

Statistika

PREDAVANJA 10

Univerzitet u Novom Sadu, FTN

U praksi je često potrebno doneti zaključke koji važe za veću grupu ljudi ili objekata. Umesto da ispitujemo celu grupu, koju nazivamo **populacija**, posmatraćemo samo jedan njen podskup koji nazivamo **uzorak**.

Obeležje je numerička karakteristika X elemenata ω populacije Ω . U terminima verovatnoće obeležju odgovara slučajna promenljiva. Na slučajan način se bira n elemenata $\omega_1, \omega_2, \dots, \omega_n$ i registruju vrednosti obeležja $x_1 = X(\omega_1), x_2 = X(\omega_2), \dots, x_n = X(\omega_n)$.

U praksi je često potrebno doneti zaključke koji važe za veću grupu ljudi ili objekata. Umesto da ispitujemo celu grupu, koju nazivamo **populacija**, posmatraćemo samo jedan njen podskup koji nazivamo **uzorak**.

Obeležje je numerička karakteristika X elemenata ω populacije Ω . U terminima verovatnoće obeležju odgovara slučajna promenljiva. Na slučajan način se bira n elemenata $\omega_1, \omega_2, \dots, \omega_n$ i registruju vrednosti obeležja $x_1 = X(\omega_1), x_2 = X(\omega_2), \dots, x_n = X(\omega_n)$.

U praksi je često potrebno doneti zaključke koji važe za veću grupu ljudi ili objekata. Umesto da ispitujemo celu grupu, koju nazivamo **populacija**, posmatraćemo samo jedan njen podskup koji nazivamo **uzorak**.

Obeležje je numerička karakteristika X elemenata ω populacije Ω . U terminima verovatnoće obeležju odgovara slučajna promenljiva. Na slučajan način se bira n elemenata $\omega_1, \omega_2, \dots, \omega_n$ i registruju vrednosti obeležja $x_1 = X(\omega_1), x_2 = X(\omega_2), \dots, x_n = X(\omega_n)$.

U praksi je često potrebno doneti zaključke koji važe za veću grupu ljudi ili objekata. Umesto da ispitujemo celu grupu, koju nazivamo **populacija**, posmatraćemo samo jedan njen podskup koji nazivamo **uzorak**.

Obeležje je numerička karakteristika X elemenata ω populacije Ω . U terminima verovatnoće obeležju odgovara slučajna promenljiva. Na slučajan način se bira n elemenata $\omega_1, \omega_2, \dots, \omega_n$ i registruju vrednosti obeležja $x_1 = X(\omega_1), x_2 = X(\omega_2), \dots, x_n = X(\omega_n)$.

Kažemo da je n -dimenzionalna slučajna promenljiva $(X_1, X_2, \dots, X_n)$ **prost slučajni uzorak obima n** , ako su slučajne promenljive $X_1, X_2, \dots, X_n$ nezavisne i imaju istu raspodelu kao i posmatrano obeležje X (tj. $F_{X_1} = F_{X_2} = \dots = F_{X_n} = F$), a n -torku $(x_1, x_2, \dots, x_n)$ nazivamo **realizovana vrednost** n -dimenzionalne slučajne promenljive $(X_1, X_2, \dots, X_n)$.

Funkcija raspodele prostog slučajnog uzorka za $(t_1, t_2, \dots, t_n) \in \mathbb{R}^n$ je

$$F_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = F(t_1)F(t_2) \cdots F(t_n).$$

Statistika je funkcija uzorka $Y = g(X_1, X_2, \dots, X_n)$.

Kažemo da je n -dimenzionalna slučajna promenljiva $(X_1, X_2, \dots, X_n)$ **prost slučajni uzorak obima n** , ako su slučajne promenljive $X_1, X_2, \dots, X_n$ nezavisne i imaju istu raspodelu kao i posmatrano obeležje X (tj. $F_{X_1} = F_{X_2} = \dots = F_{X_n} = F$), a n -torku $(x_1, x_2, \dots, x_n)$ nazivamo **realizovana vrednost** n -dimenzionalne slučajne promenljive $(X_1, X_2, \dots, X_n)$. Funkcija raspodele prostog slučajnog uzorka za $(t_1, t_2, \dots, t_n) \in \mathbb{R}^n$ je

$$F_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = F(t_1)F(t_2) \cdots F(t_n).$$

Statistika je funkcija uzorka $Y = g(X_1, X_2, \dots, X_n)$.

Kažemo da je n -dimenzionalna slučajna promenljiva $(X_1, X_2, \dots, X_n)$ **prost slučajni uzorak obima n** , ako su slučajne promenljive $X_1, X_2, \dots, X_n$ nezavisne i imaju istu raspodelu kao i posmatrano obeležje X (tj. $F_{X_1} = F_{X_2} = \dots = F_{X_n} = F$), a n -torku $(x_1, x_2, \dots, x_n)$ nazivamo **realizovana vrednost** n -dimenzionalne slučajne promenljive $(X_1, X_2, \dots, X_n)$. Funkcija raspodele prostog slučajnog uzorka za $(t_1, t_2, \dots, t_n) \in \mathbb{R}^n$ je

$$F_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) = F(t_1)F(t_2) \cdots F(t_n).$$

Statistika je funkcija uzorka $Y = g(X_1, X_2, \dots, X_n)$.

Aritmetička sredina uzorka ($X_1, X_2, \dots, X_n$) je slučajna promenljiva $\bar{X}_n$ definisana sa

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

a njenu realizovanu vrednost obeležavamo malim slovom:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n).$$

Uzoračka disperzija ($X_1, X_2, \dots, X_n$) je slučajna promenljiva $\bar{S}_n^2$ definisana sa

$$\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2,$$

a njenu realizovanu vrednost obeležavamo malim slovom:

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 = \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - \bar{x}_n^2.$$

Aritmetička sredina uzorka ($X_1, X_2, \dots, X_n$) je slučajna promenljiva $\bar{X}_n$ definisana sa

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

a njenu realizovanu vrednost obeležavamo malim slovom:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n).$$

Uzoračka disperzija uzorka ($X_1, X_2, \dots, X_n$) je slučajna promenljiva $\bar{S}_n^2$ definisana sa

$$\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2,$$

a njenu realizovanu vrednost obeležavamo malim slovom:

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 = \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - \bar{x}_n^2.$$

Korigovana uzoračka disperzija $\hat{S}_n^2$ uzorka $(X_1, X_2, \dots, X_n)$ se definiše

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{S}_n^2,$$

korigovana uzoračka disperzija realizovanog uzorka $(x_1, x_2, \dots, x_n)$

je

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

Korigovana uzoračka disperzija $\hat{S}_n^2$ uzorka $(X_1, X_2, \dots, X_n)$ se definiše

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{S}_n^2,$$

korigovana uzoračka disperzija realizovanog uzorka $(x_1, x_2, \dots, x_n)$

je

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

Korigovana uzoračka disperzija $\hat{S}_n^2$ uzorka $(X_1, X_2, \dots, X_n)$ se definiše

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{S}_n^2,$$

korigovana uzoračka disperzija realizovanog uzorka $(x_1, x_2, \dots, x_n)$

je

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

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Deskriptivna statistika

Uočimo interval $[a, b) \subseteq \mathbb{R}$ koji sadrži sve vrednosti realizovanog uzorka $(x_1, x_2, \dots, x_n)$. Neka je $m_0 < m_1 < \dots < m_k$ i $a := m_0$ i $b := m_k$. Time je interval $[a, b)$ razbijen na disjunktne intervale $I_1 = [m_0, m_1)$, $I_2 = [m_1, m_2)$, $\dots$, $I_k = [m_{k-1}, m_k)$, čija je unija jednaka tom intervalu $[a, b)$.

Širine intervala se označavaju sa h_i , pa je $h_i := m_i - m_{i-1}$, $i \in \{1, 2, \dots, k\}$.

Sredine intervala se označavaju sa x_i , pa je $x_i := \frac{m_{i-1} + m_i}{2}$, $i \in \{1, 2, \dots, k\}$.

Frekvencija f_i se definiše kao broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.

$$f_1 + f_2 + \dots + f_k = n.$$

Korigovane frekvencije dobijamo iz $f' = \frac{f_i}{h_i}$.

Često se u praksi ne čuvaju sve vrednosti uzorka, već samo granice intervala I_i , odnosno deobne tačke m_i , $i = 0, 1, \dots, k$ i frekvencije f_i , $i = 1, 2, \dots, k$. Takav uzorak zovemo **intervalni uzorak**.

aritmetička sredina

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

uzoračka disperzija

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}_n^2 = \frac{1}{n} (x_1^2 f_1 + x_2^2 f_2 + \dots + x_k^2 f_k) - \bar{x}_n^2$$

standardna devijacija

$$\bar{s}_n = \sqrt{\frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}_n^2} = \sqrt{\frac{1}{n} (x_1^2 f_1 + x_2^2 f_2 + \dots + x_k^2 f_k) - \bar{x}_n^2}$$

gde je $n = \sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k$.

Često se u praksi ne čuvaju sve vrednosti uzorka, već samo granice intervala l_i , odnosno deobne tačke m_i , $i = 0, 1, \dots, k$ i frekvencije f_i , $i = 1, 2, \dots, k$. Takav uzorak zovemo **intervalni uzorak**.

aritmetička sredina

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

uzoračka disperzija

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}_n^2 = \frac{1}{n} (x_1^2 f_1 + x_2^2 f_2 + \dots + x_k^2 f_k) - \bar{x}_n^2$$

standardna devijacija

$$\bar{s}_n = \sqrt{\frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}_n^2} = \sqrt{\frac{1}{n} (x_1^2 f_1 + x_2^2 f_2 + \dots + x_k^2 f_k) - \bar{x}_n^2}$$

gde je $n = \sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k$.

Grafičko predstavljanje realizovanog uzorka

Histogram se konstruiše tako što se nad svakim od intervala $I_i = [m_{i-1}, m_i)$ konstruiše pravougaonik čija je visina jednaka $\frac{f_i}{h_i}$ za sve $i \in \{1, 2, \dots, k\}$, gde je f_i frekvencija, a h_i širina i -tog intervala.

Poligon je izlomljena linija koja polazi od tačke $(x_0, 0)$, prolazi kroz tačke $(x_i, \frac{f_i}{h_i})$ i završava u tački $(x_{k+1}, 0)$, gde su x_i sredine intervala I_i , $i \in \{0, 1, \dots, k, k+1\}$, a I_0 i I_{k+1} se nalaze redom ispred i iza intervala $I_1, I_2, \dots, I_k$.

Empirijska (uzoračka) funkcija raspodele F_n^* obeležja X je funkcija definisana za svako x na sledeći način:

$$F_n^*(x) = \frac{N_x}{n}$$

gde N_x predstavlja broj elemenata uzorka koji imaju vrednost obeležja manju od x , a n je obim realizovanog uzorka.

Grafičko predstavljanje realizovanog uzorka

Histogram se konstruiše tako što se nad svakim od intervala $I_i = [m_{i-1}, m_i)$ konstruiše pravougaonik čija je visina jednaka $\frac{f_i}{h_i}$ za sve $i \in \{1, 2, \dots, k\}$, gde je f_i frekvencija, a h_i širina i -tog intervala.

Poligon je izlomljena linija koja polazi od tačke $(x_0, 0)$, prolazi kroz tačke $(x_i, \frac{f_i}{h_i})$ i završava u tački $(x_{k+1}, 0)$, gde su x_i sredine intervala I_i , $i \in \{0, 1, \dots, k, k+1\}$, a I_0 i I_{k+1} se nalaze redom ispred i iza intervala $I_1, I_2, \dots, I_k$.

Empirijska (uzoračka) funkcija raspodele F_n^* obeležja X je funkcija definisana za svako x na sledeći način:

$$F_n^*(x) = \frac{N_x}{n}$$

gde N_x predstavlja broj elemenata uzorka koji imaju vrednost obeležja manju od x , a n je obim realizovanog uzorka.

Grafičko predstavljanje realizovanog uzorka

Histogram se konstruiše tako što se nad svakim od intervala $I_i = [m_{i-1}, m_i)$ konstruiše pravougaonik čija je visina jednaka $\frac{f_i}{h_i}$ za sve $i \in \{1, 2, \dots, k\}$, gde je f_i frekvencija, a h_i širina i -tog intervala.

Poligon je izlomljena linija koja polazi od tačke $(x_0, 0)$, prolazi kroz tačke $(x_i, \frac{f_i}{h_i})$ i završava u tački $(x_{k+1}, 0)$, gde su x_i sredine intervala I_i , $i \in \{0, 1, \dots, k, k+1\}$, a I_0 i I_{k+1} se nalaze redom ispred i iza intervala $I_1, I_2, \dots, I_k$.

Empirijska (uzoračka) funkcija raspodele F_n^* obeležja X je funkcija definisana za svako x na sledeći način:

$$F_n^*(x) = \frac{N_x}{n}$$

gde N_x predstavlja broj elemenata uzorka koji imaju vrednost obeležja manju od x , a n je obim realizovanog uzorka.

Realizovana empirijska funkcija raspodele:

$$f_n^*(x) = \begin{cases} 0 & \text{za } x \leq x_{(1)} \\ \frac{n_{x_i}}{n} & \text{za } x_{(i)} < x \leq x_{(i+1)} \quad i \in \{1, 2, \dots, n-1\} \\ 1 & \text{za } x > x_{(n)} \end{cases}$$

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ je permutacija $(x_1, x_2, \dots, x_n)$ čije su komponente u neopadajućem poretku.

Za intervalni uzorak:

$$f_n^*(x) = \begin{cases} 0 & \text{za } x \leq x_1 \\ \frac{n_{x_i}}{n} & \text{za } x_i < x \leq x_{i+1} \quad i \in \{1, 2, \dots, k-1\} \\ 1 & \text{za } x > x_k \end{cases}$$

gde su $x_1, x_2, \dots, x_k$ sredine intervala $I_i = [m_{i-1}, m_i)$,

$$n_{x_i} = \sum_{j=1}^i f_j = f_1 + f_2 + \dots + f_i \text{ odnosno } n_{x_{i+1}} = n_{x_i} + f_{i+1}, \quad n_{x_1} = f_1$$

i $i \in \{1, 2, \dots, k\}$.

Realizovana empirijska funkcija raspodele:

$$f_n^*(x) = \begin{cases} 0 & \text{za } x \leq x_{(1)} \\ \frac{n_{x_i}}{n} & \text{za } x_{(i)} < x \leq x_{(i+1)} \quad i \in \{1, 2, \dots, n-1\} \\ 1 & \text{za } x > x_{(n)} \end{cases}$$

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ je permutacija $(x_1, x_2, \dots, x_n)$ čije su komponente u neopadajućem poretku.

Za intervalni uzorak:

$$f_n^*(x) = \begin{cases} 0 & \text{za } x \leq x_1 \\ \frac{n_{x_i}}{n} & \text{za } x_i < x \leq x_{i+1} \quad i \in \{1, 2, \dots, k-1\} \\ 1 & \text{za } x > x_k \end{cases}$$

gde su $x_1, x_2, \dots, x_k$ sredine intervala $I_i = [m_{i-1}, m_i)$,

$$n_{x_i} = \sum_{j=1}^i f_j = f_1 + f_2 + \dots + f_i \text{ odnosno } n_{x_{i+1}} = n_{x_i} + f_{i+1}, \quad n_{x_1} = f_1$$

i $i \in \{1, 2, \dots, k\}$.

Modus označavamo sa Mo . Modus predstavlja elemenat realizovanog uzorka oje ima najveću frekvenciju. tj. najfrekventniji elemenat (koji se najviše puta ponavlja).

$Mo = m_{s-1} + h_s \frac{r_1}{r_1 + r_2}$. Ako su širine svih intervala iste onda je interval sa najvećom frekvencijom modalni interval, a ako su širine intervala različite onda je interval sa najvećom korigovanom frekvencijom modalni interval. Sa $I_s = (m_{s-1}, m_s)$ predstavljamo modalni interval, $h_s = m_s - m_{s-1}$ je širina modalnog intervala, $r_1 = f_s - f_{s-1}$ predstavlja razliku između frekvencije modalnog (najfrekventnijeg) intervala f_s i frekvencije intervala koji prethodi modalnom f_{s-1} , $r_2 = f_s - f_{s+1}$ predstavlja razliku između frekvencije modalnog intervala f_s i frekvencije intervala posle modalnog f_{s+1} .

Obeležje može imati jedan ili više modusa.

Modus označavamo sa Mo . Modus predstavlja elemenat realizovanog uzorka oje ima najveću frekvenciju. tj. najfrekventniji elemenat (koji se najviše puta ponavlja).

$Mo = m_{s-1} + h_s \frac{r_1}{r_1 + r_2}$. Ako su širine svih intervala iste onda je interval sa najvećom frekvencijom modalni interval, a ako su širine intervala različite onda je interval sa najvećom korigovanom frekvencijom modalni interval. Sa $I_s = (m_{s-1}, m_s)$ predstavljamo modalni interval, $h_s = m_s - m_{s-1}$ je širina modalnog intervala, $r_1 = f_s - f_{s-1}$ predstavlja razliku između frekvencije modalnog (najfrekventnijeg) intervala f_s i frekvencije intervala koji prethodi modalnom f_{s-1} , $r_2 = f_s - f_{s+1}$ predstavlja razliku između frekvencije modalnog intervala f_s i frekvencije intervala posle modalnog f_{s+1} . Obeležje može imati jedan ili više modusa.

Medijana, koju označavamo sa Me , za *prosti* uzorak, nakon što elemente uzorka poređamo po veličini u neopadajući niz je:

$$Me = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ neparno} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & n \text{ parno} \end{cases}.$$

Ako je uzorak *intervalni*, bilo da su širine intervala jednake ili različite, a obim uzorka veličine n , onda se medijana računa na sledeći način

$$Me = m_{l-1} + h_l \frac{\frac{n}{2} - n_{x_{l-1}}}{f_l}$$

gde $l = (m_{l-1}, m_l)$ predstavlja medijalni interval. Medijalni interval l je interval sa najmanjom kumulativnom frekvencijom većom od $\frac{n}{2}$, f_l predstavlja frekvenciju medijalnog intervala,

$h_l = m_l - m_{l-1}$ je širina medijalnog intervala, $n_{x_{l-1}} = \sum_{i=1}^{l-1} f_i$

predstavlja kumulativnu frekvenciju intervala l_{l-1} koji prethodi medijalnom intervalu l .

Primer

U domu zdravlja je vršeno merenje krvnog pritiska u mmHg kod 20 pacijenata. Odrediti modus, medijanu, aritmetičku sredinu i uzoračku disperziju za uzorak:

87	103	130	160	180	195	132	145	211	105
145	153	152	138	87	99	93	119	129	145

Poredjajmo vrednosti iz uzorka u neopadajući niz:

87	87	93	99	103	105	119	129	130	132
138	145	145	145	152	153	160	180	195	211

$$Mo = x_{(12)} = x_{(13)} = x_{(14)} = 145.$$

$$Me = \frac{1}{2}(x_{(10)} + x_{(11)}) = \frac{1}{2}(132 + 138) = 135.$$

$$\bar{x}_{20} = \frac{1}{20}(87 + 87 + 93 + \dots + 195 + 211) = \frac{2708}{20} = 135.4,$$

$$\bar{s}_{20}^2 = \frac{1}{20} \sum_{i=1}^{20} x_i^2 - \bar{x}_{20}^2 = \frac{1}{20} 389526 - 135.4^2 = 1143.14.$$

Nakon merenja mase 100 učenika, rezultati merenja su grupisani u intervale $[60, 62)$, $[62, 64)$, $[64, 66)$, $[66, 68)$, $[68, 70)$. U tabeli su prikazani dobijeni podaci.

masa [kg]	$[60, 62)$	$[62, 64)$	$[64, 66)$	$[66, 68)$	$[68, 70)$
broj učenika f_i	5	18	42	27	8

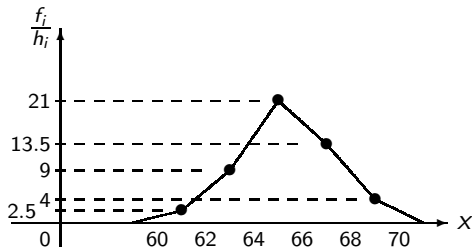
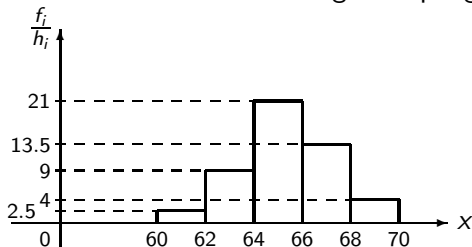
- (a) Nacrtati histogram, poligon, odrediti empirijsku funkciju raspodele i skicirati njen grafik.
- (b) Odrediti modus i medijanu.
- (c) Izračunati aritmetičku sredinu uzorka i uzoračku disperziju.

Dobijeni rezultati merenja učenika grupisani su u 5 intervala. Svi intervali su iste širine: $h_i = 2$, $i = 1, 2, 3, 4, 5$. Izračunavanjem sredina intervala, po formuli $x_i = \frac{m_{i-1} + m_i}{2}$, $i \in \{1, \dots, k\}$, dobijamo vrednosti $x_1 = 61$, $x_2 = 63$, $x_3 = 65$, $x_4 = 67$, $x_5 = 69$. Empirijska funkcija raspodele je $f_n^*(x) = \frac{n_{x_i}}{n}$, gde je $x \in (x_i, x_{i+1}]$, $i = 1, 2, 3, 4$.

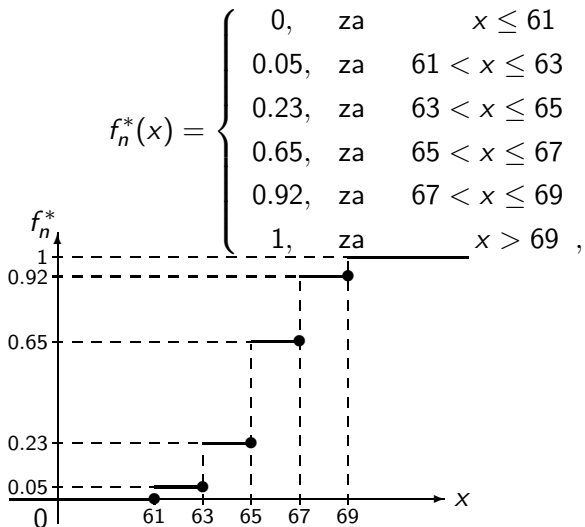
Svi dobijeni rezultati prikazani su u tabeli.

masa [kg] I_k	[60, 62)	[62, 64)	[64, 66)	[66, 68)	[68, 70)
br. uč. f_i	5	18	42	27	8
sred. int. x_i	61	63	65	67	69
$\frac{f_i}{h_i} = \frac{f_i}{2}$	2.5	9	21	13.5	4
kum. fr. n_{x_i}	5	23	65	92	100
$f_n^* = \frac{n_{x_i}}{n} = \frac{n_{x_i}}{100}$	0.05	0.23	0.65	0.92	1

Sad možemo nacrtati histogram i poligon frekvencija.



Realizovana empirijske funkcija raspodele i njen grafik su



Modalni interval je $[64, 66)$, $h = 2$, $r_1 = 42 - 18 = 24$,
 $r_2 = 42 - 27 = 15$, dakle $Mo = 64 + 2 \frac{24}{24+15} = 65.23 \text{ kg}$.
 Obim uzorka je $n = 5 + 18 + 42 + 27 + 8 = 100$, a najmanja
 kumulativna frekvencija veća od $\frac{n}{2} = 50$ je $n_{x_3} = 65$, pa je
 medijalni interval $[64, 66)$, širina mu je $h_l = 66 - 64 = 2$,
 $k_{l-1} = 23$, $f_l = 42$, dakle $Me = 64 + 2 \frac{50-23}{42} = 65.29 \text{ kg}$.

Aritmetička sredina uzorka je

$$\bar{x}_{100} = \frac{1}{100}(61 \cdot 5 + 63 \cdot 18 + 65 \cdot 42 + 67 \cdot 27 + 69 \cdot 8) = 65.30 \text{ kg},$$

a uzoračka disperzija je

$$\begin{aligned} \bar{s}_{100}^2 &= \frac{1}{100} \sum_{i=1}^5 x_i^2 f_i - \bar{x}_{100}^2 = \frac{1}{100}(61^2 \cdot 5 + 63^2 \cdot 18 + 65^2 \cdot 42 + \\ &67^2 \cdot 27 + 69^2 \cdot 8) - 65.30^2 = \frac{1}{100}426788 - 65.30^2 = 3.79 . \end{aligned}$$