

Osnovni pojmi statistike

- **Statistika** je veda, ki proučuje množične pojave in povezave med njimi. Vključuje:
 - opazovanje
 - zbiranje
 - obdelava in analiza podatkov
- **Populacija** je množica, ki jo statistično proučujemo.
- **Statisticna enota** ali **enota** je posamezni element populacije.
- Omejitev pri raziskavah sta **čas in denar**, zato raziskav ne delamo na vsej populaciji.
- **Vzorec** je končna podmnožica populacije, katere elementi kar najbolj predstavljajo značilnosti cele populacije (**reprezentativen vzorec**). Če imajo vsi elementi populacije enako možnost, da bodo izbrani v vzorec, je le-ta **slučajen vzorec**. **Moč vzorca** označimo z **N**.
- Če gledamo neko lastnost na populaciji, ji rečemo **statistični parameter**. Če gledamo to lastnost na statistični enoti, ji rečemo **statistična spremenljivka** ali **podatek** ali **statistični znak**.
- Podatki so:
 - **opisni** (barva, mesec rojstva, državljanstvo)
 - **številski** ali **numerični** - **celoštevilski (diskretni)**
 - **zvezni**
- Na podlagi vzorca statistiki naredijo **statistično oceno** za vso populacijo. Stopnja verjetnosti za celotno populacijo je odvisna od izbranega vzorca in metode raziskave.

PRIMER 1:

Statistični parameter: drugi tuj jezik – to lastnost gledamo na vsej populaciji

Populacija: dijaki vseh slovenskih šol (140 šol)

Vzorec: Dijaki samo 15-h šol (**ČAS IN DENAR!**). Samo 1 šola ni reprezentativen vzorec.

Statisticna enota: dijaki in dijakinje

Opisna statistična spremenljivka (= podatek): drugi tuj jezik (ITAL, NEM, FRAN, ANG)

Numerična spremenljivka: število ur učenja tujega jezika na teden

PRIMER 2: (Populacija, Enota, Parameter, Spremenljivka)

Starostna skupina	Skupaj	Moški	Ženske
0-9	190 286	97 844	92 424
10-19	257 091	131 640	125 451
20-19	293 619	149 936	143 683
30-39	295 959	147 120	148 839
40-49	310 428	156 604	153 824
50-59	244 195	121 525	122 670
60-69	201 633	91 931	109 702

70-79	138 609	49 208	89 401
80-89	40 497	11 397	29 100
90-99	6 361	1 441	4 920
100 in več	69	11	58

Table: Prebivalci RS po starostnih skupinah in spolu, 31.12.2000

Populacija: Vsi prebivalci RS,

Enota: Vsak prebivalec RS

Parametri (lastnost populacije!) – število vseh prebivalcev skupaj, število prebivalcev po posameznih starostnih skupinah

Spremenljivke: (na enoti!) starost in spol : vsakemu prebivalcu lahko to določiš

- Starost = **zvezna** številska spremenljivka, njena vrednost je izražena številčno v dopoljenih letih
- Spol = opisna spremenljivka (vrednost izražena opisno)

Urejanje in grupiranje podatkov

Podatke moramo urediti

Urejanje **numeričnih** podatkov: **uredimo jih po velikosti** - naredimo **rangirno vrsto**

Primer 3:

V družini Novak imajo 15 vnukov. Njihove starosti (v letih) so

10, 2, 8, 17, 12, 17, 8, 28, 4, 2, 19, 22, 8, 1, 15

Rangirna vrsta: _____

So podatki zvezni ali diskretni?

Za diskretne statistične spremenljivke definiramo **frekvenco f**. Frekvenca statistične spremenljivke je število diskretnih statističnih enot iste vrednosti.

Primer 3: Vnuki družine Novak. Frekvenca vnukov, starih 17 let je _____

Če je **diskretnih** podatkov zelo **veliko** ali če so podatki **zvezni** => združujemo jih v skupine ali frekvenčne razrede. (celotem interval vrednosti statistične spremenljivke razdelimo na več manjših)

Glej zgornji **Primer 2** – starostni frekvenčni razredi

S tem združevanjem v frekvenčne razrede delamo **grupiranje** (združevanje) podatkov.

Torej: če imamo veliko podatkov, le-te združujemo (grupiramo):

Numerične podatke – grupiramo v **frekvenčne razrede**

Opisne podatke – združujemo v **skupine**

Glej zgornji **Primer 2:** razvrščanje po spolu – dve skupini, razvrščanje po starosti – 11 frekvenčnih razredov

POMEMBNO pri grupiranju:

- Razredi/skupine morajo biti **enolično določeni (= enolično opredeljeni)** t.j. vsaka statistična enote pade v natanko določen (samo en!!) frekvenčni razred/skupino.
- Skupine morajo biti **enovite (= homogene)** t.j. vključujejo sorodne enote (vsak starostni razred iz približno enako starih ljudi).
- **Bistvo grupiranja:** vsak podatek mora pasti v natanko en razred, zato so razredi ali navzgor odprti [.,.) ali navzdol odprti (.,].
- Vse podatke zajamemo z določenim st. frekvenčnih razredov, ki so običajno enako široki, a ni nujno!

x_{\max} – podatek z največjo vrednostjo,

x_{\min} – podatek z najmanjšo vrednostjo

enako široki frekvenčni razredi: širina razreda = $(x_{\max} - x_{\min}) / \text{št.razredov}$

- **Spodnja meja** k-tega frekvenčnega razreda s_k
- **Zgornja meja** k-tega frekvenčnega razreda z_k
- **Širina k-tega frekvenčnega razreda** $i_k = z_k - s_k$
- **Sredina razreda** $x_k = \frac{1}{2}(s_k + z_k)$
- **Tabela frekvenčne porazdelitve** je razdelitev vseh podatkov (enot populacije) v 5-20 frekvenčnih razredov. Posameznemu frekvenčnemu pripišem število posameznih enot s to vrednostjo (= **frekvenca f**).
- **Absolutna frekvenca** razreda f_k (ali tudi kar **frekvenca**) je število podatkov (enot) v danem frekvenčnem razredu. Vsota vseh absolutnih frekvenc = številu vseh enot N.
- **Relativna frekvenca** razreda $f'_k = \frac{f_k}{N}$ je relativni delež razreda znotraj vzorca (N je število vseh enot v vzorcu), podajamo jo v %. Vsota vseh relat. frekvenc = 1 = 100%.

ZGLED: Frekvenčne porazdelitve in relativne frekvenčne porazdelitve 84-tih ponesrečencev.

Starostni razred v letih (frekvenčni razred)	Spodnja meja frekvenč. razreda s_k	Zgornja meja frekv. razreda z_k	Širina frekv. razreda $i_k = z_k - s_k$	Sredina frekv. razreda $x_k = \frac{1}{2}(s_k + z_k)$	(Absolutna) frekvenca f (Število ponesreč. f _k)	Relativna frekvenca f' = f _k /N (v %)
0 - 9	0	10	10	5	12	12/84 x 100 = 14.3
10 - 19	10	20	10	15	22	26.2
20 - 29	20	30	10	25	14	16.7
30 - 39	30	40	10	35	9	10.7
40 - 49	40	50	10	45	6	7.1
50 - 59	50	60	10	55	13	15.5
60 - 69	60	70	10	65	0.95 x 84 = 8	9.5

N = 84 Σ