



POSLOVNA STATISTIKA

Zašto statistika. Podaci. Deskriptivna statistika

1

ZAŠTO STATISTIKA?

- Zamislimo da smo u situaciji da odlučujemo da li će se lansirati neki novi proizvod naše kompanije.
- Da bismo doneli informisanu odluku, potrebno je prikupiti informacije o troškovima proizvodnje, da li konkurencija već nudi tako nešto, ispitati tržište da li postoji zahtev za tim proizvodom,...
- Zatim te podatke treba analizirati i interpretirati, kako bi doneli odgovarajuću odluku.
- U ovakvim situacijama od velikog značaja biće naše statističke veštine.

PRIMERI PRIMENE

- Glavni pokretač zahteva za statističkim znanjem su tehnologije dostupno za snimanje podataka.
- Primeri uključuju tehnologiju koju Google koristi da bi pratio kako korisnici Interneta pristupaju veb lokacijama. i zatim koristi ove podatke za sortiranje i određivanje prioriteta rezultata za budući Internet pretrage kao i u izboru ponude reklama.
- Lanci prodavnica koriste kartice potrošača za preuzimanje podataka o svakoj kupovini i koriste ih za upravljanje distribucijom proizvoda, donošenje odluka o marketingu i prodaji, i praćenje dnevne, pa čak i prodaje na satu.
- Policijske uprave prikupljaju i koriste podatke da stanovnicima grada pružite mape koje prenose informacije o počinjenim zločinima i njihovoj lokaciji.

ŠTA JE STATISTIKA?

STATISTIKA je nauka o prikupljanju, organizovanju, predstavljanju, analiziranju i tumačenje podataka radi lakšeg donošenja efikasnijih odluka.

- Statistika je mnogo više od predstavljanja numeričkih informacija, ona se odnosi na prikupljanje i obradu informacija da podstakne diskusiju i dodatna pitanja i pruži osnovu za donošenje odluka.

VRSTE STATISTIKE

- Kada koristimo statistiku za generisanje informacija za donošenje odluka iz podataka, koristimo:
 - deskriptivnu (opisnu) statistiku,
 - inferencijalnu (analitičku) statistiku.
- Njihova primena zavisi od pojave koja se ispituje i vrstu dostupnih podataka.

DESKRIPTIVNA STATISTIKA

- Velike količine **nesređenih podataka**, kao:
 - plate zaposlenih,
 - podaci o popisu stanovništva,
 - glasovi na izborima

nam ne mogu dati tražene informacije.

- Međutim, deskriptivna statistika nam pomaže da te podatke organizujemo u smislenu formu. Definišemo je na sledeći način:

DESKRIPTIVNA STATISTIKA predstavlja metode organizovanja, rezimiranja i predstavljanja podataka na informativan način.

INFERENCIJALNA STATISTIKA - POPULACIJA

- Ponekad se odluke moraju donositi na osnovu ograničenog skupa podataka.
- Ako bismo želeli da napravimo istraživanje u vezi sa zadovoljstvom klijenata neke banke, jedna opcija je da se kontaktiraju svi klijenti te banke. U ovom slučaju ispitujemo celu **populaciju** klijenata banke.

POPULACIJA (ili generalni skup) je skup svih individua ili objekata od interesa ili merenja dobijenih od svih individua ili objekata od interesa.

INFERENCIJALNA STATISTIKA - UZORAK

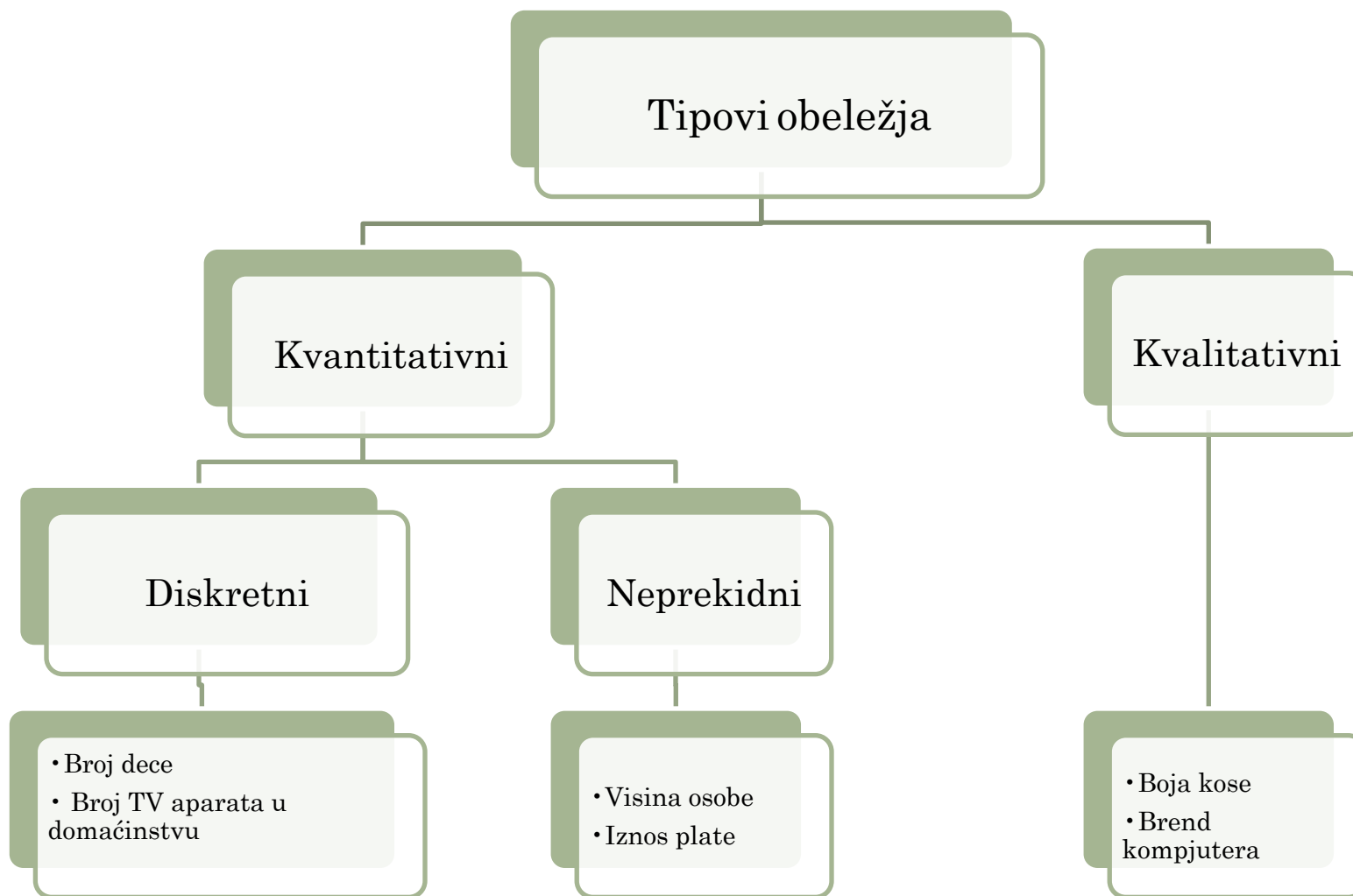
- U prethodnom slučaju morali bismo izdvojiti značajnu količinu ljudskih i finansijskih resursa za realizaciju istraživanja.
- Inferencijalna statistika nam omogućava da istraživanje izvedemo na podskupu populacije od interesa i da zatim izvedemo zaključak o celoj populaciji.
- Na ovaj način štedimo resurse, na račun smanjene preciznosti u oceni osobine koju istražujemo.

UZORAK je podskup (ili deo) populacije od interesa.

STATISTIČKI PODACI

- PODACI su skup činjenica koji su dobijeni kao rezultat posmatranja individualnih objekata.
- OBELEŽJE je karakteristika (osobina) koja se analizira.
- TIP obeležja:
 - Kvantitativno - vrednost obeležja se izražava brojevima (visina, težina, iznos plate,...), obično je rezultat merenja.
 - Kvalitativno – kada se opisuje nekim atributom (pol, boja kose, boja očiju, marka automobila,..)

TIPOVI OBELEŽJA



KVANTITATIVNA I KVALITATIVNA OBELEŽJA

- Kvantitativna obeležja su numeričkog tipa, kao što je prikazano na prethodnom grafikonu, mogu biti
 - Diskretna, obično je dobijaju prebrajanjem: broj dece, broj soba u kući, broj prodatih automobila u mesecu (ne može se prodati 35.68 automobila)
 - Neprekidna, su obično rezultat merenja: kvadratura kuće, težina jabuka u gajbi, prosečna ocena studenta,...
- Kvalitativna obeležja predstavljaju kategorije, za koje obično vršimo prebrajanja koliko ima posmatranja u svakoj od kategorija. Često ih izražavamo u procentima. Na primer. Koji procenat ljudi ima plave oči, a koji ima braon.

NIVOI MERENJA

- Podaci se mogu klasifikovati prema nivoima merenja.
- Nivo merenja određuje kako se podaci trebaju sumirati i predstaviti.
- Takođe indukuje koje vrste vrstu statističkih analiza koje se mogu izvesti na datim podacima.
- Razlikujemo:
 - nominalni,
 - ordinalni,
 - intervalni i
 - racio nivo merenja.

NOMINALNI NIVO

- Podaci dobijeni na ovom nivou predstavljaju se kategorijama ili imenima.
- Ne mogu se porediti.
- Mogu se samo klasifikovati i prebrajati.
- Npr. vrsta voćnih karamela u kesici, broj na dresu sportiste (jasno je da nema smisla vršiti matematičke operacije sa brojevima dresova).

Vrsta	Procenat
Jagoda	33.33%
Višnja	26.67%
Limun	17.78%
Narandža	13.33%
Kajsija	8.89%

ORDINALNI

- Vrednosti prikupljene na ovom nivou mogu se samo rangirati i prebrajati.
- Nije smisleno vršiti matematičke operacije na ovim podacima (ne može student biti na 3.25 godini studija)
- Primer su:
 - godina studija
 - spisak zemalja rangiranih po sreći njihovih stanovnika

Rangiranje	Zemlja
1.	Finska
2.	Danska
3.	Norveška
4.	Island
5.	Švajcarska
6.	Holandija
7.	Švedska
8.	Novi Zeland
9.	Kanada
10.	Australija

INTERVALNI

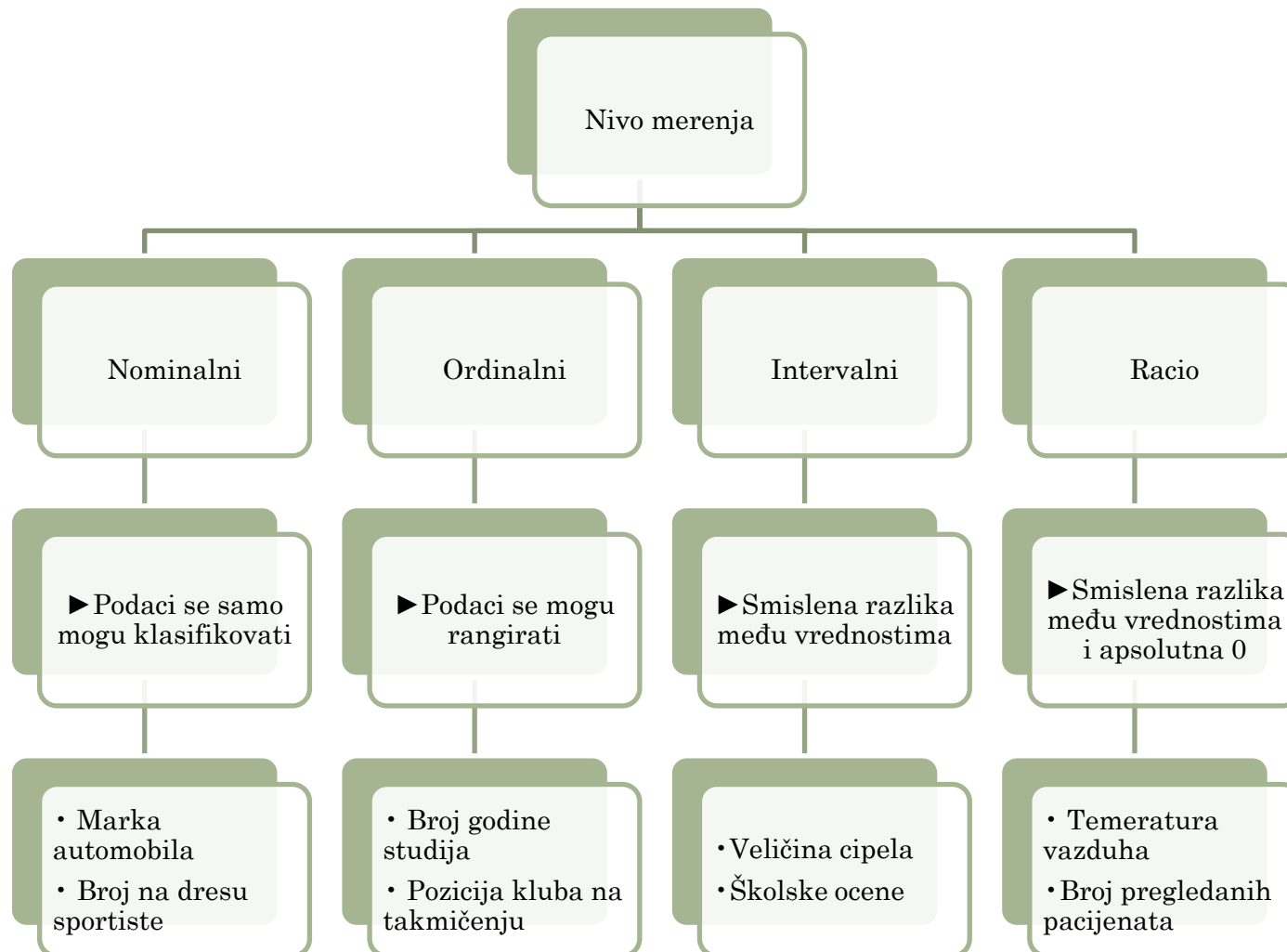
- Za podatke merene na intervalnom nivou razlika na skali imaju značenje, tj. u saglasnosti su sa merenom osobinom.
- Merenje je zasnovano na skali sa poznatom jedinicom mere
- Nema apsolutnu nulu, ako je i postojena, ona je proizvoljna
- Primer je veličina cipela

EU veličine	UK veličine	US veličine	Dužina stopala (cm)
34,5	1,5	4	21
35	2	4,5	21,5
35,5	2,5	5	22
36	3	5,5	22,5
36,5	3,5	6	23
37,5	4	6,5	23,5
38	4,5	7	24
38,5	5	7,5	24,5
39	5,5	8	25
40	6	8,5	25,5
40,5	6,5	9	26
41	7	9,5	26,5

RACIO

- Racio nivo merenja ima sve osobine intervalne skale i još ima apsolutnu 0. To znači da je poznata skala i jedinica mere i da 0 ima smislenu interpretaciju.
- Primeri: količina ušteđevine, broj proizvedenih jedinica, promene cena akcija, udaljenosti između ekspozitura,...

NIVOI MERENJA



PITANJE:

- 1. Koji je nivo merenja za sledeća obeležja?
 - a. IQ studenta.
 - b. Udaljenost od škole
 - c. Brojevi na dresovima fudbalskog kluba
 - d. Grad rođenja studenta
 - e. Godina studija studenta
 - f. Broj sati koje student provede učeći na nedeljnom nivou

DESKRIPTIVNA STATISTIKA

Formiranje distribucija frekvencije

FORMIRANJE DISTRIBUCIJA FREKVENCije

- Osnova statističkih ispitivanja su empirijski podaci .
- Te vrednosti treba urediti tako da daju adekvatnu sliku o prirodi pojave koja se izučava
- **Primer1:** U toku 30 dana beležena je prodaja mobilnih telefona i dobijeni su sledeći podaci:

dan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
broj prodatih mobilnih telefona	10	12	8	10	12	11	9	10	15	9	11	9	11	11	13
dan	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
broj prodatih mobilnih telefona	9	10	12	11	13	11	14	10	12	8	13	13	11	10	12

NEUREĐENI STATISTIČKI PODACI

- **Neuređeni statistički podaci** predstavljaju niz numeričkih vrednosti dobijenih kao rezultat posmatranja svake jedinice u izabranom skupu. U slučaju našeg primera, neuređeni statistički podaci su:
- 10, 12, 8, 10, 12, 11, 9, 10, 15, 9, 11, 9, 11, 11, 13, 9, 10, 12, 11, 13, 11, 14, 10, 12, 8, 13, 13, 11, 10, 12

UREĐENA STATISTIČKA SERIJA

- **Uređena statistička serija** je niz numeričkih podataka sređenih po veličini od najmanje do najveće vrednosti. Uređena statistička serija za podatke iz primera 1 je:
- 8, 8, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 14, 15
- Bez obzira na svoje prednosti (uočavanje minimuma, maksimuma, lakoća grupisanja...) nije zahvalna za prikazivanje velikog broja podataka.

KLASA

- **Klasa** je grupa istih vrednosti obeležja.
- Ako su obeležja kvalitativnog tipa, svaka kategorija predstavlja jednu klasu.
- Ako su vrednosti obeležja celi brojevi, klase mogu biti izolovane tačke, tj. različite vrednosti obeležja. U primeru prodaje mobilnih telefona bilo je 8 klasa.

KLASA

- Kada su vrednosti obeležja realni brojevi, koristimo klase u obliku intervala: x_1-x_2, x_2-x_3, \dots
- U tom slučaju gornja i donja granica svakog intervala moraju biti nedvosmisleno određeni, tako da svaka vrednost može pripadati tačno jednoj klasi
- Pri čemu ćemo podrazumevati da klasa x_1-x_2 predstavlja interval $[x_1, x_2)$ tj., Donja granica je uključena u interval, dok gornja nije.
- Obično se biraju intervali jednake širine.
- U skladu sa tim, oznaka intervala od npr. 11-12, 12-13, ... podrazumeva da 11.9 pripada prvom, a 12 drugom intervalu.
- Klasni intervali mogu biti otvoreni i zatvoreni.
- Kod zatvorenog tipa poznata je donja i gornja granica, kod otvorenog donja ili gornja granica nisu poznate: „manje od 5“ ili „više od 5“.
- Intervalne klase se mogu formirati i za cele vrednosti obeležja.

KLASNA SREDINA

- Klasna sredina je sredina klasnog intervala . U obračunima i nekim grafičkim prikazima, klasna sredina zamenjuje ceo interval.

klasni interval	sredina intervala
1 - 3	2
3 - 5	4
5 - 7	6
7 - 9	8
9 - 11	10

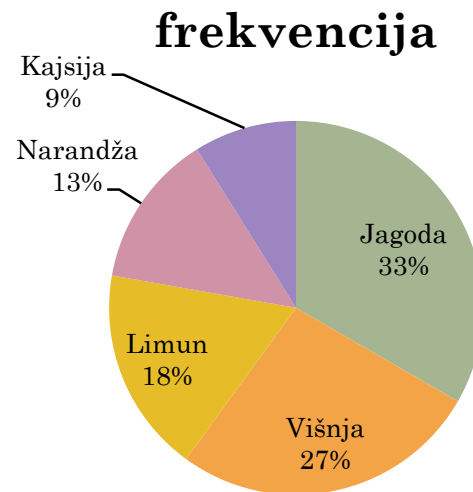
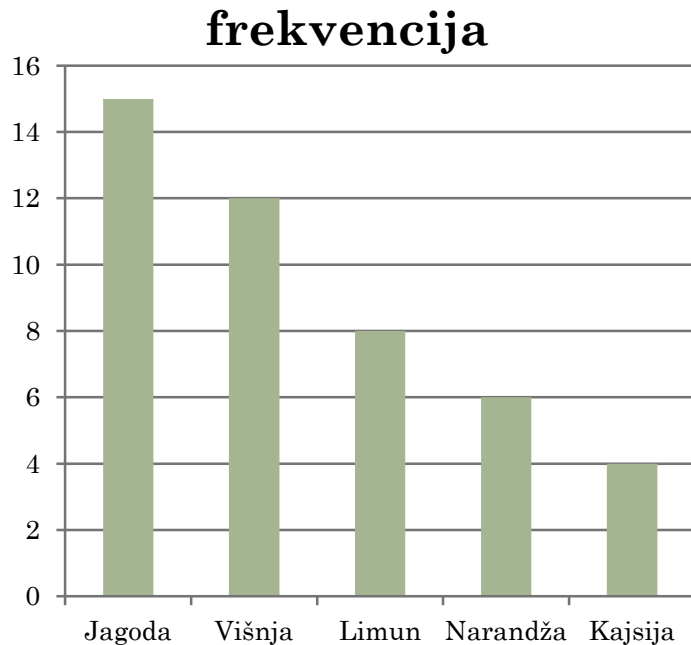
KREIRANJE TABLICE FREKVENCIJA

- Tablica frekvencija služi da se prikažu različite klase i broj posmatranja za svaku od njih.

- | Vrsta | frekvencija |
|----------|-------------|
| Jagoda | 15 |
| Višnja | 12 |
| Limun | 8 |
| Narandža | 6 |
| Kajsija | 4 |
| Ukupno: | 45 |

VIZUELNI PRIKAZ KVALITATIVNIH PODATAKA

- Vrlo često se kvalitativni podaci prikazuju u obliku stubića ili grafikona koji prikazuju udeo, tzv. PIE



KLASNA SREDINA

- Klasna sredina je sredina klasnog intervala, dobija se kao aritmetička sredina donje i gornje granice intervala.
- U obračunima, klasna sredina zamenjuje ceo interval.

klasni interval	sredina intervala
1 - 3	2
3 - 5	4
5 - 7	6
7 - 9	8
9 - 11	10

APSOLUTNA FREKVENCIJA

- **Apsolutna frekvencija** je broj podataka koji pripada jednoj klasi, apsolutna frekvencija i -te klase označava se sa f_i i važi

$$f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i = n$$

- gde je n broj jedinica u statističkoj seriji, a k broj klasa.

RELATIVNE FREKVENCije

- Relativne frekvencije su definisane sa

$$p_i = \frac{f_i}{n}$$

- Dakle, predstavljaju količnik apsolutne frekvencije i ukupnog broja jedinica u statističkoj seriji.
- Prikazuje frekvenciju klase u odnosu na ukupni broj posmatranja
- Važi:

$$\sum_{i=1}^k p_i = 1$$

RELATIVNE FREKVENCije U PROCENTIMA

- Kada se relativne frekvencije pomnože sa 100 dobijaju se frekvencije izražene u procentima

$$p_i = \frac{f_i}{n} \cdot 100$$

- Važi

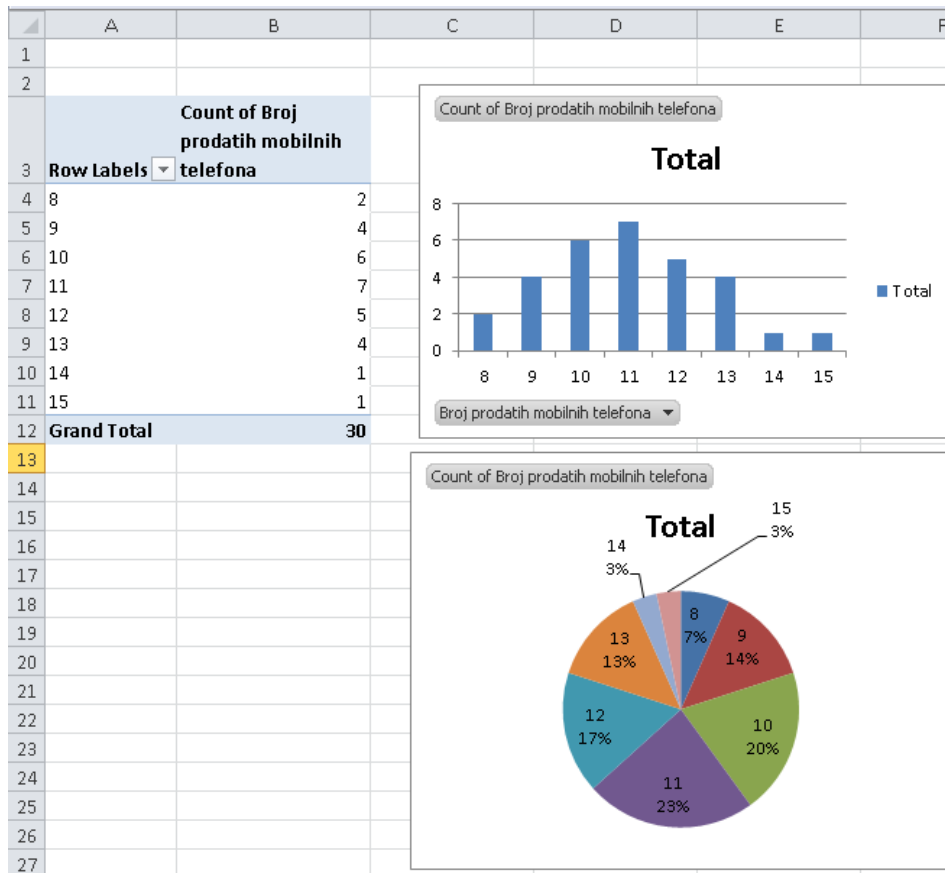
$$\sum_{i=1}^k p_i = 100$$

DISTRIBUCIJA FREKVENCIJA

- DISTRIBUCIJA FREKVENCIJA je raspodela frekvencija prema klasama i pokazuje koliko jedinica statističke serije pripada određenoj klasi. Može biti prikazana tabelarno ili grafički

Broj prodatih telefona	Frekvencija (apsolutna)	Relativna frekvencija	Relativna frekvencija u procentima
8	2	0.067	6.67%
9	4	0.133	13.33%
10	6	0.200	20.00%
11	7	0.233	23.33%
12	5	0.167	16.67%
13	4	0.133	13.33%
14	1	0.033	3.33%
15	1	0.033	3.33%
Ukupno	30	1	100%

GRAFIČKI PRIKAZ DISTRIBUCIJE FREKVENCIJA



- Oba grafikona, pita i stubići služe da ilustruju tabele apsolutne i relativne frekvencije.
- Pita se koristi kada želimo da uporedimo relativne frekvencije, tj da vidimo udeo svake klase u celini.
- Stubići su korisni kada želimo da poredimo apsolutne frekvencije.

KLASE U FORMI INTERVALA

- U prethodnim primerima smo prikazali distribucije frekvencija kada se radi o kvalitativnim podacima (karamele u kesici) gde su kategorije predstavljale klase.
- U primeru koji opisuje prodaju mobilnih telefona po danima, podaci su bili celobrojni. Tu smo razlikovali 8 klasa, za svaku celobrojnu vrednost.
- U realnim situacijama, imaćemo i vrednosti koje su realni brojevi, zbog čega je potrebno koristiti klase u formi intervala.
- Serije statističkih podataka koji su celi brojevi, kada su u pitanju na primer cene, plate, merenja se organizuju koristeći klase u formi intervala.

KLASE U FORMI INTERVALA-PRIMER

- **Primer 2:** Posmatramo neuređenu seriju podataka o cenama prodatih polovnih automobila u toku prethodno meseca izraženih u valuti EUR
- Prvo što vidimo je da je minimum 588, a maksimum 6584.

2774	4296	4402	1926	1640	4460	6086	5168	4740
3508	4414	1992	2596	2532	4682	2118	5332	5274
3634	4504	5626	2820	3482	6584	3348	5982	2852
2080	2856	646	3106	3544	2216	3614	1868	5888
2546	3778	704	3296	3864	2590	4112	4126	4294
3058	2332	964	4142	4700	2688	4472	4166	3946
6164	2640	2288	4232	4844	3812	5856	5712	5004
3902	4530	2970	3000	4892	3904	2538	5978	1566
5384	2646	3018	3098	738	4140	3434	1820	3076
2412	3520	3276	4696	1956	4908	3594	3072	4678
2684	3838	3922	4996	2476	3212	3910	3914	5400
886	4714	4254	588	3636	3360	4398	4480	4444
1508	5732	4860	2230	3648	3654	4964	5390	5194
3242	1464	3408	2248	3814	3830	5402	2650	5484
1740	2928	3752	3064	3876	4168	6420	4500	3674
2348	3252	4020	3376	3880	5278	754	4558	5684
2824	3524	4330	3644	4394	1684	2440	5252	4868
3618	3830	4462	3794	5292	3926	2802	3002	3280
4830	4238	4778	4890	2922	4118	4350	3504	3642
3092	3532	670	5772	3462	4676	2236	4116	4974

1. KORAK: ODREĐIVANJE BROJA KLASA

- Koristan princip za određivanje broja klasa je pravilo „2 na k“ – ideja je da se odabere najmanje k za koje je zadovoljeno da je 2^k veća od broja posmatranja (n).
- Kako u našem slučaju imamo da je $n=180$,
 - $k=7$ ne može da se kristi jer je $2^7=128 < 180$,
 - Za $k=8$ imamo da je $2^8=256 > 180$, stoga ćemo podatke organizovati u 8 klasa

2. KORAK: ODREĐIVANJE ŠIRINE KLASNOG INTERVALA

- Širinu klasnog intervala određujemo po pravilu

$$l \geq \frac{\text{Maksimalna vrednost} - \text{Minimalna vrednost}}{k}$$

- U našem slučaju imamo

$$l \geq \frac{6584 - 588}{8} = 749.5$$

- Kako je pravilo, da širina intervala bude zaokružena na neki zgodan broj koji predstavlja umnožak od 10 ili 100, mi ćemo koristiti širinu
 - $l = 800$.

3. ODREĐIVANJE GRANICA KLASE I KREIRANJE KLASNOG INTERVALA

- Po dogovoru, pri prikazu interval podrazumevaćemo da je donja granica uključena, a gornja nije. Tj. $l - u$ znači interval $[l, u)$
- Budući da smo zaokružili širinu intervala na veći broj, sa osam klasa pokrivamo veći raspon nego što je potrebno.
 - Pokrivamo 8 klasa po 800, a to je 6400.
 - Stvarni opseg je $6584 - 588 = 5996$, dakle imamo viška $6400 - 5996 = 404$
 - Pošto pokrivamo opseg od minimuma do maksimuma, jasno je da ovaj višak treba raspodeliti na krajnje intervale (repove) $588 - (404/2) = 386$, što znači da ćemo kao donju granicu prvog intervala odrediti 400. Tako dobijamo klasne intervale prikazane u tabeli.

Klasni intervale
400 - 1200
1200 - 2000
2000 - 2800
2800 - 3600
3600 - 4400
4400 - 5200
5200 - 6000
6000 - 6800

4. KORAK: ODREŠIVANJE APSOLUTNIH FREKVENCIJA

- Sada se vrši razvrstavanje vrednosti po intervalima i određivanje frekvencija

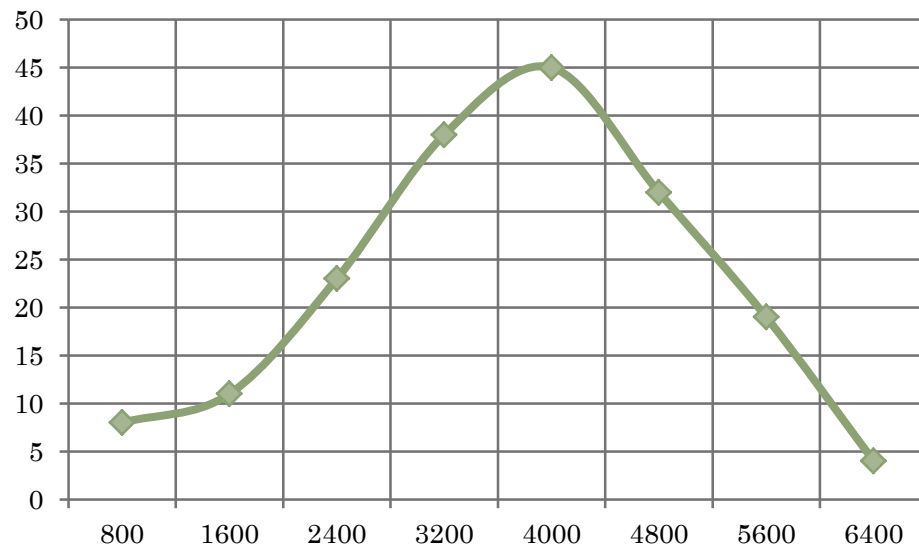
klasni interval	frekvencija	frekvencija	relativna frekvencija
400 - 1200	//////	8	0.044
1200 - 2000	//////	11	0.061
2000 - 2800	//////	23	0.128
2800 - 3600	//////	38	0.211
3600 - 4400	//////	45	0.250
4400 - 5200	//////	32	0.178
5200 - 6000	//////	19	0.106
6000 - 6800	////	4	0.022
Ukupno		180	1.000

POLIGON I HISTOGRAM

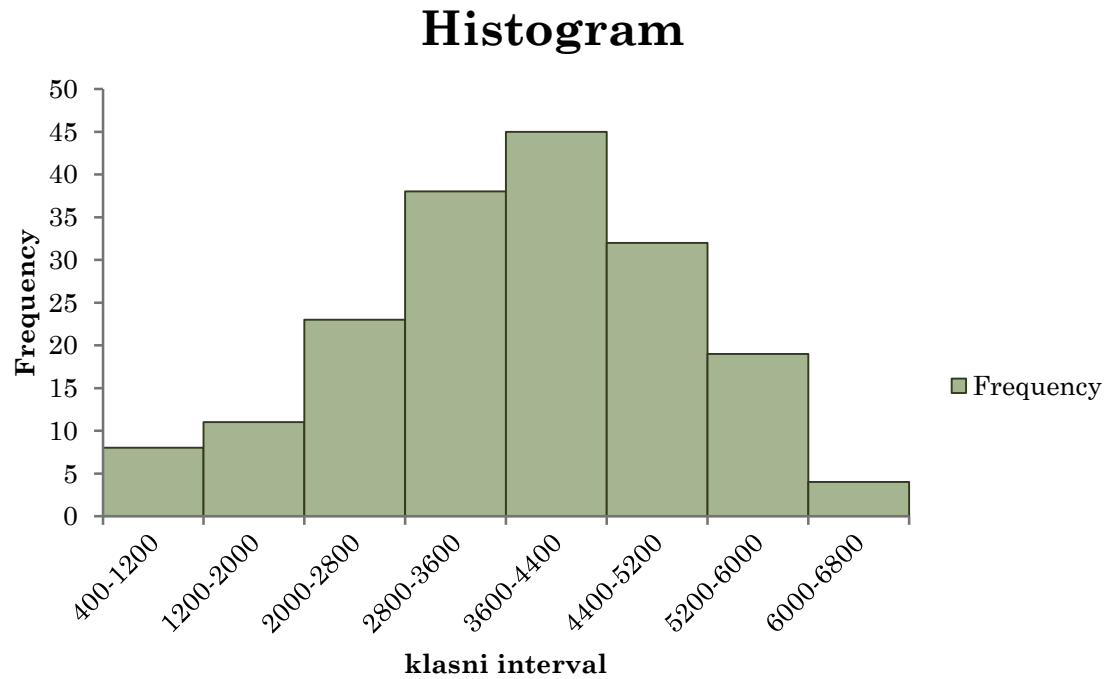
- **Grafik distribucije frekvencije** je način prikazivanja podataka u dvodimenzionalnoj ravni. Na x-osu se nanose obeležja, a na y-osi frekvencije.
- **Poligon** je grafički prikaz distribucije frekvencija pomoću tačaka koje se spajaju u izlomljenu liniju. U slučaju intervalnih klasa umesto intervala, kao predstavnik klase koristi se njena sredina
- **Histogram** je grafički prikaz distribucije frekvencije pomoću stubaca, kod kojih je širina proporcionalna širini klasnog intervala.
- Oba prikaza imaju svoje prednosti, histogram svojom visinom pokazuje frekvenciju klase, ali je poligon zgodan za poređenje više distribucija frekvencije.

POLIGON

klasni interval	klasna sredina	frekvencija
400 - 1200	800	8
1200 - 2000	1600	11
2000 - 2800	2400	23
2800 - 3600	3200	38
3600 - 4400	4000	45
4400 - 5200	4800	32
5200 - 6000	5600	19
6000 - 6800	6400	4



HISTOGRAM



KUMULATIVNE FREKVENCije

- **Kumulativne frekvencije** se izračunavaju sabiranjem (kumuliranjem) frekvencija od najmanje do najveće vrednosti obeležja.
- Kumulativna frekvencija „ispod“ izražava broj jedinica sa vrednošću obeležja $\leq x_i$ za klase sa diskretnim vrednostima, odnosno $\leq u_i$ u intervalnim klasama, gde je gornja granica klase u_i

$$K_i = \sum_{j=1}^i f_j = K_{i-1} + f_i$$

KUMULATIVNE FREKVENCije

- Kumulativna frekvencija „iznad“ izražava broj jedinica sa vrednošću obeležja $\geq x_i$ za klase sa diskretnim vrednostima, odnosno $\geq l_i$ u intervalnim klasama, gde je donja granica klase l_i

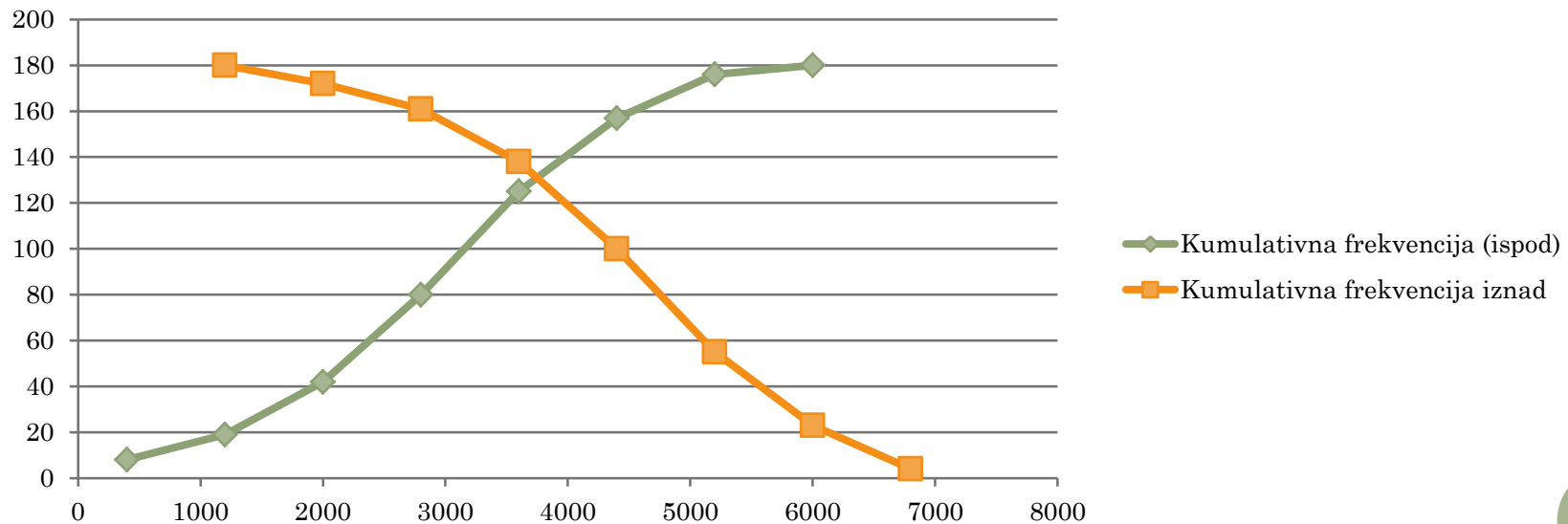
$$K'_i = \sum_{j=i}^n f_j = K'_{i-1} - f_{i-1}$$

PRIKAZ KUMULATIVNIH FREKVENCIJA

- Pri grafičkom prikazivanju kumulativnih frekvencija sa intervalnim klasama, kumulativna frekvencija „ispod“ dobija se unošenjem tačaka (u_i, K_i) , dok se kod kumulativna frekvencija „iznad“ unose tačke (l_i, K'_i) u koordinatni sistem. S obzirom da kod klasa sa diskretnim vrednostima $l_i = u_i = x_i$
- Presek kumulanti ima koordinate $\left(M_e, \frac{n}{2}\right)$, vrednost M_e se zove medijana i to je vrednost obeležja u odnosu na koju je statistički skup podeljen na dva jednaka dela.
- Kumulativne relativne frekvencije se dobijaju analogno (umesto f piše se p), u tom slučaju presek kumulanti ima koordinate $\left(M_e, \frac{1}{2}\right)$

PRIKAZ KUMULATIVNIH FREKVENCIJA

Klasni intervali	frekvencija	Kumulativna frekvencija (ispod)	Kumulativna frekvencija iznad
400 - 1200	8	8	180
1200 - 2000	11	19	172
2000 - 2800	23	42	161
2800 - 3600	38	80	138
3600 - 4400	45	125	100
4400 - 5200	32	157	55
5200 - 6000	19	176	23
6000 - 6800	4	180	4



ZADACI ZA SAMOSTALNI RAD

- 1. Mašina za pakovanje kafe puni kesice od 100 gr. Za kontrolu kvaliteta pakovanja posmatrano je 20 kesica i dobijeni su sledeći podaci
- 101, 102, 98, 98, 97, 99, 99, 100, 100, 102, 101, 10, 99, 98, 97, 100, 100, 100, 99, 97
- a) Srediti podatke u rastući niz
- b) Formirati distribucije apsolutnih frekvencija i predstaviti grafički preko histograma
- c) Formirati distribucije relativnih frekvencija i predstaviti grafički preko poligona

- 2. U toku 30 dana prati se prodaja ulja u nekoj prodavnici. Dobijeni su sledeći podaci:
- 37, 38, 37, 6, 20, 11, 22, 23, 3, 11, 4, 39, 17, 23, 37, 21, 21, 10, 12, 20, 15, 2, 20, 15, 22, 3, 18, 45, 36, 9
- Za podatke o prodatom ulju u prodavnici formirati distribuciju relativnih frekvencija za a) 5 klasa, b) 3 klase

- 3. Anketirani su kupci o broju članova u porodici i dobijeni su sledeći podaci
- 1 2 3 4 4 3 2 4 5 1 4 3 2 2 3 4 4 4 5 6 5 4 4 3 1 3 3 4 2 4
3 2 2 4 3 1
- Srediti podatke u rastući niz.
- Formirati distribuciju apsolutnih i relativnih frekvencija
- Prikazati grafički distribuciju relativnih frekvencija (poligonom)
- Formirati distribuciju kumulativnih frekvencija („ispod“ i „iznad“).
- Koliko kupaca ima manje od 4 člana?
- Koliko kupaca ima više od 3 člana?