Formule iz statistike

1. Za interval $I_i = [m_{i-1}, m_i)$, $i \in \{1, 2, \ldots, k\}$, **sredina intervala** je $x_i = \frac{m_{i-1}+m_i}{2}$, a **širina intervala** je $h_i = m_i - m_{i-1}$.

2. **Frekvencija** f_i je broj elemenata realizovanog uzorka koji pripadaju intervalu $I_i = [m_{i-1}, m_i)$.
   Obim uzorka $n = \sum\limits_{i=1}^{k} f_i$. **Korigovana frekvencija** $f_i' = \dfrac{f_i}{h_i}$.

3. **Aritmetička sredina uzorka**
   - Za prost uzorak $\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$; za grupisan uzorak $\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^{k} f_i x_i$.

4. **Uzoračka disperzija**
   - Za prost uzorak $\bar{s}_n^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x}_n)^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \bar{x}_n^2$
   - za grupisan uzorak $\bar{s}_n^2 = \dfrac{1}{n}\sum\limits_{i=1}^{k} f_i(x_i - \bar{x}_n)^2 = \dfrac{1}{n}\sum\limits_{i=1}^{k} f_i x_i^2 - \bar{x}_n^2$

5. **Standardna devijacija** ili **standardno odstupanje** $\bar{s}_n = \sqrt{\bar{s}_n^2}$.

6. **Korigovana uzoračka disperzija** $\tilde{s}_n^2 = \dfrac{n}{n-1}\bar{s}_n^2$

7. **Realizovana empirijska funkcija raspodele** f_n^* je data sa:

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

   gde je n_{x_i} **kumulativna frekvencija** $n_{x_i} = f_1 + f_2 + \ldots + f_i$, $i \in \{1, 2, \ldots, k\}$.

8. **Modus**
   - kod prostog uzorka predstavlja vrednost obeležja koje ima najveću frekvenciju;
   - kod intervalnog uzorka $Mo = m_{s-1} + h_s \dfrac{r_1}{r_1 + r_2}$.
     $I_s = [m_{s-1}, m_s)$ je modalni interval, h_s je širina modalnog intervala, $r_1 = f_s' - f_{s-1}'$ i $r_2 = f_s' - f_{s+1}'$.

9. **Medijana** je broj koji uzorak deli na dva jednaka dela.
   - Kod prostog uzorka, nakon što elemente istog poređamo u neopadajući niz:
$$Me = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ neparno} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & n \text{ parno} \end{cases}.$$
   - kod intervalnog uzorka $Me = m_{l-1} + h_l \dfrac{\frac{n}{2} - n_{x_{l-1}}}{f_l}$
     $I_l = [m_{l-1}, m_l)$ je medijalni interval, h_l je širina medijalnog intervala, $n_{x_{l-1}}$ je kumulativnu frekvencija intervala I_{l-1}, f_l je frekvencija medijalnog intervala i n je veličina uzorka.

10. **Metod momenata** $E(X) = \bar{x}_n;\ D(X) = \tilde{s}_n^2$

11. **Metod maksimalne verodostojnosti** Neka je $(x_1, x_2, \ldots, x_n)$ realizovani uzorak. **Funkcija verodostojnosti** je

$$L(x_1, \ldots, x_n; \theta) = \begin{cases} \mathrm{p}(x_1, \theta)\mathrm{p}(x_2, \theta)\ldots\mathrm{p}(x_n, \theta), & \text{diskretno obeležje} \\ \varphi_X(x_1, \theta)\varphi_X(x_2, \theta)\ldots\varphi_X(x_n, \theta), & \text{neprekidno obeležje.} \end{cases}$$

12. **Intervali poverenja**
   - **Interval poverenja za nepoznati parametar m obeležja sa normalnom $\mathcal{N}(m, \sigma)$ raspodelom ako je standardna devijacija σ obeležja poznata**

$$\mathbf{I} = \left(\bar{x}_n - a\frac{\sigma}{\sqrt{n}}, \bar{x}_n + a\frac{\sigma}{\sqrt{n}}\right), \quad a = \phi^{-1}(\frac{1+\beta}{2})$$

- **Interval poverenja za nepoznati parametar m obeležja sa normalnom $\mathcal{N}(m,\sigma)$ raspodelom ako je standardna devijacija σ obeležja nepoznata**

$$\mathbf{I} = \left(\bar{x}_n - a\frac{\bar{s}_n}{\sqrt{n-1}}, \bar{x}_n + a\frac{\bar{s}_n}{\sqrt{n-1}} \right).$$

  Ako je $n \geq 30$, $a = \phi^{-1}(\frac{1+\beta}{2})$, a ako je $n < 30$, $a = t_{n-1,\frac{1+\beta}{2}}$.

- **Interval poverenja za nepoznatu disperziju σ^2 obeležja sa normalnom raspodelom $\mathcal{N}(m,\sigma)$**

  dvostrani: $\mathbf{I} = \left(\frac{n\bar{s}_n^2}{a}, \frac{n\bar{s}_n^2}{b} \right)$, **jednostrani:** $\mathbf{I} = \left(0, \frac{n\bar{s}_n^2}{c} \right)$, $a = \chi^2_{n-1,\frac{1+\beta}{2}}$, $b = \chi^2_{n-1,\frac{1-\beta}{2}}$, $c = \chi^2_{n-1,\beta}$.

- **Interval poverenja za nepoznatu proporciju p obeležja sa binomnom $\mathcal{B}(n,p)$ raspodelom**

$$\mathbf{I} = \left(\hat{p} - a\sqrt{\frac{\hat{p}\hat{q}}{n-1}}, \hat{p} + a\sqrt{\frac{\hat{p}\hat{q}}{n-1}} \right), \quad a = \phi^{-1}(\frac{1+\beta}{2})$$

  Interval poverenja za nepoznat broj pozitivnih realizacija obeležja sa binomnom $\mathcal{B}(n,p)$ raspodelom **u populaciji obima** N

$$\mathbf{I} = \left(N(\hat{p} - a\sqrt{\frac{\hat{p}\hat{q}}{n-1}}), N(\hat{p} + a\sqrt{\frac{\hat{p}\hat{q}}{n-1}}) \right), \quad a = \phi^{-1}(\frac{1+\beta}{2})$$

13. **Parametarski testovi jednog uzorka -** Nulta hipoteza: $\mathbf{H_0}(\theta = \theta_0)$. Pomoću intervala poverenja, sa nivoom poverenja $\beta = 1 - \alpha$. Za testiranje disperzije koristi se jednostrani interval poverenja.

14. **Parametarski testovi dva uzorka -** Nulta hipoteza: $\mathbf{H_0}(\theta_1 = \theta_2)$.

- **Test jednakosti srednjih vrednosti dva uzorka sa Normalnom raspodelom**

  1) Ako je $n_1 + n_2 \geq 30$:

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{Se(\bar{x}_1 - \bar{x}_2)}, \quad Se(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\bar{s}_1^2}{n_1} + \frac{\bar{s}_2^2}{n_2}} \quad a = \phi^{-1}(\frac{1+\beta}{2})$$

  2) Ako je $n_1 + n_2 < 30$:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{Se(\bar{x}_1 - \bar{x}_2)}, \quad Se(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{(n_1 - 1)\bar{s}_1^2 + (n_2 - 1)\bar{s}_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 n_2}}, \quad a = t_{n_1+n_2-2,\frac{1+\beta}{2}}$$

- **Test jednakosti proporcija dva uzorka sa Binomnom raspodelom**

$$z_0 = \frac{\bar{p}_1 - \bar{p}_2}{se(\bar{p}_1 - \bar{p}_2)}, \quad Se(\bar{p}_1 - \bar{p}_2) = \sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}, \quad p = \frac{k_1 + k_2}{n_1 + n_2}, \quad q = 1 - p, \quad a = \phi^{-1}(\frac{1+\beta}{2})$$

15. **Pirsonov χ^2-test -** Nulta hipoteza: $\mathbf{H_0}(\mathbf{F} = \mathbf{F_0})$.

$$z = \sum_{m=1}^{k} \frac{(n_m - np_m)^2}{np_m}, \quad p_m = F_0(b_m) - F_0(a_m), \quad a = \chi^2_{1-\alpha,k-1-s}$$

16. **Tabele kontingencije -** Nulta hipoteza: Obeležja X i Y su nezavisna.

$$z = \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{(f'_{ij} - f_{ij})^2}{f'_{ij}}, \quad f'_{ij} = \frac{f_{i\star} \cdot f_{\star j}}{n}, \quad a = \chi^2_{1-\alpha,(k-1)\times(r-1)}$$

17. **Uzoračka korelacija i regresija**

Koeficijent korelacije $r = \frac{\bar{s}_{xy}}{\bar{s}_x \bar{s}_y}$ gde je $\bar{s}_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$; $\bar{s}_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}$;

$$\bar{s}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)$$

Jednačina linearne regresije (regresiona prava Y u zavisnosti od X):

I način: $\hat{y} = \bar{y}_n + b(x - \bar{x}_n)$, gde je $b = r\frac{\bar{s}_y}{\bar{s}_x} = \frac{\bar{s}_{xy}}{\bar{s}_x^2}$.

II način $\hat{y} = a_1 x + a_0$, gde je $a_1 = r\frac{\bar{s}_y}{\bar{s}_x} = \frac{\bar{s}_{xy}}{\bar{s}_x^2}$, $a_0 = \bar{y}_n - a_1\bar{x}_n$.