

## PRIMERI STATISTIKA / TAČKASTE OCENE

Statistika  $S$  je neka određena funkcija uzorka,  $S = h(X_1, \dots, X_n)$ . Statistike služe za ocenu parametara (numeričkih karakteristika) obeležja  $X$ . Najvažniji primeri statistika:

1. **Sredina uzorka** je  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

U slučaju kada su podaci dati u intervalima, za  $X_i$  se uzima sredina uzorka.

2. **Medijana uzorka** je

$$\tilde{X} = \begin{cases} X'_{\frac{n+1}{2}}, & \text{ako je } n \text{ neparno} \\ \frac{1}{2} (X'_{\frac{n}{2}} + X'_{\frac{n}{2}+1}), & \text{ako je } n \text{ parno} \end{cases}$$

gde je  $X'_1, \dots, X'_n$  diskretni varijacioni niz, tj. važi  $X'_1 \leq X'_2 \leq \dots \leq X'_n$ .

U slučaju kada imamo intervalni varijacioni niz, medijana se određuje kao tačka koja histogram deli na dva dela jednake površine.

3. **Mod uzorka** je statistika jednaka slučajnoj promenljivoj  $X'_m, 1 \leq m \leq n$ , iz diskretnog varijacionog niza  $X'_1, \dots, X'_n$  koja ima najveću frekvenciju (odnosno relativnu frekvenciju). Kod obeležja datih intervalno mod se ne može neprekidno odrediti – on pripada intervalu koji ima najveću frekvenciju. Taj interval se naziva modalna klasa.

Kod simetričnih raspodela je  $\bar{X} = \tilde{X} = X'_m$ .

4. **Disperzija uzorka** je  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \overline{X^2} - (\bar{X})^2$ , odnosno  $\bar{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2$ , ako je poznato matematičko očekivanje  $a$  obeležja  $X$ .

$S = \sqrt{S^2}$  ( $S = \sqrt{\bar{S}^2}$ ) je **standardna devijacija (odstupanje) uzorka**.

5. **Popravljen disperzija uzorka**  $\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ,  $\tilde{S} = \frac{n}{n-1} S$ .

Ocnom na osnovu uzorka nepoznatog parametra  $\theta$  u raspodeli obeležja  $X$  nazivamo proizvoljnu funkciju uzorka, odnosno statistiku  $\hat{\theta} = h(X_1, \dots, X_n)$ . Dakle, za realizovani uzorak  $(x_1, \dots, x_n)$  kao ocenu za  $\theta$  dobijamo broj, odnosno tačku  $\theta^* = h(x_1, \dots, x_n)$ . Zato se ovakva ocena naziva tačkastom.

## INTERVAL POVERENJA

Kod intervalnog ocenjivanja definišemo par statistika  $\hat{\theta}_1 = h_1(X_1, \dots, X_n)$  i  $\hat{\theta}_2 = h_2(X_1, \dots, X_n)$  takvih da je  $P\{\hat{\theta}_1 < \hat{\theta}_2\} = 1$  i  $P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \beta = 1 - \alpha$ , gde je  $\beta$  unapred zadata verovatnoća i naziva se nivo poverenja,  $\alpha$  se naziva rizik ili greška, a (slučajni) interval  $(\hat{\theta}_1, \hat{\theta}_2)$  se naziva interval poverenja.

### Interval poverenja za matematičko očekivanje $a$ kada je poznata disperzija $\sigma^2$

Ako  $X$  ima  $N(a, \sigma^2)$  raspodelu, tada statistika  $\frac{(\bar{X}-a)\sqrt{n}}{\sigma}$  ima  $N(0,1)$  raspodelu, pa se interval poverenja za matematičko očekivanje  $a$  za dati nivo poverenja  $\beta$  nalazi iz uslova:

$$P\left\{\left|\frac{(\bar{X}-a)\sqrt{n}}{\sigma}\right| < z\right\} = \beta = 2\Phi(z)$$

Interval poverenja je  $a \in \left(\bar{X} - \frac{z\sigma}{\sqrt{n}}, \bar{X} + \frac{z\sigma}{\sqrt{n}}\right)$ , a njegova dužina je  $2z \frac{\sigma}{\sqrt{n}}$ . Za dati nivo poverenja  $\beta$ , z nalazimo u tablici za  $N(0,1)$  raspodelu (tablici IV) iz jednakosti  $\Phi(z) = \frac{\beta}{2}$ .

Ako raspodela obeležja  $X$  nije normalna, a uzorak je veliki ( $n > 30$ ), tada statistika  $\frac{(\bar{X}-a)\sqrt{n}}{\sigma}$  ima približnu  $N(0,1)$  raspodelu pa se istim postupkom dobija aproksimacija intervala poverenja.

### Interval poverenja za matematičko očekivanje $a$ kada disperzija $\sigma^2$ nije poznata

Ako obeležje  $X$  ima  $N(a, \sigma^2)$  raspodelu, tada statistika  $\frac{(\bar{X}-a)}{s}\sqrt{n-1}$  ima Studentovu  $t$  raspodelu sa  $n-1$  stepenom slobode, pa se interval poverenja za matematičko očekivanje  $a$  za dati nivo poverenja  $\beta$  nalazi iz uslova:

$$P\left\{\left|\frac{(\bar{X}-a)}{s}\sqrt{n-1}\right| < t\right\} = \beta$$

gde broj  $t$  nalazimo u tablici za Studentovu raspodelu (tablici VI -  $P\{|t_n| > t\} = \alpha$ ).

Interval poverenja je  $a \in \left(\bar{X} - t \frac{s}{\sqrt{n-1}}, \bar{X} + t \frac{s}{\sqrt{n-1}}\right)$ , a njegova dužina je  $2t \frac{s}{\sqrt{n-1}}$ .

Kako je  $s = \sqrt{\frac{n-1}{n}}\tilde{s}$ , takođe se koristi statistika  $\frac{(\bar{X}-a)\sqrt{n}}{\tilde{s}}$  pa je interval poverenja

$a \in \left(\bar{X} - t \frac{\tilde{s}}{\sqrt{n}}, \bar{X} + t \frac{\tilde{s}}{\sqrt{n}}\right)$ , čija je dužina  $2t \frac{\tilde{s}}{\sqrt{n}}$ .

Za veliki uzorak ( $n > 30$ ), raspodela  $t_{n-1}$  je približno jednaka  $N(0,1)$ , tako da kad je uzorak veliki prakrično nema razlike između slučaja kada je disperzija poznata i kada nije poznata.

Takođe, ako je uzorak veliki, a obeležje  $X$  nema normalnu raspodelu, tada  $\frac{(\bar{X}-a)}{s}\sqrt{n-1}$  ima približno  $N(0,1)$  raspodelu.