდება, როგორც მოსალოდნელი იყო.

მესამე კომპონენტისათვის გვაქვს:

$$\hat{\chi}_3^2 = 221.6833,$$

რაც ძალიან დიდია 1-ის ტოლი თავისუფლების ხარისხის χ^2 განაწილებისათვის. ამ სტატისტიკების საერთო მნიშვნელობა

$$\hat{\chi}^2 = 0.0945+0.3443+221.6833 = 222.1221$$

აჩვენებს, რომ სრული გადახრა ჰიპოთეზისაგან კონცენტრირებულია ჯამის მესამე კომპონენტში. ამიტომ, გადახრა ჰიპოთეტურ და რეალურ სიხშირეებს შორის გამოწვეულია განხილულ ნიშნებს შორის დამოკიდებულებით და ჰიპოთეზა იმის შესახებ, რომ ყვავილის ფერი და მისი მტვრის ფორმა მემკვიდრეობით ერთმანეთისაგან დამოუკიდებლად გადაეცემა უნდა უკუვაგდოთ. ♦

ბ) განვიხილოთ შემთხვევა, როდესაც $P_{i\bullet}$ და $P_{\bullet j}$ მარგინალური განაწილებები უცნობია.

ამ შემთხვევაში ნიშანთა დამოუკიდებლობის H_0 ჰიპოთეზა, რომელიც (12.8) ტოლობის შესრულებას ნიშნავს, არ აზუსტებს უცნობ პარამეტრთა (ანუ მარგინალურ განაწილებათა) მნიშვნელობას და საჭირო ხდება მათი შეფასება მოცემული შერჩევის საშუალებით. თუ

$$\hat{P}_{i\bullet} = \frac{n_{i\bullet}}{n}, \qquad \hat{P}_{\bullet j} = \frac{n_{\bullet j}}{n}. \tag{12.13}$$

შეფასებებს ჩავსვამთ (12.10) ფორმულით მოცემულ გამოსახულებაში, მარგინალური განაწილებების ნაცვლად, მივიღებთ

$$\hat{\chi}^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - \dfrac{n_{i\bullet}\, n_{\bullet j}}{n}\right)^2}{\dfrac{n_{i\bullet}\, n_{\bullet j}}{n}} \tag{12.14}$$

სტატისტიკას.

ორი შერჩევის დამოუკიდებლობის ჰიპოთეზის შემოწმება.

ორგანზომილებიან ($k\times r$-ზე) ნიშანთა შეუღლების 5 ცხრილის სახის მონაცემები არის n დამოუკიდებელი დაკვირვების შედეგი ალბათურ მოდელზე, რომლის ერთობლივი განაწილების კანონია (12.6) **და $P_{i\bullet}$, $P_{\bullet j}$ მარგინალური განაწილებები უცნობია.**

ნულოვანი ჰიპოთეზა: $H_0 : P_{1j} = P_{2j} = \cdots = P_{kj}$, $\quad j=1,2,\ldots,r$

ალტერნატიული ჰიპოთეზა: H_a : H_0 არ არის მართებული

სტატისტიკის მნიშვნელობა $$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} \frac{\left(n_{ij} - \dfrac{n_{i\bullet}\, n_{\bullet j}}{n}\right)^2}{\dfrac{n_{i\bullet}\, n_{\bullet j}}{n}}$$

უარყოფის არე $\chi^2 \geq \chi^2_{\alpha,(k-1)(r-1)}$

მაგალითი 6. სოციოლოგს სურს შეამოწმოს ჰიპოთეზა იმის შესახებ, რომ ოჯახში ბავშვების რაოდენობა არ არის დამოკიდებული ოჯახის შემოსავალზე. შემთხვევითი შერჩევის შედეგები, რომელიც მიღებულია 385 ოჯახზე დაკვირვებით, მოცემულია 9 ცხრილში:

ცხრილი 9.

ბავშვების რაოდენობა	შემოსავალი (ათასობით $-ში)				
	0–6 Aჯგუფი	6–12 Bჯგუფი	12–18 Cჯგუფი	18-ზე მეტი Dჯგუფი	სულ
0	10	9	18	24	61
1	8	12	25	31	76
2	24	28	23	28	103
3	26	24	20	6	76
4 ან მეტი	32	22	18	7	79
სულ	100	95	104	96	395

არის თუ არა ეს მონაცემები საკმარისად თვალსაჩინო იმისათვის, რომ α=0.01 ნდობის დონით მივიღოთ ჰიპოთეზა ოჯახში ბავშვების რაოდენობის ოჯახის შემოსავლისაგან დამოუკიდებლობის შესახებ?

ამოხსნა: χ^2 სტატისტიკის დაკვირვებული მნიშვნელობის გამოსათვლელად ვისარგებლოდ შემდეგი ცხრილით

	n_{ij}^0	$n_{ij}^e = \frac{n_{i\bullet} n_{\bullet j}}{n}$	$n_{ij}^0 - n_{ij}^e$	$\left(n_{ij}^0 - n_{ij}^e\right)^2$	$\left(n_{ij}^0 - n_{ij}^e\right)^2 / n_{ij}^e$
A0	10	15.44	-5.44	29.63	1.92
A1	8	19.24	-11.24	126.35	6.57
A2	24	26.08	-2.08	4.31	0.17
A3	26	19.24	6.76	45.69	2.37
A4	32	20.00	12.00	144.00	7.20
B0	9	14.67	-5.67	32.16	2.19
B1	12	18.28	-6.28	39.42	2.16
B2	28	24.77	3.23	10.42	0.42
B3	24	18.28	5.72	32.74	1.79
B4	22	19.00	3.00	9.00	0.47
C0	18	16.06	1.94	3.76	0.23
C1	25	20.01	4.99	24.90	1.24
C2	23	27.12	-4.12	16.97	0.63
C3	20	20.01	-0.01	0.00	0.00
C4	18	20.80	-2.80	7.84	0.38
D0	24	14.83	9.17	84.17	5.68
D1	31	18.47	12.53	156.98	8.50
D2	28	25.03	2.97	8.80	0.35
D3	6	18.47	-12.47	155.52	8.42
D4	7	19.20	-12.20	148.84	7.75
					58.44

მესამე სვეტში მოთავსებული რიცხვები წარმოადგენს ცხრილის $(i;j)$ უჯრაში ოჯახების მოსალოდნელ რაოდენობის შეფასებას.

ამრიგად, სტატისტიკის დაკვირვებული მნიშვნელობა 58.44-ის ტოლია, ხოლო თავისუფლების ხარისხი ტოლია $(r-1)(k-1)=12$. კრიტიკული მნიშვნელობაა $\chi^2_{12,\ 0.01}=26.2170$.

გადაწყვეტილება: რადგან $\hat{\chi}^2 > \chi^2_{12,\ 0.01}$, ჩვენ უარვყოფთ H_0-ჰიპოთეზას, ანუ ვასკვნით, რომ ოჯახში ბავშვების რაოდენობა და ოჯახის შემოსავალი არ წარმოადგენს დამოუკიდებელ კლასიფიკაციებს.♦

კითხვები თვითშემოწმებისათვის

1. ჩამოაყალიბე არაპარამეტრული ჰიპოთეზა დისკრეტული განაწილებისთვის
2. განსაზღვრე χ^2 განაწილება დისკრეტული განაწილებისთვის
3. განსაზღვრე გადაწყვეტილების მიღების წესი არაპარამეტრულ ჰიპოთეზაში დისკრეტული განაწილებისთვის
4. ჩამოაყალიბე არაპარამეტრული ჰიპოთეზის შემოწმების ალგორითმი უწყვეტი განაწილების ფუნქციისთვის
5. უწყვეტი პოპულაციის მნიშვნელობათა სიმრავლის დაყოფის ორი: ტოლ ალბათობათა და ტოლ მნიშვნელობათა დაყოფის ხერხები არაპარამეტრული ჰიპოთეზის შემოწმების ალგორითმში
6. განსაზღვრე χ^2 განაწილება უწყვეტი განაწილებისთვის
7. განსაზღვრე გადაწყვეტილების მიღების წესი არაპარამეტრულ ჰიპოთეზაში უწყვეტი განაწილებისთვის
8. განსაზღვრე ერთობლივი ნიშანთა (ორნიშნიანი) შეუღლების სიხშირული ცხრილი
9. განსაზღვრე ორნიშნიანი ერთობლივი ალბათური განაწილების ცხრილი
10. ჩამოაყალიბე ორი შერჩევის დამოუკიდებლობის ჰიპოთეზა, როდესაც $P_{i\bullet}$ და $P_{\bullet j}$ მარგინალური განაწილებები ცნობილია.
11. ჩამოაყალიბე ორი შერჩევის დამოუკიდებლობის ჰიპოთეზა, როდესაც $P_{i\bullet}$ და $P_{\bullet j}$ მარგინალური განაწილებები უცნობია.

სავარჯიშოები

12.1. 164 სტუდენტის სიმაღლეთა გაზომვით მიღებული შედეგები დაჯგუფებულია მონაცემების ხუთ სხვადასხვა ინტერვალში მოხვედრის მიხედვით

ინტერვალის სიგრძე დიუიმებში	სიხშირეები
61–64	10
65–68	21
69–72	83
73–76	49
77–80	1

ეთანხმება თუ არა ეს მონაცემები ჰიპოთეზას, რომ სიმაღლეთა განაწილება ნორმალურია საშუალო მნიშვნელობით 70 დიუიმი და საშუალო კვადრატული გადახრით 3 დიუიმი (1 დიუიმი=2.5სმ)

12.2. ქვემოთ მოყვანილ ცხრილში აღნიშნულია ერთ მაღაზიაში კონკრეტული საქონლის რეალიზაციათა რაოდენობები კვირის დღეების მიხედვით

დღე	ორშაბათი	სამშაბათი	ოთხშაბათი	ხუთშაბათი	პარასკევი
გაყიდული საქონლის რაოდენობა	49	35	32	39	45

შეიძლება თუ არა დავასკვნათ, რომ გაყიდული საქონელი თანაბრად არის განაწილებული კვირის დღეების მიხედვით (ანუ ეთანხმება თუ არა ეს მონაცემები თანაბარ დისკრეტულ განაწილებას)?

12.3. მოცემულია 1725 მოსწავლის კლასიფიკაცია (1) ტანსაცმლის ხარისხისა და (2) გონებრივი უნარის მიხედვით. მეორე ნიშნის შემთხვევაში მიღებული იყო ასეთი გრადაცია *A* – გონებრივად ჩამორჩენილი, *B* – ნელი და ჩლუნგი, *C* – ჩლუნგი, *D* – ნელი მაგრამ ჭკვიანი, *E* – საკმაოდ ჭკვიანი, *F* – აშკარად უნარიანი, *G* – ფრიად უნარიანი. განსაზღვროთ, არის თუ არა ჩაცმულობის ხარისხი დამოკიდებული გონებრივ უნარზე.

როგორ იცვამს	უნარი						
	*A*და *B*	*C*	*D*	***E***	*F*	*G*	ჯამი
ძალიან კარგად	33	48	113	209	194	39	636
კარგად	41	100	202	255	138	15	751
არა უშავს	39	58	70	61	33	4	265
ცუდად	17	13	22	10	10	1	73
ჯამი	130	219	407	535	375	59	1725

12.4. ცხრილში მოყვანილია 818 შემთხვევა, რომლებიც კლასიფიცირებულია ორი ნიშნის მიხედვით (1) აცრილია თუ არა ქოლერის წინააღმდეგ (2) დაავადდა თუ არა.

	არადაავადებულნი	დაავადებულნი	ჯამი
აცრილები	276	3	279

არააცრილები	473	66	539
ჯამი	749	69	818

განსაზღვრეთ ამ მონაცემების მიხედვით, ეფექტურია თუ არა აცრა ავადმყოფობის გავრცელების წინააღმდეგ.

12.5. შევადაროთ მკურნალობის ორი მეთოდი. 80 ავადმყოფიდან 40 მკურნალობს სპეციფიკური საშუალებით, 40 კი კონსერვატორულად. მკურნალობის შედეგები გაყოფილია 3 კლასად: სწრაფი განკურნება, ნელი განკურნება, უშედეგო მკურნალობა. შედეგები თავმოყრილია ცხრილში.

მკურნალობის შედეგი	მ კ უ რ ნ ა ლ ო ბ ა		ჯამი
	კონსერვატორული	სპეციფიკური	
სწრაფი განკურნება	14	22	36
ნელი განკურნება	18	16	34
უშედეგო მკურნალობა	8	2	10
ჯამი	40	40	80

განსაზღვრეთ, მოქმედებს თუ არა ახალი მეთოდით მკურნალობა მკურნალობის შედეგზე.

12.6. ვთქვათ გვაინტერესებს განვსაზღვროთ არის თუ არა არსებითი განსხვავება რეგულარული სატელევიზიო პროგრამების მაყურებელთა შორის მათი სქესის მიხედვით. აღებულია შემთხვევითი შერჩევა და ყოველ პიროვნებას ეკითხებიან თუ რომელ პროგრამას ანიჭებს ის უპირატესობას. გამოკითხვის შედეგები მოყვანილია ცხრილში

სასურველი სატელევიზიო პროგრამა	მაყურებელთა სქესი		ჯამი
	კაცები	ქალები	
1) ვესტერნი	32	18	50
2) ტრილერი	17	13	30
3) დრამა	27	33	60
4) კომედია	13	7	20
5) ვარიეტე	24	16	40
ჯამი	113	87	200

განსაზღვრეთ, ეწინააღმდეგება თუ არა მოყვანილი მონაცემები ნულოვან ჰიპოთეზას სატელევიზიო პროგრამების მაყურებელთა სქესისაგან დამოუკიდებლობის შესახებ.

12.7. მარკეტინგის მენეჯერს სურს განსაზღვროს არის თუ არა არსებითი განსხვავება რეგიონებს შორის, გარკვეული ახალი პროდუქტის მოხმარების თვალსაზრისით. ქვემოთ მოყვანილი მონაცემები მიღებულია ამ პროდუქტის მომხმარებელთა შერჩევითი მეთოდით გამოკითხვის შედეგად.

მოხმარების ხარისხი	ქალაქი				ჯამი
	ბათუმი	ქუთაისი	თბილისი	თელავი	
დაბალი 1)	22	35	0	5	62
ზომიერი 2)	84	55	8	24	171
მაღალი 3)	25	17	22	12	76
ჯამი	131	107	30	41	309

ნულოვანი ჰიპოთეზა ნიშნავს ქალაქისაგან პროდუქტის მოხმარების ხარისხის დამოუკიდებლობას. იანგარიშეთ χ^2 სტატისტიკა. განსაზღვრეთ ძირითადი სტატისტიკის თავისუფლების ხარისხი. განსაზღვრეთ ძირითადი სტატისტიკის კრიტიკული მნიშვნელობა. განსაზღვრეთ, უნდა მივიღოთ, თუ უკუვაგდოთ ნულოვანი ჰიპოთეზა.

ლექცია XIV, XV უმცირეს კვადრატთა მეთოდი და წრფივი რეგრესია

§1. უმცირეს კვადრატთა მეთოდის წარმოდგენა

ვთქვათ რაიმე სტატისტიკური პოპულაციის აღმწერი X და Y შემთხვევითი სიდიდეები დაკავშირებულია $Y=\phi(X)$ ფუნქციონალური დამოკიდებულებით, თანაც ϕ – **ჩვენთვის უცნობი ფუნქციაა. სტატისტიკის ერთ-ერთი ძირითადი ამოცანაა X დაY სიდიდეებზე ცდისეული დაკვირვებების საფუძველზე განისაზღვროს.** ϕ **დამოკიდებულება.**

დავუშვათ რომ არსებობს საშუალება ცდისეულ $X = x_i$ მნიშვნელობაზე (X სიდიდის i-ურ რეალიზაციაზე) განისაზღვროს, გაიზომოს (ცდისეულად) Y სიდიდე. ე.ი. ზომადია შემდეგი სიდიდე:

$$\{Y/X = x_i\}.$$

ვთქვათ ტარდება n დამოუკიდებელი ცდა. i -ური ცდის ჩატარებით Y სიდიდის რეალიზაცია აღვნიშნოთ y_i -ით, მაშინ y_i x_i სიდიდესთან შეიძლება ფუნქციონალურად ასე დავაკავშირით:

$$y_i = \phi(x_i) + \varepsilon_i,$$

სადაც ვუშვებთ რომ ε_i – i -ური ცდის **"შემთხვევითი ცდომილებაა"**, ისე რომ ε_i-ცდომილების მოლოდინი 0-ის ტოლია. დავუშვათ, რომ

$$E\{Y/X= x_i\} =\phi(x_i).$$

ე.ი. Y შემთხვევითი სიდიდის "საშუალო" მნიშვნელობა (მათ.ლოდინი), როცა X სიდიდე მიიღებს თავის x_i მნიშვნელობას ტოლია $\phi(x_i)$-ის. ამ **პირობით მათემატიკურ ლოდინს**(როგორც x-ზე დამოკიდებულ ფუნქციას),

$$E\{Y/X = x\} =\phi(x)$$

ეწოდება Y სიდიდის რეგრესიის წირი X სიდიდეზე.

$Y=\phi(X)$ – ფუნქციონალური დამოკიდებულების აღდგენა, შეფასება, აგება შერჩევითი $\{x_i,y_i\}_{i=1}^{n}$ სიდიდეებით – გამოყენებითი სტატისტიკის უმნიშვნელოვანესი ამოცანაა.მას *<u>რეგრესიული ანალიზის ამოცანა</u>* **ეწოდება.**

შევნიშნოთ, რომ $E\varepsilon_i=0$, ე.ი. შემთხვევითი ცდომილების "საშუალო" ნულის ტოლია, რაც რეალობის ბუნებრივი წარმოდგენაა. **განვიხილავთ შემთხვევას, როცა $D\varepsilon_i=\sigma^2$, $i=\overline{1,n}$ და ε_i – ნორმალური კანონითაა განაწილებული: ε_i – $N(0,\sigma^2)$. ε_i დამოუკიდებელია ε_j -გან $(i \neq j)$.**

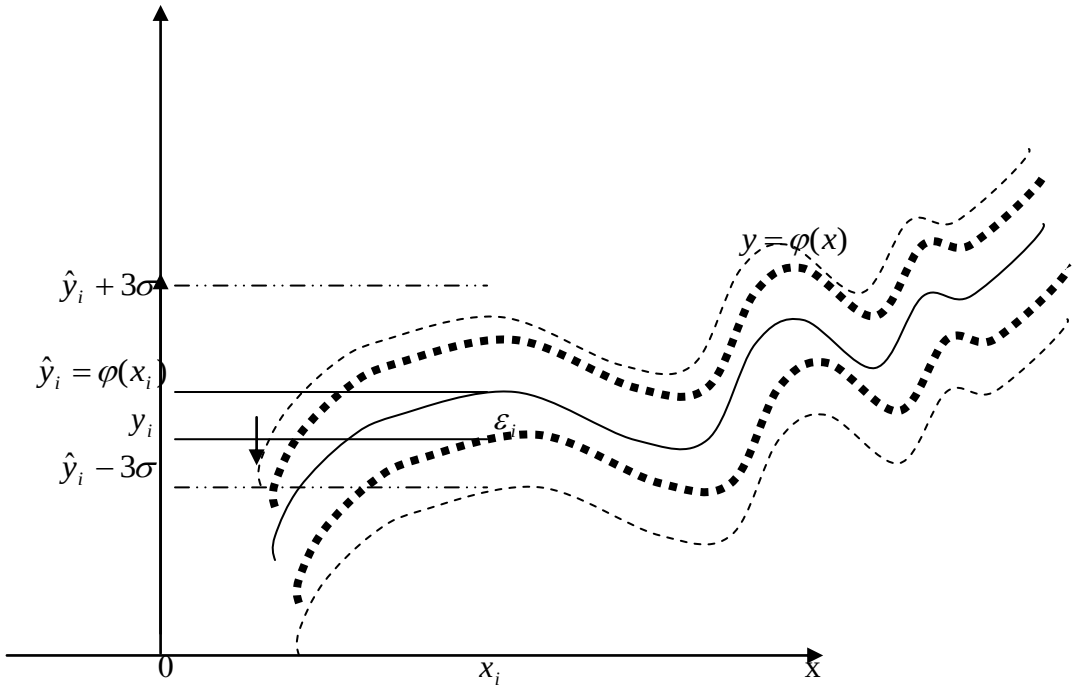
ვთქვათ $X= x_i$ რეალიზაციაზე Y სიდიდის **მოდელური, ფუნქციონალური მნიშვნელობაა**

$$\hat{y}_i = \phi(x_i).$$

ხოლო y_i – i - ური ცდის რეალიზაციაა Y სიდიდეზე, რომელიც **ზოგადად განსხვავდება მისი მოდელური (რეგრესიული)** $\hat{y}_i$ **მნიშვნელობისგან** და განსხვავება ტოლია i -ური ცდის შემთხვევითი ცდომილების:

$$y_i - \hat{y}_i = \phi(x_i) + \varepsilon_i - \phi(x_i) = \varepsilon_i.$$

თუ საკოორდინატო სიბრტყეზე მოვნიშნავთ (x_i, y_i) წერტილებს და გავატარებდით სავარაუდო $y=\phi(x)$ - წირს, მაშინ (x_i, y_i) წერტილები გაბნეული იქნებიან წირის გასწვრივ ε_i "მცირე" ცდომილების რიგით. ე.ი. y_i –შემთხვევითი სიდიდე განაწილებულია ნორმალური კანონით, მათემატიკური მოლოდინით: $E y_i = \hat{y}_i$, დისპერსიით: $D y_i=\sigma^2$. *"3σ "-ს წესის გათვალისწინებით შეიძლება წარმოვადგინოთ შემდეგი გრაფიკული ინტერპრეტაცია:*



<u>ჩვენ განვიხილავთ შემდეგ რეგრესიულ ამოცანას</u>: $\{x_i,y_i\}_{i=1}^{n}$ -ცდისეულ მონაცემებზე დაყრდნობით აიგოს ისეთი ϕ- ფუნქცია, რომელიც "საუკეთესოდ" აღწერდეს (ეთანადებოდეს) ცდისეულ შედეგებს გარკვეული აზრით.

ეს ამოცანა ცნობილია ექსპერიმენტულ მონაცემთა "გაგლუვების" ამოცანის სახით, რომლის ზოგადი ამოცანიდან გამოვყოფთ ერთ, მეტად გავრცელებულ ამოცანას, როდესაც არ არსებობს $y=\phi(x)$ ფუნქციის სახე, თუმცა ის ეკუთვნის გარკვეულ ფუნქციათა **F** კლასს, სადაც ფუნქციები განსხვავდებიან ერთი ან რამოდენიმე პარამეტრის რიცხვითი მნიშვნელობებით. ვთქვათ

$$\phi(x)=\phi(x;\alpha_1,\alpha_2,...,\alpha_m),$$

სადაც $\alpha_1,\alpha_2,...,\alpha_m$ –რაიმე პარამეტრებია. მაშინ $\phi(x)$– საუკეთესო ფუნქციის პოვნა დაიყვანება $\alpha_1,\alpha_2,...,\alpha_m$ პარამეტრთა შეფასებაზე, გამომდინარე ცდისეული მონაცემების გამოყენებით. ე.ი. გამოიკვეთა პარამეტრული შეფასების სტატისტიკური ამოცანა. და რა თქმა უნდა პარამეტრთა შეფასებანი აიღება მათი დასაშვებ მნიშვნელობათა არიდან.

"პარმეტრთა საუკეთესო შეფასება" შეიძლება გაგებულ იქნას სხვადასხვა აზრით. ჩვენ მომავალში ამ ცნებას დავუკავშირებთ უმცირეს კვადრატთა მეთოდს, რომელიც პრაქტიკაში როგორც ყოველთვის გვაძლევს დამაკმაყოფილებელ და ზოგჯერ საკმაოდ კარგ შედეგებს-შეფასებებს. რაც მთავარია მივყავართ უმარტივეს გამოთვლებამდე.

<u>**განსაზღვრება 1:**</u> $\alpha_1,\alpha_2,...,\alpha_m$ უცნობი პარამეტრები, რომლებიც ცხადად იძლევიან Y= $\phi(x;\alpha_1,\alpha_2,...,\alpha_m)$ დამოკიდებულებას , განისაზღვრებიან "საუკეთესოდ" <u>*უმცირეს კვადრატთა მათოდის აზრით*</u> , თუ y_i – ექსპერიმენტალური მნიშვნელობებისგან $\hat{y}_i=\phi(x_i,\alpha_1,\alpha_2,...,\alpha_m)$– მოდელური, გაგლუვებულ (რეგრესიული) მნიშვნელობათა გადახრების $|y_i-\hat{y}_i|$, **i**=$\overline{1,n}$, კვადრატების ჯამი მინიმალურია: ე.ი. მინიმალურია:

$$\varepsilon^2 = \sum_{i=1}^{n} [y_i - \phi(x_i, \alpha_1, \alpha_2, ..., \alpha_m)]^2$$

– სიდიდე.

თუ $\{x_i, y_i\}$–რეალიზაციებია, მაშინ

$$\varepsilon^2 = \varepsilon^2(\alpha_1, \alpha_2, ..., \alpha_m) \qquad (14.1)$$

და უმცირეს კვადრატთა მეთოდის ამოცანა მიიყვანება მრავალცვლადიანი ფუნქციის ექსტრემუმის წერტილის პოვნის ამოცანაზე. თუ $(\alpha_1^*, \alpha_2^*, ..., \alpha_m^*)$ **(14.1)** ფუნქციის მინიმუმის წერტილია:

$$\varepsilon^2(\alpha_1^*, \alpha_2^*, ..., \alpha_m^*) = \min_{\langle \alpha_1, \alpha_2, ..., \alpha_m \rangle} \sum_{i=1}^{n} [y_i - \phi(x_i, \alpha_1, \alpha_2, ..., \alpha_m)]^2 ,$$

მაშინ მას ეწოდება $(\alpha_1, \alpha_2, ..., \alpha_m)$ -პარამეტრთა უმცირეს კვადრატთა მეთოდის აზრით საუკეთესო შეფასება: განიხილება <u>შეფასება:</u>

$$\phi^*(x) = \phi(x; \alpha_1^*, \alpha_2^*, ..., \alpha_m^*) \qquad (14.2)$$

რომელსაც რეგრესიული წირი ეწოდება და ყოველი **X=x** რეალიზაციისას კეთდება **"პროგნოზი"**

Y სიდიდის მნიშვნელობისთვის:

$$Y = \phi^*(x). \qquad (14.3)$$

<u>განმარტება 2:</u> Y სიდიდის მნიშვნელობის პროგნოზი, როცა ცნობილია X=x მნიშვნელობა, განისაზღვრება (14.3) ტოლობით.

პროგნოზი საშუალებას იძლევა ცდის ჩაუტარებლად X=x მნიშვნელობაზე შევაფასოთ Y სიდიდის მნიშვნელობა, რომელიც საუკეთესო იქნება უმცირეს კვადრატთა მეთოდის აზრით.

უმცირეს კვადრატთა მეთოდის ამოცანის ამოხსნის ერთი რეალიზაცია

თუ ϕ ფუნქციას გააჩნია უწყვეტი პირველი რიგის კერძო წარმოებულები, $\frac{\partial \varepsilon^2}{\partial \alpha_k}$; $k = \overline{1, m}$

მაშინ გამოვიყენებთ რა ექსტრემუმის არსებობის აუცილებელ პირობას (ცხადია მინიმუმი არსებობს: $\varepsilon^2 \geq 0$), მაშინ მივიღებთ ექსტრემუმის წერტილთა პოვნის განტოლებათა სისტემას, (m განტოლება,m უცნობი):

$$\left\{ \frac{\partial \varepsilon^2}{\partial \alpha_k} = 0 \Leftrightarrow \sum_{i=1}^{n} [y_i - \phi(x_i, \alpha_1, \alpha_2, ..., \alpha_m)] \frac{\partial \phi(x_i, \alpha_1, \alpha_2, ..., \alpha_m)}{\partial \alpha_k} = 0 . \right. \qquad (14.4)$$

k= 1,2,…,m

თუ (14.4) სისტემის ამოხსნა $(\alpha_1^*, \alpha_2^*, ..., \alpha_m^*), \varepsilon^2$ –ფუნქციის მინიმუმის წერტილია, მაშინ იგი წარმოადგენს უმცირეს კვადრატთა მეთოდის ამოცანის ამოხსნას, ანუ $(\alpha_1, \alpha_2, ..., \alpha_m)$ პარამეტრთა შემფასებელს:

$$\alpha_k^* = \alpha_k(x_1y_1, x_2y_2, ..., x_ny_n), \quad k = \overline{1, m}. \tag{14.5}$$

რა თქმა უნდა საჭირო იქნება α_k^* –შემფასებელთა სტატისტიკური გამოკვლევა: ნდობის ინტერვალთა აგება, კორელაციური ანალიზი, სტატისტიკურ ჰიპოთეზთა შემოწმება. ყოველივე ეს ამოცანები გადაწყდება მოცემული **F** ფუნქციათა კლასის მაგალითებზე: წრფივი, კვადრატული და ა.შ. აქ ჩვენ განვიხილავთ მხოლოდ წრფივ შემთხვევას.

§2. წრფივი რეგრესიული ანალიზის ელემენტარული შეფასებანი.

განვიხილოთ F ფუნქციათა წრფივი კლასის მაგალითი. ვთქვათ $\phi(x) = kx + b$ – წრფივი რეგრესიაა: მაშინ

$$\varepsilon^2 = \sum_{i=1}^{n} (y_i - kx_i - b)^2. \quad \varepsilon^2 = \varepsilon^2(\mathrm{k,b}). \tag{14.6}$$

განვიხილოთ განტოლებათა სისტემა:

$$\begin{cases} \dfrac{\partial \varepsilon^2}{\partial k} = 0 \\ \\ \dfrac{\partial \varepsilon^2}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} -\sum_{i=1}^{n} (y_i - kx_i - b)x_i = 0 \\ \\ -\sum_{i=1}^{n} (y_i - kx_i - b) = 0, \end{cases} \tag{14.7}$$

საიდანაც ვღებულობთ ერთადერთ ამოხსნას:

$$k^* = \frac{n\sum_{i=1}^{n} x_iY_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\widetilde{R}_{xy}}{S_x^{'2}}$$

$$b^* = \bar{y} - k^*\bar{x}. \tag{14.8}$$

სადაც $\widetilde{R}_{xy} = \widetilde{R}_{xy} = \overline{xy} - \bar{x}\bar{y}$–შერჩევითი კოვარიაციაა **X** და **Y** სიდიდეებს შორის;

$S_x^2 = S_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\overline{x_n})^2$ - შერჩევითი დისპერსიაა

რადგან ε^2-ს გააჩნია მინიმუმი, ამიტომ $(\mathrm{k^*,b^*}) - \varepsilon^2$-ის ექსტრემუმის ერთადერთი წერტილი იქნება, ამიტომ ის მინიმუმის წერტილია:

$$\varepsilon^2(k^*, b^*) = \inf_{(k,b)\in R^2} \varepsilon^2(k, b).$$

ე.ი. **k*** და **b*** სტატისტიკები **k** და **b** პარამეტრების საუკეთესო შეფასებებია(უმცირეს კვადრატთა აზრით).
მაშინ $\phi(x) = kx + b$ – წრფივი რეგრესია საბოლოოდ მიიღებს ასეთ სახეს:

$$Y = \phi(X) \leftrightarrow Y = \frac{\widetilde{R}_{xy}}{S_x^2} X + \bar{y} - \frac{\widetilde{R}_{xy}}{S_x^2} \bar{x}$$

იქიდან, რომ **შერჩევითი კორელაციის კოეფიციენტი** განისაზღვრება როგორც

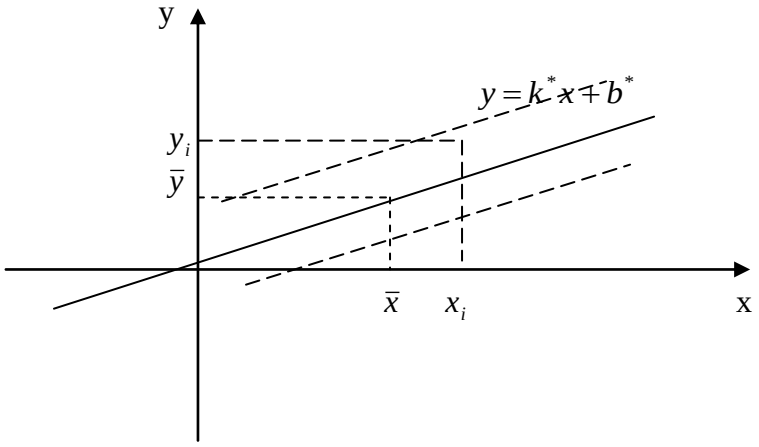
$$\widetilde{\rho}_{xy} = \frac{\widetilde{R}_{xy}}{S_x S_y} \ ,$$

მაშინ $Y=\phi(X)$ – **წრფივი რეგრესიის წირი მიიღებს სახეს:**

$$Y = \bar{y} + \widetilde{\rho}_{xy} \frac{S_y}{S_x} (X - \bar{x}). \qquad (14.9)$$

მიღებული წრფე ყველა y=kX+b წრფეებიდან ყველაზე სუკეთესოდ უმცირეს კვადრატთა მეთოდის აზრით აღწერს $Y \leftarrow X$ ***დამოკიდებულებას.***

ე.ი. ***წრფივი რეგრესიული წირის ასაგებად საჭიროა*** $\bar{y}, \bar{x}, \widetilde{\rho}_{xy}, S_x, S_y$ – ***შერჩევითი პარამეტრების გამოთვლა.*** შევნიშნოთ, **რომ წრფივი რეგრესიის წირი გადის** $(\bar{x}, \bar{y})$ **წერტილებზე:**



<u>შენიშვნა:</u> საკოორდინატო სიბრტყეზე (x_i, y_i), $i = 1,2,...,n$ წერტილთა სიმრავლეს მონაცემთა **გაბნევის დიაგრამა** ეწოდება.

§3 წრფივი რეგრესიის პარამეტრთა შეფასება ნდობის ინტერვალთა მეთოდით

იმისთვის, **რომ** k^* და b^* შემფასებლების საფუძველზე გავაკეთოთ დასკვნები k და b პარამეტრების შესახებ, ვისარგებლოთ დაშვებით, **რომ** Y**-ის განაწილება** ფიქსირებული x_j**-თვის არის ნორმალური.** შეიძლება ვაჩვენოთ, რომ k და b უცნობი პარამეტრების k^* და b^* ***შემფასებლები ნორმალურად იქნება განაწილებული (ფიქსირებული*** $(x_1, x_2, ..., x_n)$ ***რეალიზაციისთვის) შემდეგი პარამეტრებით:***

$$k^* \sim N\left(k, \frac{\sigma^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2}\right), \qquad b^* \sim N\left(b, \frac{1}{n} + \frac{(\bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \sigma^2\right).$$

ახლა k და b პარამეტრებისთვის $(1-\alpha)$**-** დონის ნდობის ინტერვალების აგების

შესახებ.:

1. *თუ* σ^2 *ცნობილია, მაშინ ნდობის ინტერვალთა აგების ცნობილი მიდგომის გამოყენებით,* $(1-\alpha)$ *დონის ნდობის ალბათობით გვექნება შემდეგი ნდობის ინტერვალები* k *და* b *პარამეტრებისთვის.*

$$P\left\{k^* - z_{\alpha/2}\left(\frac{\sigma^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}\right)^{1/2} \le k \le k^* + z_{\alpha/2}\left(\frac{\sigma^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}\right)^{1/2}\right\} = 1-\alpha,$$

$$P\left\{b^* - z_{\alpha/2}\left(\frac{1}{n}+\frac{(\bar{x})^2\sigma^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}\right)^{1/2} \le b \le b^* + z_{\alpha/2}\left(\frac{1}{n}+\frac{(\bar{x})^2\sigma^2}{\sum_{j=1}^{n}(x_j-\bar{x})^2}\right)^{1/2}\right\} = 1-\alpha.$$

2. თუ σ^2 არ არის ცნობილი, მაშინ აგებული ნდობის ინტერვალებში σ^2-ის ნაცვლად ჩაისმება მისი მიახლოება, ჩაუნაცვლებელი სტატისტიკა:

$$S_\sigma^{'2} = \frac{1}{n-2}\sum_{j=1}^{n}(y_j - k^* x_j - b^*)^2.$$

თუმცა არსებობს უფრო ზუსტი მიდგომაც და გვექნება შემდეგი ინტერვალები:

$$P\left\{k^* - t_{n-2,\alpha/2}\frac{S_\sigma^{'}}{\sqrt{\sum_{j=1}^{n}(x_j-\overline{x_n})^2}}\frac{1}{\sqrt{n-2}} \le k \le k^* + t_{n-2,\alpha/2}\frac{S_\sigma^{'}}{\sqrt{\sum_{j=1}^{n}(x_j-\overline{x_n})^2}}\frac{1}{\sqrt{n-2}}\right\} = 1-\alpha,$$

$$P\left\{b^* - t_{n-2,\alpha/2}\frac{S_\sigma^{'}(\overline{x_n})^{1/2}}{\sqrt{\sum_{j=1}^{n}(x_j-\overline{x_n})^2}}\frac{1}{\sqrt{n-2}} \le b \le b^* + t_{n-2,\alpha/2}\frac{S_\sigma^{'}(\overline{x_n})^{1/2}}{\sqrt{\sum_{j=1}^{n}(x_j-\overline{x_n})^2}}\frac{1}{\sqrt{n-2}}\right\} \approx 1-\alpha.$$

შევნიშნოთ, რომ აგებული ნდობის ინტერვალების ბაზაზე შეიძლება განხილული იყოს სტატისტიკური პარამეტრული ჰიპოთეზები b *და* k *(ასევე* σ*) პარამეტრების მიმართაც (აქ არ განვიხილავთ).*

§4 მაგალითების განხილვა წრფივი რეგრესიის გამოყენებაზე

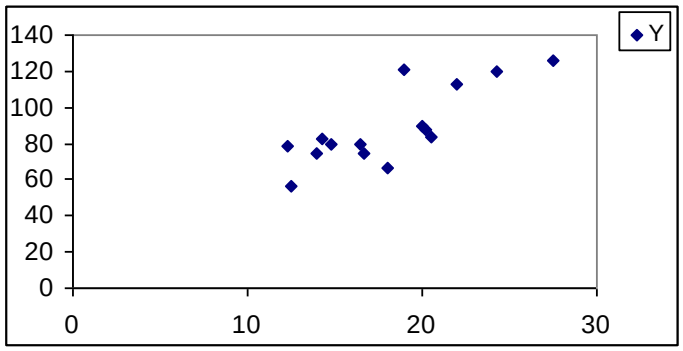
განვიხილოთ მაგალითი: უძრავი ქონების კომპანიის აგენტს სურს განჭვრიტოს ერთი ოჯახისათვის განკუთვნილი საცხოვრებელი სახლის გასაყიდი ფასი. გულმოდგინე განსჯის შემდეგ მან დაასკვნა, რომ სახლის გასაყიდი ფასი ყველაზე მჭიდროდ დაკავშირებულია მის ფართობთან. ამ ჰიპოთეზის შესამოწმებლად მან შემთხვევით შეარჩია ადრე გაყიდული 15 სახლის მონაცემები, რომლებიც თავმოყრილია ქვემომოყვანილ ცხრილში, სადაც გამოყენებულია

აღნიშვნები: X – სახლის ფართობი ($\times 10 მ^2$), Y – სახლის გასაყიდი ფასი (/1000), რადგანაც ჩვენს მიზანს შეადგენს სახლის ფასის (დამოკიდებული ცვლადი) პროგნოზირება მისი ფართობის (პრედიქტორის) მეშვეობით.

ცხრილი 1

x	Y	x	Y	x	Y
20.0	89.5	14.3	82.5	22.0	112.6
14.8	79.9	27.5	126.3	19.0	120.8
20.5	83.1	16.5	79.3	12.3	78.5
12.5	56.9	24.3	119.9	14.0	74.3
18.0	66.6	20.2	87.6	16.7	74.8

1. Chart Wizard – (XY Scatter) ფუნქციის გამოყენებით შევადგინოთ გაბნევის დიაგრამა



გაბნევის დიაგრამიდან ჩანს, ***Y* ცვლადის *X*-ისაგან დამოკიდებულებაში თავს იჩენს წრფივობის ტენდენცია და შესაძლებელია, რომ მარტივმა რეგრესიამ კარგად აღწეროს იგი. მაშასადამე, ამოცანის მათემატიკურ მოდელად შეიძლება ავირჩიოთ** (აქ $B_1 = k^*$ და $B_0 = b^*$)

$$Y_i = B_0 + B_1 x_i + \varepsilon_i ,$$

2. გამოვთვალოთ სიდიდეები: $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} Y_i$, $\sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} Y_i^2$, $\sum_{i=1}^{n} x_i Y_i$,

ამისათვის შევადგინოთ შემდეგი სახის ცხრილი:

ცხრილი 2.

n	x_i	Y_i	x_i^2	Y_i^2	$x_i\, Y_i$
1	**20.0**	**89.5**	**400.00**	**8010.25**	**1790.00**
2	**14.8**	**79.9**	**219.04**	**6384.01**	**1182.52**
3	**20.5**	**83.1**	**420.25**	**6905.61**	**1703.55**
4	**12.5**	**56.9**	**156.25**	**3237.61**	**711.25**
5	**18.0**	**66.6**	**324.00**	**4435.56**	**1198.80**
6	**14.3**	**82.5**	**204.49**	**6806.25**	**1179.75**
7	**27.5**	**126.3**	**756.25**	**15951.69**	**3473.25**
8	**16.5**	**79.3**	**272.25**	**6288.49**	**1308.45**
9	**24.3**	**119.9**	**590.49**	**14376.01**	**2913.57**
10	**20.2**	**87.6**	**408.04**	**7673.76**	**1769.52**

11	**22.0**	**112.6**	**484.00**	**12678.76**	**2477.20**
12	**19.0**	**120.8**	**361.00**	**14592.64**	**2295.20**
13	**12.3**	**78.5**	**151.29**	**6162.25**	**965.55**
14	**14.0**	**74.3**	**196.00**	**5520.49**	**1040.20**
15	**16.7**	**74.8**	**278.89**	**5595.04**	**1249.16**
Σ	**272.6**	**1332.6**	**5222.24**	**124618.42**	**25257.97**

გვექნება:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{272.6}{15} = 18.17, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1332.6}{15} = 88.84.$$

ცხრილში ჩატარებული გამოთვლების მეშვეობით ადვილად **გამოითვლება SS_X, SS_{XY} და SS_Y სიდიდეებიც, თუ ვისარგებლებთ გამარტივებული ფორმულებით:**

$$\boldsymbol{SS_X} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}, \quad \boldsymbol{SS_Y} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}, \quad \boldsymbol{SS_{XY}} = \sum x_i Y_i - \frac{\left(\sum x_i\right)\left(\sum Y_i\right)}{n}$$

მივიღებთ:

$$SS_X = 268.19, \quad SS_Y = 6230.24, \qquad SS_{XY} = 1040.18.$$

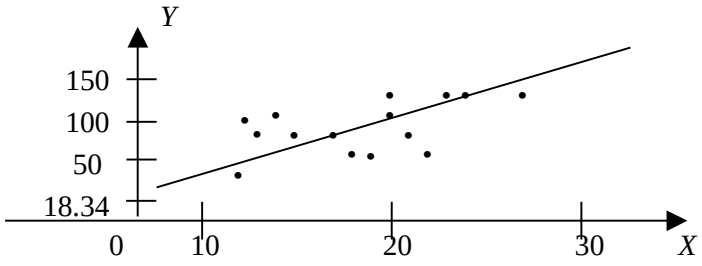
3. გამოვთვალოთ $B_0 = b^*$ და $B_1 = k^*$ შეფასებები (14) ფორმულებით. მივიღებთ:

$$\boldsymbol{b_1} = \frac{SS_{XY}}{SS_X} = \frac{1040.18}{268.19} = \mathbf{3.88}, \quad \boldsymbol{b_0} = \bar{Y} - \boldsymbol{b_1}\,\bar{x} = \mathbf{88.84 - 3.88 \cdot 18.17 = 18.34.}$$

ამრიგად **რეგრესიის წრფის შეფასებაა**:

$$\hat{Y} = \mathbf{18.34 + 3.88}x$$

მონაცემები გვაძლევს შემდეგ გრაფიკულ სურათს



ნახ. 14.1

კოეფიციენტების ინტერპრეტაცია. b_1-ის გამოთვლილი მნიშვნელობაა 3.88 ($1000-ში), რაც ნიშნავს, რომ *სახლის ფართობის ყოველი ერთი ერთეულით გაზრდა იწვევს სახლის ფასის საშუალოდ $3880-ით გაზრდას.* ამასთან b_0 =18.34. როგორც ადრე ვთქვით, b_0 არის ის წერტილი, რომელშიც რეგრესიის შეფასებული წრფე კვეთს Y ღერძს, ე.ი., როცა $x=0$, $\hat{Y}$ =18.34 ($1000-ში), ანუ ნულოვანი ფართობის სახლის საშუალო ფასია $1834, რაც აზრს მოკლებულია. *როგორც წესი, ჩვენ არ შეგვიძლია მივცეთ $\hat{Y}$ -ს ინტერპრეტაცია, როდესაც X ცვლადის მნიშვნელობა არ ეკუთვნის X-ის დაკვირვებულ $x_1,x_2,...,x_n$ მნიშვნელობათა გაბნევის დიაპაზონს,* ანუ როცა $x<x_{min}=\min(x_1,x_2,..,x_n)$ ან $x>x_{max}=\max(x_1,x_2,..,x_n)$. განხილულ მაგალითში $x_{min}=12.3$, $x_{max}=27.5$. რადგან $x=0$ არ ეკუთვნის [12.3;27.5] ინტერვალს, ამიტომ ჩვენ არ შეგვიძლია $\hat{Y}$ მნიშვნელობის საიმედო ინტერპრეტირება, როცა $x=0$.

4. გამოვთვალოთ *SSE* შემდეგი ფორმულით:

$$SSE = SS_Y - \frac{(SS_{XY})^2}{SS_X} = 6230{,}24 - \frac{(1040.18)^2}{268.19} = 21995.88,$$

საიდანაც მიიღება უცნობი დისპერსიის σ^2-ის σ^{2^*} შეფასება

$$\sigma^{2^*} = S^2 = \frac{SSE}{n-2} = \frac{2199{,}588}{13} = 168.91.$$

5. გამოვთვალოთ S_{b_0} და S_{b_1} შესაბამისი ვარიაციის კოეფიციენტების ფორმულებით:

$$S^2_{b_0} = 168{,}91 \cdot \frac{522{,}24}{15 \cdot 268{,}19} = 219.26, \quad S^2_{b_1} = 168{,}91 \cdot \frac{1}{268{,}19} = 0.794, \quad S_{b_1} = 0.794.$$

6. გამოვთვალოთ დეტერმინაციის კოეფიციენტი:

$$R^2 = 1 - \frac{SSE}{SST} = 0.648.$$

დეტერმინაციის კოეფიციენტის გამოთვლილი მნიშვნელობა გვიჩვენებს, რომ მარტივი რეგრესიით აიხსნება სახლის გასაყიდი ფასის ვარიაციის მხოლოდ 64.8%, დანარჩენი 35.2% აუხსნელი რჩება. საჭიროა შეირჩეს სხვა მოდელი.

7. თუ დავრჩებით ამ მოდელში, მაშინ ავაგოთ ნდობის ინტერვალი B_1-ისათვის.

ნდობის დონედ ავიღოთ 0.95. მაშინ ზემოდ ჩაწერილი ფორმულების თანახმად 0.95 ნდობის ალბათობით **აგებული ნდობის ინტერვალია**

[1.49, 6.27].

ანალოგიურად გამოითვლება B_0 **-ის ნდობის ინტერვალი.**

8. ავაგოთ ნდობის ინტერვალი $E(Y|X=x^*)$-ისათვის, $x^*=20$. ნდობის დონედ ავიღოთ $1-\alpha=0.95$, $\alpha/2=0.025$. ვისარგებლოთ ნდობის ინტერვალის ფორმულით პროგნოზის მნიშვნელობისთვის, რომელიც მისი გამოთვლის პარამეტრების ნდობის ინტერვალების მატივი, წრფივი გარდაქმნებით მიიღება. ადრე ჩატარებული გამოთვლებით, რადგან $b_0=18.34$, $b_1=3.88$, ამიტომ **რეგრესიის წრფის შეფასებაა**

$$\hat{Y}(x) = 18.34 + 3.88x$$

და

$$\hat{Y}(x^*) = \hat{Y}(20) = 18.34 + 3.88 \cdot 20 = 95.94.$$

ამრიგად $\hat{Y}(20) = 95.94$.

$\bar{x} = 18.17$, $SS_X = 268.19$, $t_{0,025,\ 13} = 2.16$

ახლა გამოვთვალოთ

$$t_{13,0.025} \cdot S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum x_i^2 - (\bar{x})^2}} = 2.160 \cdot 13 \sqrt{\frac{1}{15} + \frac{(20 - 18.17)^2}{268.19}} = 7.9$$

მაშასადამე 0.95 დონის ნდობის ინტერვალი იქნება

95.94 ± 7.9, ანუ $88.04 < E(Y|X=20) < 103.84$.

კითხვები თვითშემოწმებისათვის

1. ჩამოაყალიბე უმცირესი კვადრატების მეთოდის გამოყენების იდეა ორი სტატისტიკური პოპულაციის ფუნქციონალური დამოკიდებულების დასადგენად
2. რა არის რეგრესიის წირი და "შემთხვევითი" შეცდომა რეგრესიულ მოდელში
3. ჩამოაყალიბე რეგრესიული ანალიზის ამოცანა, როგორც პარამეტრული შეფასების მეთოდი ორი სტატისტიკური პოპულაციის ფუნქციონალური დამოკიდებულების დასადგენად
4. უმცირეს კვადრტთა მეთოდის აზრით საუკეთესო შემფასებელი და პროგნოზის ცნება რეგრესიულ ამოცანებში
5. რეგრესიული ამოცანის ერთი გადაწყვეტის შესახებ
6. წრფივი რეგრესია და უმცირეს კვადრატთა მეთოდი: რეგრესიის პარამეტრების შეფასება უმცირესი კვადრატებით
7. წრფივი რეგრესიის პარამეტრების შეფასება ნდობის ინტერვალების მეთოდით
8. ამოხსენი მაგალითები წრფივი რეგრესიის გამოყენებაზე EXCEL-ის ფუნქციებიის საშუალებით

§ 5. EXCEL-ის წრფივი რეგრესიის ამოცანები

მარტივ წრფივ რეგრესიულ მოდელთან დაკავშირებით განვიხილოთ **EXCEL** -ის შემდეგი სტატისტიკური ფუნქციები:

DEVSQ, INTERCEPT, LINEST, PSQ.

მოვიყვანოთ აღნიშნული სტატისტიკური ფუნქციების ჩაწერისას გამოყენებული აღნიშვნების ინტერპრეტაცია და შინაარსობრივი დატვირთვა
ინტერპრეტაცია და შინაარსობრივი დატვირთვა

აღნიშვნა	შინაარსი
Known y's	ცნობილი დამოკიდებული ცვლადის მნიშვნელობები
Known x's	ცნობილი დამოუკიდებული ცვლადის მნიშვნელობები
x	დამოუკიდებელი ცვლადის მნიშვნელობა
Const	**ამასთან თუ const=0, მაშინ იგულისხმება რეგრესიის წრფე გადის კოორდინატთა სათავეზე.**

DEVSQ ფუნქციის ჩაწერის სინტაქსი ასეთია:

DEVSQ (Number1,Nuvber2).

ფუნქცია **DEVSQ** გამოითვლის დაკვირვებული მნიშვნელობების მათი საშუალო მნიშვნელობიდან გადახრების კვადრატების ჯამს (14.7) ფორმულის საფუძველზე.

INTERCEPT ფუნქციის ჩაწერის სინტაქსი ასეთია:

INTERCEPT (Known y's, Known x's).

INTERCEPT ფუნქცია გამოითვლის, უმცირეს კვადრატთა მეთოდის საფუძველზე, B_0 კოეფიციენ-ტის b_0 შეფასებას (14.6) ფორმულით.

LINEST ფუნქციის ჩაწერის სინტაქსი ასეთია:

LINEST(Known y's, Known x's,const,Stats).

LINEST ფუნქცია გამოითვლის, უმცირეს კვადრატთა მეთოდის საფუძველზე, B_1 კოეფიციენტის b_1 შეფასებას.

FORECAST ფუნქციის ჩაწერის სინტაქსი ასეთია:

FORECAST (x,Known y's, Known x's).

FORECAST ფუნქცია გამოითვლის შეფასებული რეგრესიის წრფის საფუძველზე საპროგნოზო მნიშვნელობას.

PSQ ფუნქციის ჩაწერის სინტაქსი ასეთია:

PSQ (Known y's, Known x's,).

PSQ ფუნქცია გამოითვლის პირსონის კორელაციის კოეფიციენტის კვადრატს, დეტერმინაციის კოეფიციენტის მნიშვნელობას (14.19) ფორმულით.

სავარჯიშოები

შემდგომი ამოცანები შეგიძლიათ ამოხსენით Excel-ის გამოყენებით

14.1. გამოსარკვევია როგორაა დამოკიდებული შრომის ნაყოფიერება (Y) სამუშაოთა მექანიზაციის დონისაგან (X) 14 სამრეწველო საწარმოს მონაცემებით. სამუშაოთა მექანიზაციის დონე წარმოადგენს ამხსნელ ცვლადს. მონაცემები მოყვანილია ცხრილში

საწარმო	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X	20	24	28	30	31	33	34	37	38	40	41	43	45	48
Y	32	30	36	40	41	47	56	54	60	55	61	67	69	76

ა) ააგეთ ნდობის ინტერვალი $E(Y \mid x^*)$-თვის, როცა $x^* = 40$.
ბ) ააგეთ ნდობის ინტერვალი $Y(x^*)$-ისათვის, როცა $x^* = 40$.

14.2. ფირმებმა, რომელთა ბიზნესია სწრაფი ბეჭდვა-გამრავლება, რეკლამისათვის გამოყოფილი თანხა (X) დახარჯეს ქალაქის ავტობუსების გაჩერებებზე განცხადებების გამოკვრაში. ამოცანა მდგომარეობს გავარკვიოთ რამდენად ეფექტური იყო ასეთი რეკლამა, როგორ იმოქმედა მან პროდუქციის გასაღების დონეზე (Y). მონაცემები მოცემულია ცხრილში:

ფირმები	A	B	C	D	E
X (ათასი)	2	4	5	7	3
Y (ათასი)	10	40	30	50	20

შეაფასეთ თვიური გაყიდვა, რომელიც $4500 სარეკლამო ხარჯებით არის გამოწვეული.

14.3. აგრონომი, რომელიც მუშაობს „აგრიკო“-ში აწარმოებს ექსპერიმენტებს თხევადი სასუქით, ტოლი სიდიდის მიწის ნაკვეთებზე. (X) სასუქის რაოდენობა (ტონა), (Y) შემოსავალი (ასი ბუშელი). მონაცემები მოცემულია ცხრილში:

ნაკვეთი	A	B	C	D
X	2	1	3	4
Y	7	3	8	10

შეაფასეთ, თუ რამდენი მოსავალი მოვა 3 ტონა სასუქის გამოყენებით.

14.4. კბილის პასტის გაყიდვები, როგორც ჩანს, ძლიერადაა დამოკიდებული რეკლამის დონეზე. რეკლამის წლიურ ხარჯებსა და გაყიდვებს შორის დამოკიდებულება სხვადასხვა პასტის მწარმოებლისათვის მოცემულია ცხრილში:

	A	B	C	D
წლიური ხარჯი რეკლამაზე (მილიონი /)	2	4	3	1
წლიური გაყიდვები (მილიონი /)	50	70	60	20

ა) საშუალოდ რა მოცულობის გაყიდვებია მოსალოდნელი, თუ რეკლამაზე გაწეულია $1.9 მილიონის ხარჯები?

ბ) რას უდრის 95% ნდობის ინტერვალი პასტების საშუალო გაყიდვებისათვის, იმ ჯგუფისათვის, რომელთა რეკლამაზე დაიხარჯა $1.9 მილიონი?

14.5. შემდეგ ცხრილში წარმოდგენილია ამერიკის შეერთებულ შტატების (X) პირადი წმინდა შემოსავლებისა და (Y) სამომხმარებლო ხარჯების (მილიარდ $-ში) მონაცემები 1964-1968 წლებში

X	401 537	433	466	492
Y	438 520	473	512	547

14.6. სხვადასხვა ფინანსური ინსტიტუტის მიერ დადგენილი საპროცენტო განაკვეთები შესაძლოა ასახავდეს იმას, თუ რამდენად მკაცრადაა დაცული მათ მიერ გაცემული სესხების შეფასების სტანდარტები: რაც უფრო დაბალია საპროცენტო განაკვეთი (X), მით უფრო მაღალია სტანდარტები (და, მაშასადამე, უფრო დაბალია სესხების დაუბრუნებლობის (დეფოლტების) (Y) განაკვეთი. ამ თვალსაზრისის შემოწმების მიზნით შეგროვებულია შემთხვევით 9 ფინანსური კომპანიის მონაცემები, რომლებიც წარმოდგენილია ცხრილით:

X	7,0	6,6	6,0	8,5	8,0	7,5	6,5	7,0	8,0
Y	38	40	35	46	48	39	36	37	44

მოძებნეთ 95%-იანი საპროგნოზო ინტერვალი დეფოლტების განაკვეთისათვის, როდესაც საპროცენტო განაკვეთი 8%-ის ტოლია.

14.7. ქვემომოყვანილი მონაცემები წარმოადგენს დაკვირვებებს კვების პროდუქტებზე საცალო ფასების ინდექსისა (Y) და სამრეწველო წარმოების ინდექსის (X) მნიშვნელობებზე 10 წლის მანძილზე.

წელი	1	2	3	4	5	6	7	8	9	10
Y	100	101	113	115	113	113	111	112	115	120

X	64	75	81	84	91	85	96	99	100	93

ააგეთ 95%-იანი ნდობის ინტერვალი ფასების ინდექსის საშუალო მნიშვნელობისათვის და 99%-იანი საპროგნოზო ინტერვალი ფასების ინდექსისათვის, როდესაც სამრეწველო ინდექსის მნიშვნელობაა x^*=98.

14.8. ქვემომოყვანილი მონაცემები წარმოადგენს ჯამური სამომხმარებლო კრედიტისა ($ მილიონობით) (X) და ფინანსურ და სამრეწველო გაკოტრებათა რაოდენობის (ათასობით) (Y) დაკვირვებულ მნიშვნელობებს გარკვეული 10 წლის განმავლობაში

წელი	1	2	3	4	5	6	7	8	9	10
Y	21	23	28	31	32	39	43	45	46	52
X	92	80	76	89	111	110	127	137	150	141

დანართი (ცხრილები)

A0. EXCEL-ის სტატისტიკური ფუნქციები

A1. სტანდარტული ნორმალური განაწილების ფუნქციის მნიშვნელობები

A2. სტანდარტული ნორმალური განაწილების ზედა α-კრიტიკული წერტილები (z_α)

A3. χ^2 განაწილების ზედა α-კრიტიკული წერტილები $(\chi^2_{n,\alpha})$

A4. *t* განაწილების ზედა α-კრიტიკული წერტილები $(t_{n,\alpha})$

A0. EXCEL-ის სტატისტიკური ფუნქციები

AVERAGE(x_1:x_n)	$\frac{1}{n}\sum_{i=1}^{n} x_i$
BINOMDIST(k,n,p,0)	$P_n\{S_n = k\} = C_n^k\ p^k (1-p)^{n-k}$
BINOMDIST(k,n,p,1)	$P\{S_n \le k\} = \sum_{i=0}^{k} C_n^i\ p^I (1-p)^{n-I}$
CHIDIST(x,n)	$P\{\chi^2(n) > x\}$
CHIINV(p,n)	$P\{\chi^2(n) > x\} = p \rightarrow x$
COMBIN(n,r)	$C_n^r = \frac{A_n^r}{r!} = \frac{n!}{r!(n-r)!}$
CONFIDENCE(α,σ,n)	$z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$
CORREL(x_1:x_n, y_1:y_n)	$\rho_{x,y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$
COUNT(x_1:x_n)	n – მონაცემთა რაოდენობა
COVAR(x_1:x_n, y_1:y_n)	$\mathrm{cov(X,Y)} = \frac{1}{n}\sum (x_i - \mu_X)(y_i - \mu_Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$
CRITBINOM(n,p,α)	$\alpha = P\{S_n \le k\} = \sum_{i=0}^{k} C_n^i\ p^i (1-p)^{n-i} \rightarrow k$
DEVSQ(x_1:x_n)	$\sum_{i=1}^{n} (x_i - \bar{x})^2$
EXPONDIST(x,λ,0)	$f(x;\lambda) = \lambda e^{-\lambda x}$
EXPONDIST(x,λ,1)	$F(x;\lambda) = 1 - e^{-\lambda x}$
FACT(n)	$n!$
FDIST(x,n_1,n_2)	$P\{\mathrm{F}(n_1,n_2) > x\}$
FINV(α,n_1,n_2)	$\alpha = P\{\mathrm{F}(n_1,n_2) > x\} \rightarrow x$
FORECAST(x,y_1:y_n,x_1:x_n)	$y = b_0 + b_1 x$
FREQUENCY(x_1:x_n,k)	მონაცემთა რაოდენობა რომლებიც არ აღემატება k-ს
FTEST(x_1:x_n, y_1:y_m)	$P\left\{F_{n_x,n_y} > s_x'^2 / s_y'^2\right\} + P\left\{F_{n_x,n_y} < \frac{1}{s_x'^2 / s_y'^2}\right\}$
HYPGEOMDIST(k,n,A,N)	$P(k,n,A,N) = \frac{C_n^k \cdot C_{N-n}^{A-k}}{C_N^A}$
INTERCEPT(y_1:y_n,x_1:x_n)	b_0-ის შეფასება წრფივი რეგრესიისათვის
KURT(x_1:x_n)	$\frac{\sum (x-\bar{x})^3}{ns^3}$, ასიმეტრიის კოეფიციენტი

LARGE(x_1:x_n,k)	უდიდესი k-ური წევრი
LINEST(y_1:y_n,x_1:x_n,1,)	b_1 შეფასება, როდესაც მოდელში b_0 ნულის ტოლია
LINEST(y_1:y_n,x_1:x_n,0,)	b_1 შეფასება, როდესაც მოდელში b_0 არ უდრის ნულს
MAX(x_1:x_n)	$\max(x_1;x_2;...;x_n)$
MEDIAN(x_1:x_n)	$\tilde{x}$ =median(x_1;x_2;...;x_n)
MIN(x_1:x_n)	$\min(x_1;x_2;...;x_n)$
MODE(x_1:x_n)	მოდა
NORMDIST(x,μ,σ,0)	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
NORMDIST(x,μ,σ,1)	$F(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$
NORMINV(p,μ,σ)	$p=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow x$
NORMSDIST(x)	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$
NORMSINV(p)	$p=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt \rightarrow x$
PERCENTILE(x_1:x_n,p)	$P=p\cdot100$ რიგის პროცენტილი ($0\le p\le1$)
PERMUT(n,r)	$A_n^r=\frac{n!}{(n-r)!}$
POISSON(k,λ,0)	$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$
POISSON(k,λ,1)	$P\{X\le k\}=\sum_{i=1}^{k}\frac{\lambda^i}{i!}e^{-\lambda}$
PROB(x_1:x_n, p_1:p_n,a,b)	$P\{a\le x\le b\}$
QUARTILE(x_1:x_n,k)	k=0 Q_0=min(x_1;x_2;...;x_n) k=1 Q_1=პირველი კვარტილი (25 პროცენტილი) k=2 Q_2=მედიანა, მეორე კვარტილი (50 პროცენტილი) k=3 Q_3=მესამე კვარტილი (75 პროცენტილი) k=4 Q_4=max(x_1;x_2;...;x_n)
RSQ(y_1:y_n,x_1:x_n)	R^2
SKEW(x_1:x_n)	$\frac{\sum(x-\bar{x})^4}{ns^4}-3$, ექსცესის კოეფიციენტი
SLOPE(y_1:y_n,x_1:x_n)	b_1-ის შეფასება წრფივი რეგრესიისათვის

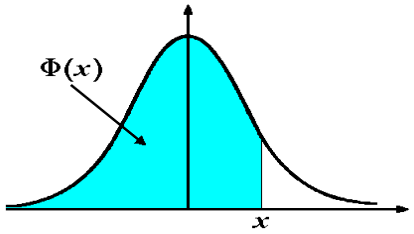
SMALL(x_1:x_n,k)	უმცირესი k-ური წევრი
STANDARDIZE(x,μ,σ)	$\frac{x-\mu}{\sigma}$
STDEV(x_1:x_n)	s'
STDEVP(x_1:x_n)	s
SUMPRODUCT(x_1:x_n, p_1:p_n)	$EX=\sum_{i=1}^{n} x_i\, p_i$
TDIST(x,n,1)	$P\{\mathrm{T}(n)>x\}$
TDIST(x,n,2)	$P\{\lvert\mathrm{T}(n)\rvert>x\}$
TINV(p,n)	$p= P\{\lvert\mathrm{T}(n)\rvert>x\} \rightarrow x$
TREND(y_1:y_n,x_1:x_n,x,0)	$\hat{Y} =b_1x$
TREND(y_1:y_n,x_1:x_n,x,1)	$\hat{Y} =b_1x+b_0$
TTEST(x_1:x_n,y_1:y_n,k,1)	k=1, $\mathrm{P}\{T_{n,n}>t\}$ დაწყვილებული k=2, $\mathrm{P}\{\lvert T_{n,n}\rvert>t\}$ მონაცემებისათვის
TTEST(x_1:x_n,y_1:y_m,k,2)	k=1, $\mathrm{P}\{T_{n,m}>t\}$ უცნობი, ტოლი k=2, $\mathrm{P}\{\lvert T_{n,m}\rvert>t\}$ დისპერსიების შემთხვევაში
TTEST(x_1:x_n,y_1:y_m,k,3)	k=1, $\mathrm{P}\{T_{n,n}>t\}$ უცნობი, არატოლი k=2, $\mathrm{P}\{\lvert T_{n,n}\rvert>t\}$ დისპერსიების შემთხვევაში
VAR(x_1:x_n)	$s'^2_n=\frac{1}{n-1}\sum(x_i-\bar{x})^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n(n-1)}$
VARP(x_1:x_n)	$s_n^2=\frac{1}{n}\sum(x_i-\bar{x})^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n^2}$
ZTEST(x_1:x_n,μ_0,σ)	$P\left\{Z>\frac{\bar{X}_n-\mu_0}{\sigma/\sqrt{n}}\right\}$, სადაც $Z\sim N(0,1)$

Tools/Data Analysis-ის გამოყენებული სტატისტიკური ფუნქციები

Anova: Single Factor,
Correlation
Covariance
Descriptive Statistics
F-Test Two-Sample for Variances
Histogram
Random Number Generation
Regression
t-Test: Paired Two Sample for Means
t-Test: Two Sample Assuming Equal Variances
t-Test: Two Sample Assuming Unequal Variances
z-Test: Two Sample for Means

A1. სტანდარტული ნორმალური განაწილების ფუნქციის მნიშვნელობები.

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

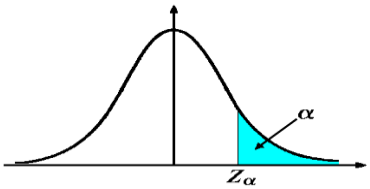


x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0.00	0.500	**0.33**	0.629	**0.66**	0.745	**0.99**	0.838	**1.32**	0.906	**1.65**	0.950
0.01	0.503	**0.34**	0.633	**0.67**	0.748	**1.00**	0.841	**1.33**	0.908	**1.66**	0.951
0.02	0.507	**0.35**	0.636	**0.68**	0.751	**1.01**	0.843	**1.34**	0.909	**1.67**	0.952
0.03	0.511	**0.36**	0.640	**0.69**	0.754	**1.02**	0.846	**1.35**	0.911	**1.68**	0.953
0.04	0.515	**0.37**	0.644	**0.70**	0.758	**1.03**	0.848	**1.36**	0.913	**1.69**	0.954
0.05	0.519	**0.38**	0.648	**0.71**	0.761	**1.04**	0.850	**1.37**	0.914	**1.70**	0.955
0.06	0.523	**0.39**	0.651	**0.72**	0.764	**1.05**	0.853	**1.38**	0.916	**1.71**	0.956
0.07	0.527	**0.40**	0.655	**0.73**	0.767	**1.06**	0.855	**1.39**	0.917	**1.72**	0.957
0.08	0.531	**0.41**	0.659	**0.74**	0.770	**1.07**	0.857	**1.40**	0.919	**1.73**	0.958
0.09	0.535	**0.42**	0.662	**0.75**	0.773	**1.08**	0.859	**1.41**	0.920	**1.74**	0.959
0.10	0.539	**0.43**	0.666	**0.76**	0.776	**1.09**	0.862	**1.42**	0.922	**1.75**	0.959
0.11	0.543	**0.44**	0.670	**0.77**	0.779	**1.10**	0.864	**1.43**	0.923	**1.76**	0.960
0.12	0.547	**0.45**	0.673	**0.78**	0.782	**1.11**	0.866	**1.44**	0.925	**1.77**	0.961
0.13	0.551	**0.46**	0.677	**0.79**	0.785	**1.12**	0.868	**1.45**	0.926	**1.78**	0.962
0.14	0.555	**0.47**	0.680	**0.80**	0.788	**1.13**	0.870	**1.46**	0.927	**1.79**	0.963
0.15	0.559	**0.48**	0.684	**0.81**	0.791	**1.14**	0.872	**1.47**	0.929	**1.80**	0.964
0.16	0.563	**0.49**	0.687	**0.82**	0.793	**1.15**	0.874	**1.48**	0.930	**1.81**	0.964
0.17	0.567	**0.50**	0.691	**0.83**	0.796	**1.16**	0.876	**1.49**	0.931	**1.82**	0.965
0.18	0.571	**0.51**	0.694	**0.84**	0.799	**1.17**	0.879	**1.50**	0.933	**1.83**	0.966
0.19	0.575	**0.52**	0.698	**0.85**	0.802	**1.18**	0.881	**1.51**	0.934	**1.84**	0.967
0.20	0.579	**0.53**	0.701	**0.86**	0.805	**1.19**	0.882	**1.52**	0.935	**1.85**	0.967
0.21	0.583	**0.54**	0.705	**0.87**	0.807	**1.20**	0.884	**1.53**	0.936	**1.86**	0.968
0.22	0.587	**0.55**	0.708	**0.88**	0.810	**1.21**	0.886	**1.54**	0.938	**1.87**	0.969
0.23	0.590	**0.56**	0.712	**0.89**	0.813	**1.22**	0.888	**1.55**	0.939	**1.88**	0.969
0.24	0.594	**0.57**	0.715	**0.90**	0.815	**1.23**	0.890	**1.56**	0.940	**1.89**	0.970
0.25	0.598	**0.58**	0.719	**0.91**	0.818	**1.24**	0.892	**1.57**	0.941	**1.90**	0.971
0.26	0.602	**0.59**	0.722	**0.92**	0.821	**1.25**	0.894	**1.58**	0.942	**1.91**	0.971
0.27	0.606	**0.60**	0.725	**0.93**	0.823	**1.26**	0.896	**1.59**	0.944	**1.92**	0.972
0.28	0.610	**0.61**	0.729	**0.94**	0.826	**1.27**	0.897	**1.60**	0.945	**1.93**	0.973
0.29	0.614	**0.62**	0.732	**0.95**	0.828	**1.28**	0.899	**1.61**	0.946	**1.94**	0.973
0.30	0.617	**0.63**	0.735	**0.96**	0.831	**1.29**	0.901	**1.62**	0.947	**1.95**	0.974
0.31	0.621	**0.64**	0.738	**0.97**	0.833	**1.30**	0.903	**1.63**	0.948	**1.96**	0.975
0.32	0.625	**0.65**	0.742	**0.98**	0.836	**1.31**	0.904	**1.64**	0.949	**1.97**	0.975

A1. (გაგრძელება)

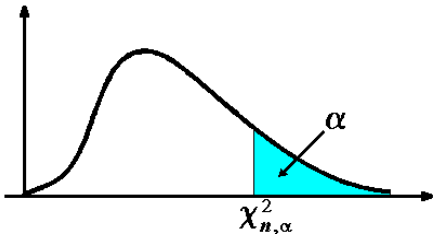
x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1.98	0.976	**2.26**	0.988	**2.54**	0.994	**2.82**	0.997	**3.10**	0.999
1.99	0.976	**2.27**	0.988	**2.55**	0.994	**2.83**	0.997	**3.11**	0.999
2.00	0.977	**2.28**	0.988	**2.56**	0.994	**2.84**	0.997	**3.12**	0.999
2.01	0.977	**2.29**	0.988	**2.57**	0.994	**2.85**	0.997	**3.13**	0.999
2.02	0.978	**2.30**	0.989	**2.58**	0.995	**2.86**	0.997	**3.14**	0.999
2.03	0.978	**2.31**	0.989	**2.59**	0.995	**2.87**	0.997	**3.15**	0.999
2.04	0.979	**2.32**	0.989	**2.60**	0.995	**2.88**	0.998	**3.16**	0.999
2.05	0.979	**2.33**	0.990	**2.61**	0.995	**2.89**	0.998	**3.17**	0.999
2.06	0.980	**2.34**	0.990	**2.62**	0.995	**2.90**	0.998	**3.18**	0.999
2.07	0.980	**2.35**	0.990	**2.63**	0.995	**2.91**	0.998	**3.19**	0.999
2.08	0.981	**2.36**	0.990	**2.64**	0.995	**2.92**	0.998	**3.20**	0.999
2.09	0.981	**2.37**	0.991	**2.65**	0.995	**2.93**	0.998	**3.21**	0.999
2.10	0.982	**2.38**	0.991	**2.66**	0.996	**2.94**	0.998	**3.22**	0.999
2.11	0.982	**2.39**	0.991	**2.67**	0.996	**2.95**	0.998	**3.23**	0.999
2.12	0.983	**2.40**	0.991	**2.68**	0.996	**2.96**	0.998	**3.24**	0.999
2.13	0.983	**2.41**	0.992	**2.69**	0.996	**2.97**	0.998	**3.25**	0.999
2.14	0.983	**2.42**	0.992	**2.70**	0.996	**2.98**	0.998	**3.26**	0.999
2.15	0.984	**2.43**	0.992	**2.71**	0.996	**2.99**	0.998	**3.27**	0.999
2.16	0.984	**2.44**	0.992	**2.72**	0.996	**3.00**	0.998	**3.28**	0.999
2.17	0.985	**2.45**	0.992	**2.73**	0.996	**3.01**	0.998	**3.29**	0.999
2.18	0.985	**2.46**	0.993	**2.74**	0.996	**3.02**	0.998	**3.30**	0.999
2.19	0.985	**2.47**	0.993	**2.75**	0.997	**3.03**	0.998	**3.31**	0.999
2.20	0.986	**2.48**	0.993	**2.76**	0.997	**3.04**	0.998	**3.32**	0.999
2.21	0.986	**2.49**	0.993	**2.77**	0.997	**3.05**	0.998	**3.33**	0.999
2.22	0.986	**2.50**	0.993	**2.78**	0.997	**3.06**	0.998	**3.34**	0.999
2.23	0.987	**2.51**	0.993	**2.79**	0.997	**3.07**	0.998	**3.35**	0.999
2.24	0.987	**2.52**	0.994	**2.80**	0.997	**3.08**	0.998	**3.36**	0.999
2.25	0.987	**2.53**	0.994	**2.81**	0.997	**3.09**	0.999	**3.37**	0.999

A2. სტანდარტული ნორმალური განაწილების ზედა α-კრიტიკული წერტილების, z_α-ს მნიშვნელობები



α	0.1	0.05	0.025	0.125	0.01	0.005	0.0025	0.001
z_α	1.28	1.64	1.96	2.24	2.33	2.57	2.81	3.08

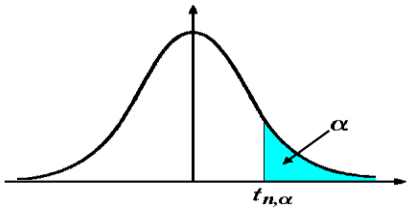
A3. χ^2 **განაწილების ზედა α-კრიტიკული წერტილები,** $\chi^2_{n,\alpha}$.



	α							
n	**0.99**	**0.975**	**0.95**	**0.9**	**0.1**	**0.05**	**0.025**	**0.01**
1	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2104
3	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449
4	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767
5	0.5543	0.8312	1.1455	1.6103	9.2363	11.0705	12.8325	15.0863
6	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119
7	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753
8	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902
9	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660
10	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093
11	3.0535	3.8157	4.5748	5.5778	17.2750	19.6752	21.9200	24.7250
12	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170
13	4.1069	5.0087	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882
14	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412
15	5.2294	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5780
16	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8453	31.9999
17	6.4077	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087
18	7.0149	8.2307	9.3904	10.8649	25.9894	28.8693	31.5264	34.8052
19	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1908
20	8.2604	9.5908	10.850	12.442	28.412	31.410	34.169	37.566

			8	6	0	4	6	3
21	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322
22	9.5425	10.9823	12.3380	14.0415	30.8133	33.9245	36.7807	40.2894
23	10.1957	11.6885	13.0905	14.8480	32.0069	35.1725	38.0756	41.6383
24	10.8563	12.4011	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798
25	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3140
26	12.1982	13.8439	15.3792	17.2919	35.5632	38.8851	41.9231	45.6416
27	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9628
28	13.5647	15.3079	16.9279	18.9392	37.9159	41.3372	44.4608	48.2782
29	14.2564	16.0471	17.7084	19.7677	39.0875	42.5569	45.7223	49.5878
30	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922

A4. ***t*** **განაწილების ზედა α-კრიტიკული წერტილები,** $t_{n,\alpha}$.



	α						
n	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0025**	**0.001**
1	3.078	6.314	12.706	31.821	63.656	127.321	318.289
2	1.886	2.920	4.303	6.965	9.925	14.089	22.328
3	1.638	2.353	3.182	4.541	5.841	7.453	10.214
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	1.476	2.015	2.571	3.365	4.032	4.773	5.894
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385

ს ა რ ჩ ე ვ ი

ლ ი ტ ე რ ა ტ უ რ ა

[1] გ. სირბილაძე, ლექციათა კურსი “მონაცემთა ანალიზი და სტატისტიკა. 2007.

[2] გ. მარი, ა. მოსიძე, სტატისტიკა. სახელმძღვანელო ESM – თბილისის სტუდენტებისთვის. თბილისი, 2005.

[3] James T. McClave, Terry Sincich, First Course in Statistics, A (9th Edition), ISBN 0131499793, 2005.

[4] В.Е. Гмурман. Руководство к решению задач по теории вероятностей и математической статистике. М., Высшая школа, 1998.