NUMERIČNE METODE ZA RAČUNANJE PARAMETROV

Numerične opisne mere (parametri)so števila izračunana iz množice podatkov z namenom, da nam pomagajo zgraditi miselno predstavo o populaciji

Delimo jih na mere, ki pomagajo določiti

1. ***sredino* podatkov**
2. ***razpršenost* podatkov**
3. ***relativno mesto* podatka v urejeni množici podatkov.**

Mere srednje vrednosti

Aritmetična sredina je povprečna vrednost množice podatkov

Če so njene vrednosti, $\left\{ x_1, x_2, x_3, \ldots x_n \right\}$
je aritmetična sredina :

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

Kadar so vrednosti podane v obliki frekvenčne distribucije izračunamo aritmetično sredino po naslednji formuli

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{n} x_i^s \cdot f_i$$

kjer pomeni :

n število razredov ali število vrednosti

$$\mathbf{N} = \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{f}_{\mathbf{i}}$$ **število podatkov**

x_i^s **sredine razredov za zvezno frekvenčno distribucijo in vrednosti za diskretno frekvenčno distribucijo**

Mediana(Me) je vrednost z lastnostjo, da ima polovica podatkov manjšo ali enako vrednost od nje

Če so podatki urejeni po velikosti in jih je 2n-1, velja

$$Me = x_n$$

Če pa je 2n podatkov urejenih po velikosti, je mediana določena

$$Me = \frac{x_n + x_{n+1}}{2}$$

Za podatke podane v obliki frekvenčne distribucije izračunamo mediano po naslednji formuli

$$\mathbf{Me} = \mathbf{r}_{\mathbf{k}-1} + \frac{\dfrac{\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{f}_{\mathbf{i}}}{2} - \sum_{\mathbf{i}=1}^{\mathbf{k}-1} \mathbf{f}_{\mathbf{i}}}{\mathbf{f}_{\mathbf{k}}} \Delta \mathbf{x}_{\mathbf{k}}$$

Medianin razred je tisti v katerem prvič nastopi :

$$\sum_{\mathbf{i}=1}^{\mathbf{k}} \mathbf{f}_{\mathbf{i}} > \frac{1}{2} \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{f}_{\mathbf{i}}$$

n **število razredov**

k **medianin razred**

$\mathbf{r_{i-1}}$ **spodnja meja medianinega razreda**

Δx_k **širina medianinega razreda**

Modus(Mo) je tisto število, ki v množici podatkov največkrat nastopi

Za zvezno frekvenčno distribucijo izračunamo modus po naslednji formuli :

$$Mo = r_{k-1} + \frac{f_k - f_{k-1}}{\left(f_k - f_{k-1}\right) + \left(f_k - f_{k+1}\right)} \Delta x_k$$

modalni razred je tisti ,pri katerem je frekvenca največja

k : modalni razred

$\mathbf{r_{k-1}}$ spodnja meja modalnega razreda

$\mathbf{\Delta x_k}$ širina modalnega razreda

Harmonična sredina je število določeno s formulo :

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

Če so podatki podani z **frekvenčno distribucijo, harmonično sredino** izračunamo po formuli

$$\mathbf{x_H} = \frac{\sum_{i=1}^{n} \mathbf{f_i}}{\sum_{i=1}^{n} \frac{\mathbf{f_i}}{\mathbf{x_i}}}$$

n število razredov v frekvenčni distribuciji

$\mathbf{f_i}$ **frekvence razredov.**

Geometrična sredina je vrednost določena s formulo:

$$x_G = \sqrt[N]{x_1 x_2 \ldots x_N}$$

Za podatke podane v obliki frekvenčne distribucije, geometrično sredino izračunamo po formuli

$$x_G = \sqrt[N]{x_1^{f_1} x_2^{f_2} \ldots x_n^{f_n}}$$

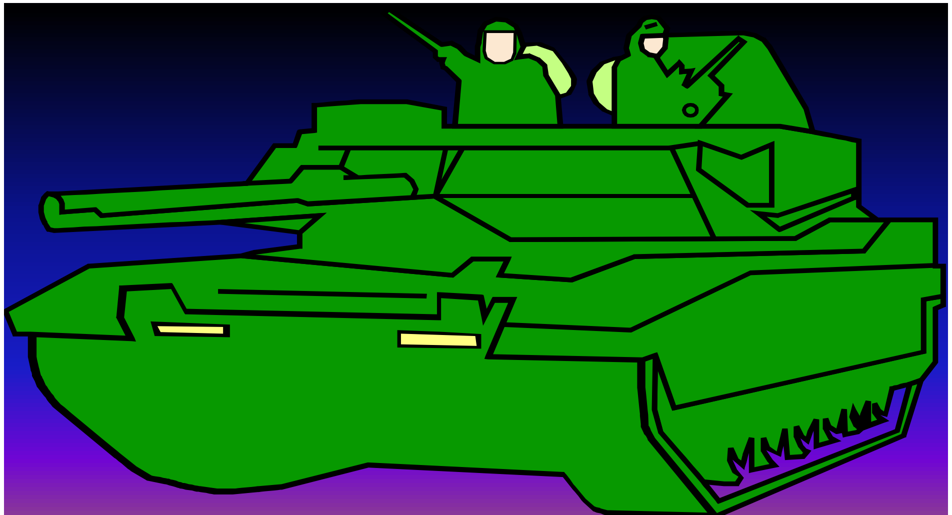
n število razredov v frekvenčni distribuciji

$\mathbf{f_i}$ **frekvence razredov**

Ranžirne vrste

Podatke razvrstimo po nekem kriteriju, tako da v tej razvrstitvi vsak podatek dobi svoje mesto

Taki razvrstitvi podatkov pravimo *ranžirana vrsta,* mesto ki ga v ranžirni vrsti zavzame neka enota , pa se imenuje *rang*.



Za zvezno frekvenčno distribucijo izračunamo rang neke enote na naslednji način :

Enoti z vrednostjo x, ki se nahaja v k-tem razredu

$r_{k-1} \leq x < r_k$ **izračunamo rang po obrazcu**

$$R_x = \sum_{i=1}^{k-1} f_i + f_k \frac{x - r_{k-1}}{r_k - r_{k-1}}$$

Kvantilni rang **enote je relativno mesto enote v ranžirni vrsti glede na število (N) enot v ranžirni vrsti. Izračunamo ga po formuli**

$$P_x = \frac{R_x - 0.5}{N}$$

$\kappa - kvantil$ **ki pripada urejeni vrsti podatkov**

$(0 < \kappa < 1)$ $\boldsymbol{x_1 \leq x_2 \leq \ldots x_N}$ **je**

$$\tilde{x}_\kappa = \begin{cases} x_k & k \text{ prvo celo število večje} \\ & \text{od } \mathrm{N}\cdot\kappa, \text{ če } \mathrm{N}\cdot\kappa \text{ ni celo število} \\ \frac{1}{2}\left(x_k + x_{k+1}\right) & k = N\cdot\kappa, \text{ če je } \mathrm{N}\cdot\kappa \text{ celo število} \end{cases}$$

$\kappa - kvantil$ **za podatke prikazane z zvezno frekvenčno porazdelitvijo je**

$$\tilde{x}_{\kappa} = r_{k-1} + \frac{N \cdot \kappa - \sum_{i=1}^{k-1} f_i}{f_k} \cdot \Delta x$$

r_{k-1} **spodnja meja razreda za katerega prvič velja**

$$\sum_{i=1}^{k} f_i \geq N \cdot \kappa$$

Δx **širina razredov**

Kvartili Q_1, Q_2, Q_3

so tiste vrednosti,ki imajo kvantilne range

0.25, 0.50, 0.75.

Mere variabilnosti

Variacijski razmik **R** je **razlika med največjo in najmanjšo vrednostjo :**

$$R = x_{max} - x_{min}$$

Varianca **meri razpršenost podatkov okrog aritmetične sredine. Iz negrupiranih N podatkov jo izračunamo po formuli**

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

Iz frekvenčne distribucije z n razredi pa varianco izračunamo

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

Pri tem je :

$$N = \sum_{i=1}^{n} f_i$$

x_i^s **sredina i-tega razreda**

Standardni odklon s **imenujemo kvadratni koren iz variance**

$$s = \sqrt{s^2}$$

Za ročno računanje obrazce za računanje variance preuredimo v obliko

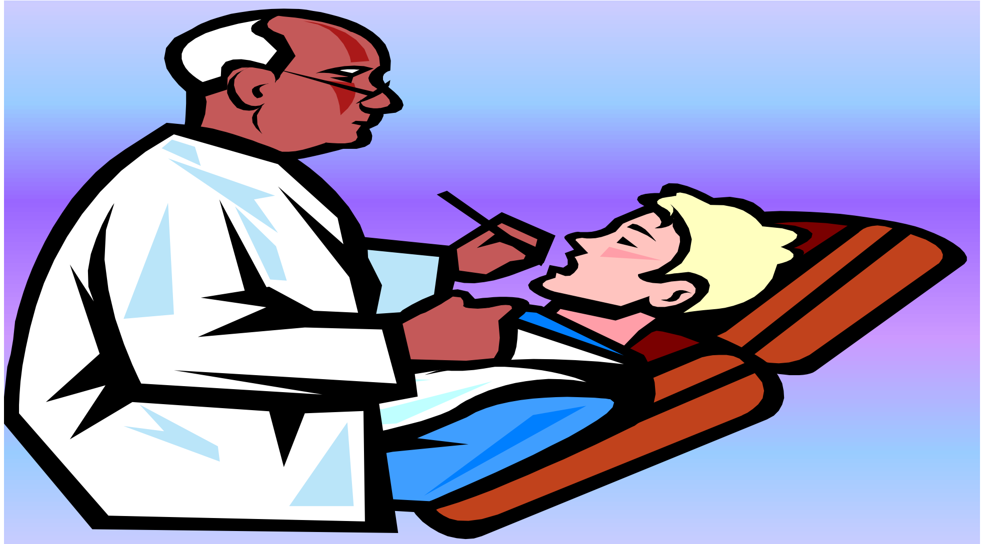
$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{n} x_i^{\;2} - N \cdot \bar{x}^2\right]$$

in za podatke podane s frekvenčno porazdelitvijo

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{n} f_i \cdot (x_i^s)^2 - N \cdot \bar{x}^2\right]$$

Variacijski koeficient **je relativna mera variabilnosti in je izražen v odstotkih, določen pa je z razmerjem med standardnim odklonom in aritmetično sredino populacije :**

$$\mathbf{K_V} = \frac{\mathbf{s}}{\bar{\mathbf{x}}} 100\%$$



Asimetrija in Sploščenost

Asimetrija meri velikost in smer asimetričnosti množice podatkov glede na aritmetično sredino

Če so $\{x_1, x_2, \ldots, x_N\}$

vrednosti iz množice podatkov in $\bar{x}$ njihova aritmetična sredina, potem količino

$$g_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^3}}$$

imenujemo koeficient asimetričnosti.

Kadar je :

$\mathbf{g}_1 = 0$ **je množica podatkov simetrična na $\bar{\mathbf{x}}$**

$\mathbf{g}_1 < 0$ **je množica podatkov asimetrična v levo**

$\mathbf{g}_1 > 0$ **je množica podatkov asimetrična v desno**

Če je množica podatkov podana v obliki frekvenčne distribucije, koeficient asimetričnosti izračunamo :

$$g_1 = \frac{\frac{1}{N}\sum_{i=1}^{n} f_i \left(x_i^s - \overline{x}\right)^3}{\sqrt{\left(\frac{1}{N}\sum_{i=1}^{n} f_i \left(x_i^s - \overline{x}\right)^2\right)^3}}$$

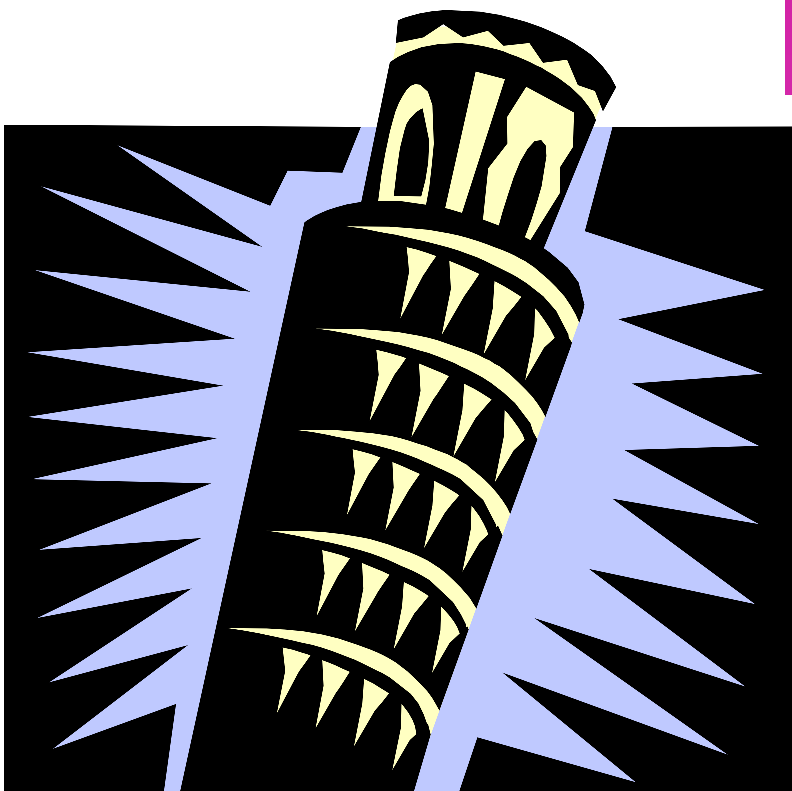
n število razredov pri zvezni frekvenčni distribuciji in število vrednosti, ki jih enote lahko zavzamejo pri diskretni frekvenčni distribuciji

f_i frekvence

x_i^s sredine razredov pri zvezni frekvenčni distribuciji in vrednosti statističnega znaka pri diskretni frekvenčni distribuciji

Približna formula

$$g1 = \frac{3.(\bar{x}\text{-Me})}{\sigma}$$



Sploščenost (eksces) je mera, ki porazdelitev dane množice podatkov(z enim maksimumom) primerja z gostoto normalne porazdelitve (pri enaki varianci) glede na absolutni maksimum

Teoretični eksces pri normalni porazdelitvi podatkov je 0.

To mero izračunamo na sledeč način

$$g_2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^2} - 3$$

Za podatke podane v obliki frekvenčne distribucije izračunamo mero sploščenosti

$$g_2 = \frac{\frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i^s - \overline{x}\right)^4}{\left[\frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i^s - \overline{x}\right)^2\right]^2} - 3$$

$g_2 > 0$

absolutni maksimum množice podatkov je večji od pripadajoče normalne porazdelitve

$g_2 < 0$

absolutni maksimum množice podatkov je manjši od pripadajoče normalne porazdelitve

Kadar sta koeficient asimetrije in sploščenosti bistveno različna od nič, je to dokaz zato, da se porazdelitev dane množice podatkov bistveno razlikuje od normalne porazdelitve.