Osnove statistike u demografiji

Predavanje 6

Uvod u teoriju vjerojatnosti

- Koncept slučajnosti (engl. randomness) je ključan u svim područjima primijenjene statistike pa tako i u demografskim istraživanjima.
- Neki događaj je slučajan ukoliko se ne može predvidjeti sa potpunom sigurnošću
- Deterministički događaji
- Obično nakon sklapanja braka očekujemo da će bračni par dobiti dijete
- Međutim, ne znamo hoće li i kada uistinu imati djecu
- Koliko će imati djece?
- Hoće li se razvesti?

Uvod u teoriju vjerojatnosti

- Vjerojatnost: početno povezana s igrama na sreću
- Teorija vjerojatnosti je osnova za izučavanje probabilističkih modela, koji čine podlogu za primjenu metoda inferencijalne statistike
- Donošenje zaključaka na osnovi ograničenog podskupa podataka izabranog iz populacije uzrok je neizvjesnosti, koja se uz pomoć teorije vjerojatnosti može kvantificirati
- Kako bi se vjerojatnost mogla kvantificirati ključno je opisati okolnosti u kojima se vjerojatnost pojavljuje
- Stoga je potrebno definirati:
    - Slučajni eksperiment
    - Slučajni događaj

Slučajni eksperiment

- Postupak koji je moguće ponavljati proizvoljan broj puta, a ima barem dva ishoda čiji je nastup u bilo kojem od pokušaja neizvjestan
- Ishodi se opisuju riječima, brojevima ili riječima i brojevima
- Prostor slučajnih događaja ili prostor uzorka (S):
  - skup svih ishoda slučajnog eksperimenta
  - Diskretan ili kontinuiran

Slučajni događaj

- Svaki podskup prostora slučajnih događaja S je slučajni događaj

Slučajni događaj

- Do novih se događaja dolazi skupovnim operacijama (unija, presjek, komplement)
- $A \cap B$
- $A \cup B$
- $\bar{A}$
- Isključivi događaji: $A \cap B = \emptyset$
- Grafički se slučajni događaji prikazuju Vennovim dijagramima

Vjerojatnost slučajnog događaja

- Mjera mogućnosti ostvarenja slučajnog događaja
- Poprima vrijednosti iz intervala [0,1].
- Definicije vjerojatnosti:
  - Klasična (vjerojatnost *a priori*)
  - Statistička (vjerojatnost *a posteriori*)
  - Subjektivna

Klasična definicija vjerojatnosti

- Slučajni eksperiment ima konačan broj jednako mogućih ishoda N
- N_A je broj ishoda u kojima se ostvaruje događaj A

$$P(A) = \frac{N_A}{N}.$$

- Naziva se još i vjerojatnost a priori jer se može izračunati i bez izvođenja eksperimenta
- Ograničenja:
  - broj ishoda slučajnog eksperimenta mora biti konačan
  - pretpostavka da su svi ishodi slučajnog eksperimenta jednako mogući

Statistička definicija vjerojatnosti

- Ako slučajni eksperiment završava s beskonačno mnogo ishoda ili ishodi nisu jednako mogući, koristi se statistička definicija vjerojatnosti
- Empirijska vjerojatnost, jer se računa polazeći od opaženih podataka.
- Slučajni eksperiment moguće ponavljati proizvoljan broj puta u nepromijenjenim uvjetima, pri čemu su pojedini pokušaji međusobno neovisni, tada se vjerojatnost događaja A definira kao granična vrijednost njegove relativne frekvencije kad broj ponavljanja slučajnog eksperimenta neomeđeno raste

$$P(A) = p \lim_{n \to \infty} \left(\frac{n_A}{n} \right)$$

- Naziva se i vjerojatnost *a posteriori* jer se može izračunati tek nakon što se izvede slučajni eksperiment

Primjer 5.4. Statistička definicija vjerojatnosti

- Bahovec i Erjavec (2015.) str. 181

Subjektivna vjerojatnost

- Osobno uvjerenje pojedinca o stupnju neizvjesnosti ishoda slučajnog događaja
- Izražena je kao broj između 0 i 1
- Određuje se u praksi u slučaju kad se vjerojatnost ne može odrediti ni jednom od ranije opisanih definicija vjerojatnosti
- Često se temelji na iskustvu, osobnom uvjerenju i analizi određene situacije
- Početnu vjerojatnost (engl. *prior*) nastoji se revidirati nakon što nova informacija postane dostupna.
- Revidirane vjerojatnosti se nazivaju posteriornim vjerojatnostima

Vjerojatnosti

- Stabilnost relativnih frekvencija: s povećanjem veličine N omjer $p = \frac{f}{N}$ se približava konstanti
- Priroda i velik broj događaja u stvarnom životu se 'ravnaju' prema određenim zakonitostima iz teorije vjerojatnosti
- Događaji su slučajni (stohastički), ali nisu kaotični.
- Iako su sami događaji na razini pojedinca potpuno nepredvidivi (slučajni), njihovo agregiranje ukazuje na određene pravilnosti
- U demografiji postoji velik broj demografskih pravilnosti (engl. demographic regularity)

Vjerojatnosti

- Proporcija živorođene djece prema spolu

Table. 2.1. Frequency of boy and girl births in Poland, 1927-32

Year	Boys	Girls	Both sexes	Proportions	
				Boys	Girls
1927	496,544	462,189	958,733	0.518	0.482
1928	513,654	477,339	990,993	0.518	0.482
1929	514,765	479,336	994,101	0.518	0.482
1930	528,072	494,739	1,022,811	0.516	0.484
1931	496,986	467,587	964,573	0.515	0.485
1932	482,431	452,232	934,663	0.516	0.484

Source: Fisz, 1963, p. 4.

Vjerojatnosti

- Demografske pravilnosti (engl. demographic regularity)
- Npr. Životni vijek žena je veći
- Važno u osiguranju (određivanje premije)

Definicija slučajne varijable

- Kako bi se svi ishodi slučajnog eksperimenta mogli izraziti brojčano, uvodi se preslikavanje X iz skupa S u skup realnih brojeva R.
- Takva se funkcija zove slučajna varijabla (engl. *random variable*).
- Označavaju se velikim slovima *X*, *Y*, *Z*,. . . , dok se njihove vrijednosti označavaju malim slovima x, y, z,. . .
- Dijele se na:
  - diskretne i
  - kontinuirane

Diskretna slučajna varijabla

- Varijabla *X* je diskretna slučajna varijabla ako poprima konačno ili prebrojivo mnogo vrijednosti s vjerojatnostima: $P(X = x_i) = p(x_i)$
- Uz uvjet: $p(x_i) > 0 \quad \text{i} \quad \sum_i p(x_i) = 1.$
- Diskretna funkcija gustoće vjerojatnosti $\{(x_i, p(x_i), i = 1, 2, \cdots\}$
- Distribucija vjerojatnosti diskretne slučajne varijable
- Funkcija distribucije vjerojatnosti $F(x) = P(X \leqslant x)$
- Vrijednost funkcije distribucije u određenoj točki računa se

$$F(x_j) = P(X \leqslant x_j) = p(x_1) + p(x_2) + \cdots + p(x_j)$$

Kontinuirana slučajna varijabla

- Poprima neprebrojivo mnogo vrijednosti na skupu realnih brojeva R
- Njezina vjerojatnosna svojstva opisuju se s pomoću funkcije gustoće vjerojatnosti $f(x)$

Karakteristike distribucija vjerojatnosti

- Vjerojatnosna svojstva slučajne varijable su potpuno određena ako je poznata njena funkcija gustoće vjerojatnosti ili ekvivalentno ako je poznata funkcija distribucije vjerojatnosti
- Ponekad se vjerojatnosna svojstva slučajne varijable opisuju pomoću karakteristika koje se nazivaju momentima slučajne varijable
- Najčešće korišteni momenti su očekivana vrijednost slučajne varijable i varijanca

Očekivana vrijednost slučajne varijable

Varijanca slučajne varijable

Odnos varijabli

- Uvjetna vjerojatnost
- Statistička nezavisnost
- Kovarijanca i korelacija

Modeli distribucija vjerojatnosti

- Distribucije vjerojatnosti za koje je poznat analitički izraz i sve njihove karakteristike
- S pomoću njih se mogu dobro aproksimirati realne pojave u vjerojatnosnom okruženju
- Razlikujemo teorijske distribucije vjerojatnosti
  - Diskretnih slučajnih varijabli
  - Kontinuiranih slučajnih varijabli

Binomna distribucija

- Binomna slučajna varijabla X je slučajna varijabla koja broji koliko se puta ostvario događaj A u n ponavljanja Bernoullijeva eksperimenta
- Uzastopni pokušaji su nezavisni
- Vjerojatnost nastupa događaja A je jednaka u svakom pokušaju

$$X \sim B(n;p)$$

Poissonova distribucija

- Poissonova varijabla je diskretna slučajna varijabla koja poprima vrijednosti 0, 1, 2, . . .
- Poissonova distribucija vjerojatnosti

Normalna ili Gaussova distribucija

- Jedna od najvažnijih kontinuiranih distribucija vjerojatnosti $X \sim N(\mu, \sigma^2)$.
- Funkcija gustoće vjerojatnosti

Standardizirana normalna distribucija

Kontinuirane distribucije vjerojatnosti usko povezane s normalnom distribucijom

- hi-kvadrat distribucija,
- *t* -distribucija i
- *F* -distribucija.

Hi-kvadrat distribucija ($\chi 2$ -distribucija)

- Poprima samo nenegativne vrijednosti
- Pozitivno je asimetrična, $a_3 > 0$

Studentova t -distribucija

F-distribucija