Osnove statistike u demografiji

Predavanje 4

Mjere disperzije

- Mjerama disperzije mjeri se stupanj homogenosti numeričkog niza, odnosno stupanj varijabilnosti podataka. Razlikujemo:
  - Potpune mjere disperzije, odnosno mjere u čijem izračunu sudjeluju sve vrijednosti (numeričke) varijable su varijanca, standardna devijacija i koeficijent varijacije, dok su ostale mjere
  - nepotpune mjere disperzije (raspon varijacije, interkvartilni raspon, koeficijent kvartilne devijacije).
- Raspon varijacije, interkvartilni raspon i standardna devijacija izraženi su u **mjernim jedinicama varijable** za koju se izračunavaju, dok su ostale mjere izražene relativno

Raspon varijacije, interkvartil i koeficijent kvartilne devijacije

- Raspon varijacije R je razlika između najveće i najmanje vrijednosti numeričke varijable

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

- Raspon varijacije središnjih 50% podataka mjeri se interkvartilom IQ (apsolutna mjera) i

$$I_Q = Q_3 - Q_1,$$

- koeficijentom kvartilne devijacije VQ (relativna mjera)

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}, \qquad 0 \leqslant V_Q \leqslant 1.$$

Raspon varijacije, interkvartil i koeficijent kvartilne devijacije

- Analogno se mogu definirati i interdecilni i interpercentilni raspon.
- Npr. razlika sedmog i trećeg decila pokazuje raspon varijacije središnjih 40% podataka, dok je razlika osamdeset i petog i petnaestog percentila raspon varijacije središnjih 70% podataka

B-P dijagram

- Stupanj varijabilnosti može se predočiti grafički B-P dijagramom (engl. Box-Plot).
- B-P dijagram se temelji na pet osnovnih pokazatelja (5 statistika, 5'S; engl. Five-Number Summaries): x_{min} (najmanjoj vrijednosti), Q_1 (prvom kvartilu), Me (medijanu), Q_3 (trećem kvartilu) i x_{max} (najvećoj vrijednosti).

Varijanca, standardna devijacija i koeficijent varijacije

- **Varijanca** σ^2 je prosječno kvadratno odstupanje vrijednosti numeričke varijable od aritmetičke sredine.
- Mjeri **varijabilnost podataka** i **reprezentativnost aritmetičke sredine**.
- Što je varijabilnost podataka **veća**, to je **reprezentativnost** prosjeka **manja** i obrnuto.

Varijanca, standardna devijacija i koeficijent varijacije

- Zbroj odstupanja vrijednosti numeričke varijable od prosjeka jednak je nuli
- Stoga se za izračun varijance koriste kvadrati odstupanja.
- Kvadriranjem se “poništava” predznak, a istovremeno, kvadriranjem veća vrijednost odstupanja dobiva veći apsolutni udio u izračunu varijance.
- **Varijanca negrupiranih podatka** definira se kao

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}.$$

Varijanca, standardna devijacija i koeficijent varijacije

- Za **distribuciju frekvencija, varijanca** je

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{x})^2}{\sum_{i=1}^{k} f_i},$$

- Varijanca kao **mjera disperzije** izražena je u **kvadratima mjernih jedinica varijable**.

$$\sigma = +\sqrt{\sigma^2},$$

- **Stoga je potrebno izračunati standardnu devijaciju σ , kao pozitivni drugi korijen iz varijance.**
- **Standardna devijacija izražava prosječno odstupanje vrijednosti varijable od aritmetičke sredine izraženo u mjernim jedinicama varijable (apsolutna mjera).**

Varijanca, standardna devijacija i koeficijent varijacije

- Ako se **standardna devijacija podijeli s aritmetičkom sredinom**, a omjer pomnoži s posto, dobiveni pokazatelj je **koeficijent varijacije** V i relativna je mjera disperzije, izražena u %.

$$V = \frac{\sigma}{\overline{x}} \cdot 100\ \%.$$

- Standardna devijacija interpretira se kao **prosječno apsolutno odstupanje**, a koeficijent varijacije kao **prosječno relativno odstupanje od prosjeka**.

Standardizirano obilježje

- **Linearna transformacija** vrijednosti numeričke varijable x_i oblika

$$z_i = \frac{x_i - \overline{x}}{\sigma}, \qquad i = 1, 2, \ldots, N$$

- naziva se **standardizacija**
- Standardizirana vrijednost varijable višestruko se koristi u **empirijskim istraživanjima**.
- Ona **pokazuje relativni položaj** vrijednosti numeričke varijable u nizu, odnosno pokazuje za koliko standardnih devijacija vrijednost numeričke varijable odstupa od aritmetičke sredine.
- Osim toga, na temelju standardiziranih vrijednosti moguće je **usporediti relativni položaj podataka u različitim numeričkim nizovima (s različitim mjernim jedinicama i na različitim razinama vrijednosti**).

Empirijsko pravilo i Čebiševljevo pravilo

- Ukoliko su vrijednosti numeričke varijable raspoređene po simetričnoj zvonolikoj unimodalnoj distribuciji (normalnoj distribuciji) primjenjuje se empirijsko pravilo

Empirijsko pravilo i Čebiševljevo pravilo

- **Za distribuciju nepoznatog oblika**, proporcija podataka iz određenog intervala određuje se pomoću **Čebiševljevog pravila** ili Čebiševljeve nejednakosti.

Mjere oblika distribucije: mjere asimetrije i mjera zaobljenosti

- Mjerama asimetrije, odnosno mjerama oblika distribucije, mjeri se način rasporeda podataka oko neke srednje vrijednosti.
- Uglavnom je to aritmetička sredina.

Položaj srednjih vrijednosti

Položaj srednjih vrijednosti

Mjera zaobljenosti

- Mjerom zaobljenosti a_4 mjeri se zaobljenost unimodalnog vrha (simetrične ili približno simetrične) distribucije i **mjera je "debljine" odnosno "težine" repova distribucije**.
- Koeficijent a_4 je potpuna mjera. Izračunava se za **negrupirane podatke** i za **distribuciju frekvencija** prema izrazima:

Mjere koncentracije

- Mjere koncentracije su pokazatelji načina razdiobe totala (zbroja vrijednosti numeričke varijable) po jedinicama niza ili po modalitetima kvalitativnih ili vrijednostima kvantitativnih varijabli.