BIOSTATISTIKA (studij Nutricionizam)
(formule za prvi parcijalni ispit)

Deskriptivna statistika

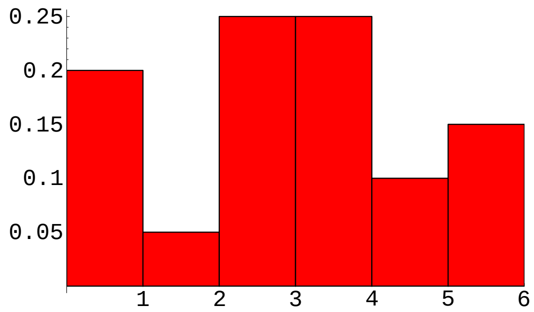
$\mathrm{Im}X = \{a_1, a_2, \ldots, a_i, \ldots\} \Rightarrow$ skup svih vrijednosti koje X može poprimiti

f_i = **frekvencija (učestalost) pojavljivanja elementa** a_i u nizu podataka

broj $f_{r_i} = \dfrac{f_i}{n}$: **relativna frekvencija** od a_i (n je broj ponavljanja pokusa)

Histogram:

- svaka 2 susjedna stupića se dodiruju i svaki ima težište u vrijednosti visina f_i ili f_{r_i}
- površina svakog stupića jednaka je relativnoj frekvenciji pa je površina ispod cijelog grafa jednaka je 1



broj razreda (okvirno: $\sqrt{n}$)

Aritmetička sredina uzorka je broj

$$\bar{x} := \frac{1}{n}(x_1 + x_2 + \ldots + x_n).$$

Ako je $\mathrm{Im}X = \{a_1, a_2, \ldots, a_k\}$ i pritom se a_i u uzroku ponavlja f_i puta, tada

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i \cdot a_i, \quad n = \sum_{i=1}^{k} f_i.$$

Medijan uzorka je broj za koji vrijedi da je 50% svih podataka manje od ili jednako njemu i 50% svih podataka veće od ili jednako njemu.

Ako je broj podataka neparan, tj $n = 2k-1,\ k \in \mathbf{N}$, tada je $m = x_{(k)}$. Za paran n ($n = 2k$), vrijedi

$$m = \frac{x_{(k)} + x_{(k+1)}}{2}.$$

Općenito, $m = x_{(\frac{n+1}{2})}$. Vrijedi

$$x_{(\frac{p}{q})} = x_{(k+\frac{r}{q})}$$
$$x_{(\frac{p}{q})} := x_{(k)} + \frac{r}{q}\left(x_{(k+1)} - x_{(k)}\right)$$

Mod je ona vrijednost statističkog obilježja koja se u uzorku javlja s najvećom frekvencijom.

- BIMODALNI UZORAK: uzorak u kojem postoje 2 vrijednosti s jednakom frekvencijom
- UNIMODALNI UZORAK: uzorak u kojem postoji samo jedan mod
- Ako svi podaci imaju istu frekvenciju pojavljivanja u uzorku, tada uzorak nema mod.

Neka je $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ uređeni niz podataka. Broj

$$d = x_{(n)} - x_{(1)}$$

naziva se **raspon uzorka.**

Donji kvartil q_L je ona vrijednost uzroka za koju vrijedi da je 25% svih podataka manje ili jednako od nje i 75% svih podataka veće ili jednako od nje.

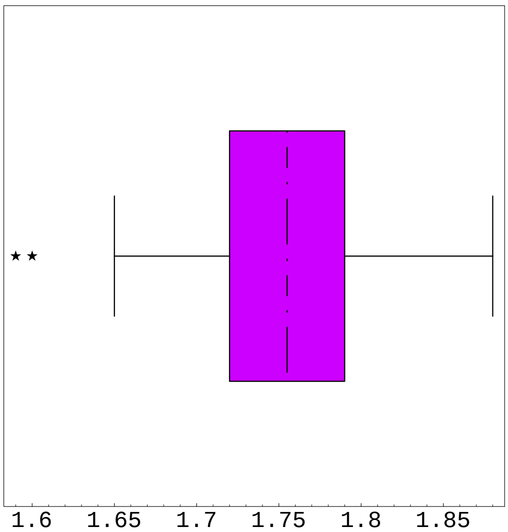
$$q_L = x_{(\frac{n+1}{4})}$$

Gornji kvartil q_U je ona vrijednost uzroka za koju vrijedi da je 75% svih podataka manje ili jednako od nje i 25% svih podataka veće ili jednako od nje.

$$q_U = x_{(\frac{3(n+1)}{4})}$$

Interkvartil: $d_q = q_U - q_L$

Uređenu petorku $(x_{(1)}, q_L, m, q_U, x_{(n)})$ zovemo **karakteristična petorka uzorka.** Pomoću nje crtamo tzv. **"box and whisker" dijagram**, odnosno dijagram pravokutnika.



Sve što je izvan $3d_q$ označava se točkom i smatra se "ekstrmnom vrijednošću", tj. "outlier".

Koeficijent kvartilne varijacije:

$$v_q = \frac{d_q}{q_L + q_U}$$

Uzoračka varijanca:

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

Uzoračka standardna devijacija:

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

Vrijedi:

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

Ako se u uzroku $x_1, x_2, \ldots, x_n$ vrijednosti $a_1, a_2, \ldots, a_k$ pojavljuju s frekvencijom $f_1, f_2, \ldots, f_k$, onda vrijedi:

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

DECILI: k-ti uzorački decil je broj $D_k = x_{\left(\frac{k(n+1)}{10}\right)}$, $k = 1, 2, \ldots, 9$ (k/10 podataka je manje ili jednako njemu)

PERCENTILI: k-ti uzorački percentil je broj $P_k = x_{\left(\frac{k(n+1)}{100}\right)}$, $k = 1, 2, \ldots, 99$ (k% podataka je manje ili jednako njemu)

Uzorački k-ti centralni moment, $k \in \mathbb{N}$:

$$\mu_k = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^k$$

Specijalno,

$$\mu_1 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) = \frac{1}{n-1} \sum_{n=1}^{n} x_i - \frac{n\bar{x}}{n-1} = \frac{n\bar{x}}{n-1} - \frac{n\bar{x}}{n-1} = 0$$

$$\mu_2 = s^2$$

$$\mu_3 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^3$$

Koeficijent asimetrije uzorka (skewness) definiran je s:

$$\alpha_3 = \frac{\mu_3}{s^3} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3 = \frac{1}{n-1} \sum_{i=1}^{k} f_i \cdot \left(\frac{a_i - \bar{x}}{s} \right)^3$$

Vrijedi:

(i) $\alpha_3 = 0 \Rightarrow$ uzorak je SIMETRIČAN

(ii) $\alpha_3 > 0 \Rightarrow$ uzorak je POZITIVNO ASIMETRIČAN

(iii) $\alpha_3 < 0 \Rightarrow$ uzorak je NEGATIVNO ASIMETRIČAN

Linearna regresija:

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

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right)$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \right)$$

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Pearsonov koeficijent korelacije:

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y} \right) = \frac{s_{xy}}{s_x \cdot s_y}, \quad -1 \le r \le 1.$$

Ako je $y_i = \alpha + \beta x_i$ za sve $i = 1, \ldots, n$ (tj. ako su x i y u egzaktnoj linearnoj vezi), tada je $r = \frac{\beta}{|\beta|}$.

1. $r = 0$ nema korelacije
2. $r = 1 (> 0)$ pozitivna korelacija (ako x raste, i y u pravilu raste)
3. $r = -1 (< 0)$ negativna korelacija (ako x raste, y u pravilu pada)

Osnove teorije vjerijatnosti

Slučajni pokus je pokus s više (mogućih) ishoda. Ishode slučajnog pokusa zovemo **događajima**. Događaje koje ne možemo razložiti na jednostavnije događaje zovemo **elementarnim događajima**.

Matematički, skup elementarnih događaja označavamo s Ω i zovemo **prostorom elementarnih događaja**, a njegove elemente (elementarne događaje) s $\omega_1, \omega_2, \ldots$:

$$\Omega = \{\omega_1, \omega_2, \ldots\}.$$

Događaji su podskupovi od Ω i označavaju se velikim slovima abecede: $A, B, C, \ldots$

Događaj C je **suprotan** događaj događaju A ako je $C = A^c = \Omega \backslash A$.

Laplaceov model vjerojatnosti: $P(A) = \frac{k(A)}{k(\Omega)}$

Neka je A n-člani skup i neka je $r \in \mathbf{N}$ takav da je $r \leq n$. **Varijacija r-tog razreda** u skupu A jest svaka uređena r-torka $(a_1, a_2, \ldots, a_r)$ čije su sve komponenete $a_1, \ldots, a_r$ međusobno različiti elementi skupa A.

$$V_n^{(r)} = n(n-1)\ldots(n-r+1)$$

Permutacija u n-članom skupu A je svaka varijacija n-tog razreda u skupu A.

$$P_n = n!$$

Neka je A skup od n elemenata i neka je $r \in \mathbf{N}$. **Varijacija s ponavljanjem r-tog razreda** u skupu A je svaka uređena r-torka elemanata iz skupa A (budući da članovi te r-torke mogu biti međusobno jednaki, može biti $r > n$)

$$\bar{V}_n^{(r)} = [k(A)]^r$$

Permutacije s ponavljanjem: Označimo sa $\bar{P}_n^{(n_1,\ldots,n_k)}$ ukupan broj permutacija od n elemenata među kojima je n_1 prve vrste, n_2 druge vrste, $\ldots$, n_k k-te vrste $(n_1 + n_2 + \ldots + n_k = n)$. Tada vrijedi

$$\bar{P}_n^{(n_1,\ldots,n_k)} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

Neka je A n-člani skup i $r \in \mathbf{N}$, $r < n$. **Kombinacija r-tog razreda** u skupu A jest svaki r-člani podskup od A.

$$C_n^{(r)} = \binom{n}{r} = \frac{n!}{(n-r)! \cdot r!} = \frac{n(n-1)\ldots(n-r+1)}{r!}$$

Neka je A n-člani skup i $r \in \mathbf{N}$. **Kombinacija s ponavljanjem r-tog razreda** u skupu A je svaka **neuređena** r-torka $(a_1, a_2, \ldots, a_r)$ elemenata iz skupa A (u njoj elementi mogu biti međusobno jednaki)

$$\bar{C}_n^{(r)} = \binom{n+r-1}{r}$$

Uvjetna vjerojatnost: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Formula potpune (totalne) vjerojatnosti:

$$P(A) = P(A \cap \Omega) = P(\cup_{i=1}^n A \cap H_i) = \sum_{i=1}^n P(A \cap H_i) = \sum_{i=1}^n P(H_i)P(A|H_i).$$

Bayesova formula (aposteriorne vjerojatnosti hipoteza):

$$P(H_i|A) = \frac{P(A \cap H_i)}{P(A)} = \frac{P(H_i)P(A|H_i)}{\sum_{i=1}^n P(H_i)P(A|H_i)}.$$

Osnovni pojmovi vezani uz slučajnu varijablu

Funkcija distribucije: $F_X(x) = P(X \leq x)$

Funkcija gustoće vjerojatnosti: $p_X(a_i) := P(X = a_i) = p_i$

Matematičko očekivanje: $\mathrm{E}[X] = \sum_{a_i \in \mathrm{Im}X} a_i \cdot p_X(a_i)$

Varijanca: $\mathrm{Var}[X] := \sum_{a_i \in \mathrm{Im}X} (a_i - \mathrm{E}X)^2 \cdot p_X(a_i) = \mathrm{E}[X^2] - (\mathrm{E}X)^2$.

Vrijedi: $\mathrm{Var}[aX + b] = a^2 \mathrm{Var}X$

Standardna devijacija: $\sigma_X := +\sqrt{\mathrm{Var}[X]}$

Čebiševljeva nejednakost: $P(|X - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathrm{Var}[X]$

Binomna slučajna varijabla

$$\mathrm{Im}X = \{0, 1, 2, \ldots, n\},$$

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p, \quad k = 0, 1, \ldots, n$$

Oznaka: $X \sim B(n, p)$.

$$\mathrm{E}[X] = np, \ \mathrm{Var}[X] = npq.$$

Slabi zakon velikih brojeva: $P\left(\left|\frac{Y_n}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2}$