

Statistika i vjerojatnost

1.

elementi



1.1. Prikaz podataka

Podatci su svuda oko nas. Tako se u e-imeniku nalaze ocjene i bilješke o napredovanju učenika, izostancima, izvannastavnim aktivnostima i sl. U zdravstvenom kartonu svake osobe nalaze se podatci o preboljenim bolestima, propisanim lijekovima, postojećim alergijama i dr. U poslovnim knjigama nalazimo podatke o ulaznim i izlaznim računima, poreznim i drugim davanjima, amortizaciji predmeta u vlasništvu tvrtke, isplaćenim plaćama, danima bolovanja, slobodnim danima, prekovremenim satima i sl. U različitim izvještajima mogu se naći podatci o gospodarstvu neke države, o različitim industrijskim granama i dr.

Već u Starom Egiptu, Kini, Rimu i drugdje skupljani su podatci vezani uz funkcioniranje države; radilo se o popisima stanovništva, popisima imovine, popisima poljoprivrednih prinosa i sl. Razvojem matematičkih i računalnih metoda, danas je statistika postala znanost kojom se koristimo u raznim područjima. Zanimljivo je da je na prostoru Hrvatske prvi poznati popis stanovništva proveden 1357. godine na području Dubrovačke Republike. Prvi popis stanovništva koji je obuhvatio područje današnje Hrvatske proveden je 1857. i podatci iz njega se smatraju prvim službenim podacima o broju stanovništva po naseljima u Hrvatskoj. Više zanimljivosti o popisima stanovništva u Hrvatskoj pronađite u prilogu “Kako smo se brojili posljednjih 150 godina” koji je objavljen na internetskim stranicama Državnog zavoda za statistiku (dzs.gov.hr).

Radi lakšeg razumijevanja, boljeg uočavanja pravilnosti i trendova, podatke razvrstavamo i prikazujemo na razne načine.

Primjer 1.

U jednom hotelu prikupljeni su dnevni podatci o tome koji napitak gost pije za doručak. Gosti su za doručak mogli piti mlijeko (M), čaj (Č), kavu (K), sok od naranče (N) ili sok od jabuke (J). Dežurni konobar bilježio je vrste napitaka i evo što je zapisao:

M	Č	Č	K	K	N	N	J	Č	K
K	M	N	N	J	Č	N	K	K	Č
K	Č	K	K	K	N	K	Č	N	Č

Ovako zapisani podatci su vrlo nepregledni. Nazivamo ih **primarni** podatci.

Bilo bi čitljivije da je vodio evidenciju koja bi na kraju izgledala ovako:

M					
Č					
K					
N					
J					

Zatim se podatci zapisuju u **tablicu frekvencija**.

vrsta napitka	broj čaša ili šalica napitka (frekvencija)
mlijeko	2
čaj	8
kava	11
sok od naranče	7
sok od jabuke	2

Svojstvo koje nas je zanimalo je vrsta napitka. To svojstvo još nazivamo **obilježje** ili **varijabla**. Napitak može biti: mlijeko, čaj, kava, sok od naranče, sok od jabuke. Znači, obilježje poprima jednu od tih pet **vrijednosti** ili **kategorija**.

Njih zapisujemo u prvom stupcu tablice. Broj koji kazuje koliko čaša ili šalica je popijeno uz taj doručak od svake vrste napitka naziva se **frekvencija** i te brojeve pišemo u drugi stupac. Dobivena tablica zove se **tablica frekvencija**.

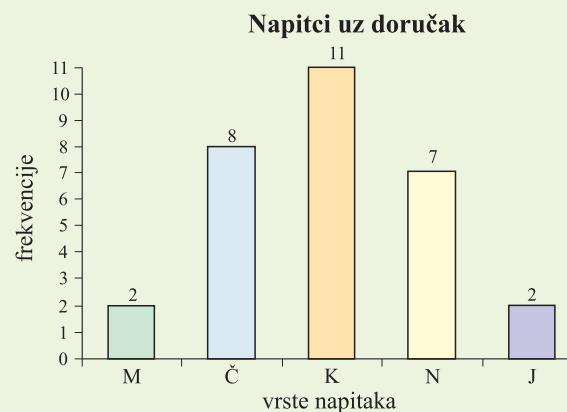
Katkad nas zanima koliki se dio svih podataka odnosi na pojedinu kategoriju. Drugim riječima, zanima nas broj koji pokazuje koliki je udio promatranog podatka u odnosu na cjelinu. Taj se broj naziva **relativna frekvencija**.

Relativnu frekvenciju računamo tako da frekvenciju podijelimo s ukupnim brojem podataka. Zbroj svih relativnih frekvencija mora biti jednak 1.

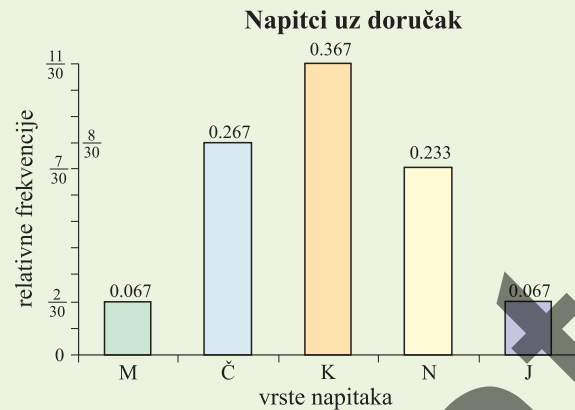
U ovom primjeru ukupan broj podataka je $n = 30$. To je zbroj svih frekvencija pa je primjerice relativna frekvencija mlijeka $\frac{2}{30}$. Često taj broj zapisujemo u obliku postotka: $\frac{2}{30} = 0.0666 \dots = 6.7 \%$.

vrsta napitka	broj čaša ili šalica napitka (frekvencija)	relativna frekvencija	relativna frekvencija u %
mlijeko	2	$\frac{2}{30}$	6.7 %
čaj	8	$\frac{8}{30}$	26.7 %
kava	11	$\frac{11}{30}$	36.7 %
sok od naranče	7	$\frac{7}{30}$	23.3 %
sok od jabuke	2	$\frac{2}{30}$	6.7 %

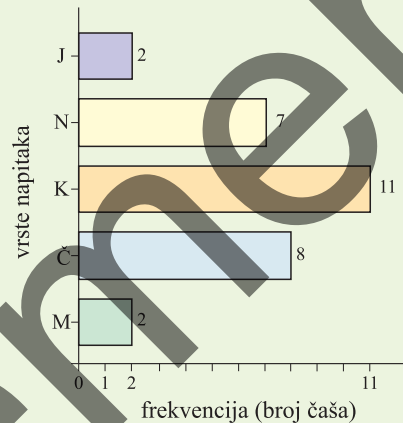
Prikazat ćemo podatke ovog primjera grafički, s pomoću **stupčastog dijagrama**. Prvo nacrtajmo stupčasti dijagram frekvencija. Svaku vrijednost obilježja i njezinu frekvenciju prikazujemo u obliku stupca. Širine svih stupaca međusobno su jednake, a visine su jednake frekvencijama.



Slično izgleda stupčasti dijagram relativnih frekvencija. Kod njega visine su stupaca jednake relativnim frekvencijama vrijednosti obilježja.



Ponekad se stupci crtaju položeno, pa takav dijagram frekvencija izgleda kao na donjoj slici.



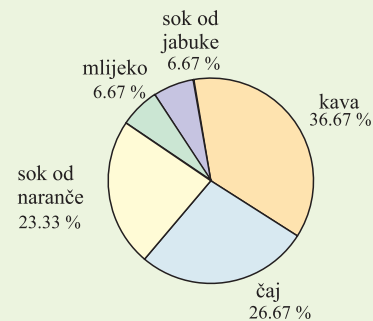
Osim stupčastim dijagramom, podatke možemo prikazati i **kružnim** ili **strukturnim dijagramom**. Svakoj kategoriji odgovara jedan kružni isječak čiji je središnji kut proporcionalan frekvenciji. Pokažimo to na primjeru napitka *kava*. Frekvencija kave je 11, a ukupni broj podataka je 30. S druge strane, središnji kut α je nepoznat, a ukupan broj stupnjeva cijelog kruga je 360° .

Postavljamo razmjer

$$11 : 30 = \alpha : 360, \quad \alpha = \frac{11 \cdot 360}{30}, \quad \alpha = 132^\circ.$$

Ostali kutovi su 24° , 96° , 84° .

Koristeći se šestarom i kutomjerom, crtamo krug i kutove u krugu.



Zamijetimo da smo pri rješavanju ovog primjera tablice i crteže izrađivali ručno, bez upotrebe posebnih informatičkih programa. Pri radu s velikim brojem podataka primjereno je koristiti se nekim od specijaliziranih digitalnih alata za proračunske tablice (kao što su Excel, Google sheets, Geogebra, LibreOffice i sl.) koji omogućavaju brzu izradu tablica i dijagrama. Ovdje ćemo opisati izradu tablice i dijagrama iz primjera 1 s pomoću Excela.

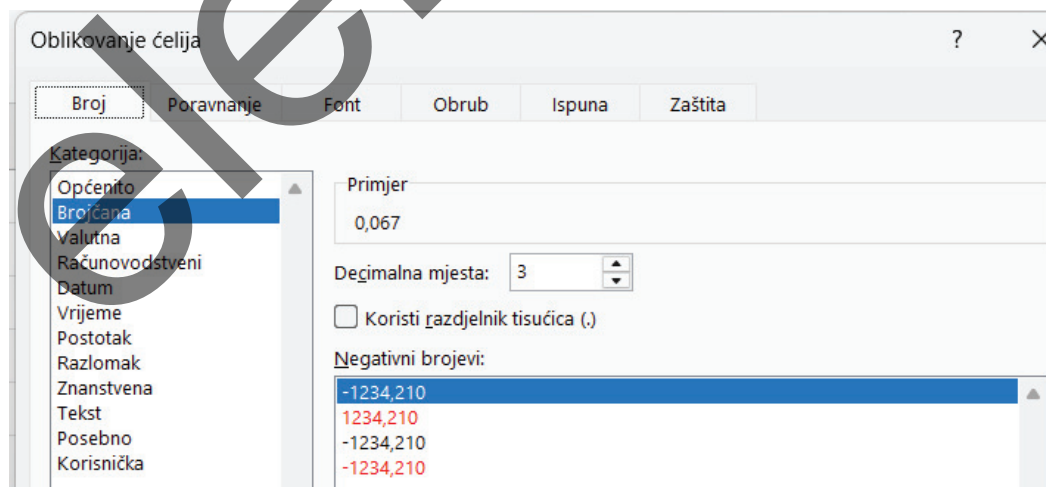
Otvorimo novu datoteku te u ćelije stupca A upišimo nazive kategorija (vrste napitaka), u ćelije stupca B frekvencije, tj. broj čaša napitaka. Budući da su vrijednosti u ćelijama B2, B3, . . . , B6 cijeli brojevi, napravimo i formatiranje ćelija, tj. oblikovanje podataka. Naredbe za oblikovanje podataka smještene su na kartici “Polazno” na alatnoj traci. Na toj kartici, unutar grupe naredbi “Broj” označene ćelije oblikujemo tako da sadrže prirodne brojeve, tj. decimalne brojeve s 0 decimalnih mjesta.

Istovremeno, u ćelije C1 i D1 upišimo naslove za ta dva stupca: “relativna frekvencija”, “relativna frekvencija u %”.

Iako u ovom primjeru imamo poznat ukupan broj podataka, napraviti ćemo ćeliju u kojoj piše koliko ukupno imamo podataka, tj. u ćeliju F2 upišemo formulu $= \text{SUM}(B2 : B6)$. Vrijednost ćelije F2 je 30. Sad na ekranu tablica ima ovaj oblik:

	A	B	C	D	E	F
1	vrsta napitka	broj čaša (frekvencija)	relativna frekvencija	relativna frekvencija u %		
2	mlijeko	2			ukupno	30
3	čaj	8				
4	kava	11				
5	sok od naranče	7				
6	sok od jabuke	2				

U ćelije stupca C upisat ćemo relativne frekvencije. Prvo u ćeliju C2 upišemo naredbu $= B2/F\$2$. Formatirajmo ćeliju C2 tako da se u njoj nalazi broj s 3 decimalna mjesta.



1.

Statistika i vjerojatnost

Zatim tu formulu prenesimo na ostale ćelije u stupcu C tako da mišem stanemo na donji rub ćelije C2 te povučemo do ćelije C6. Na taj se način u ćeliji C3 pojavljuje formula $B3/F\$2$, u ćeliji C4 formula $B4/F\$2$ itd.

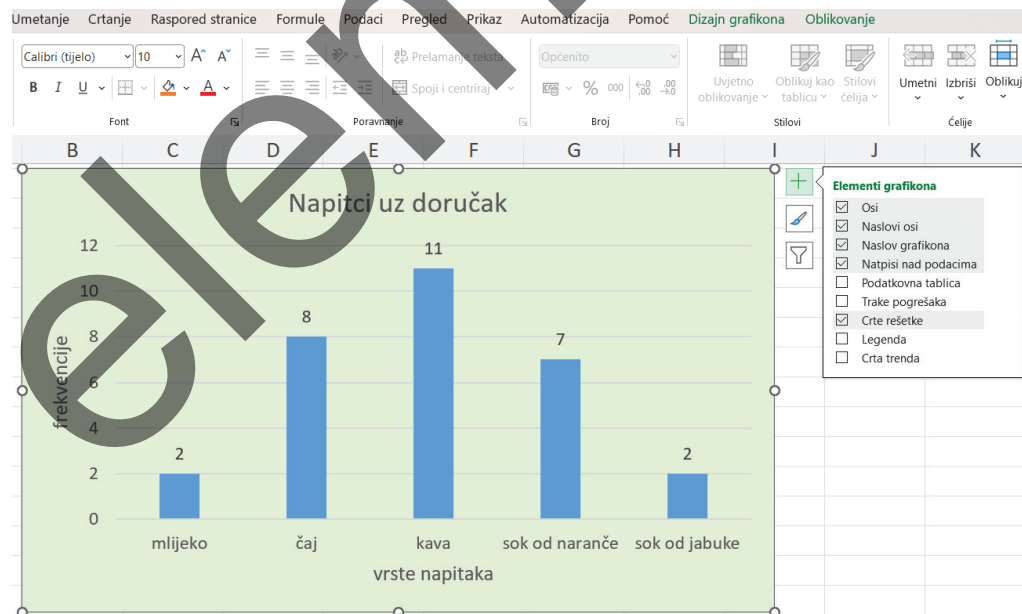
Znak \$ u imenu ćelije F2 osigurava da pri prenošenju formule iz jedne ćelije u drugu ne dolazi do promjene broja ćelije F2.

U stupac D upisat ćemo relativne frekvencije u obliku postotka ovako: u ćeliju D2 upišemo formulu $=PERCENTAGE(C2)$ i tu formulu prenesemo u sve ćelije od D3 do D6.

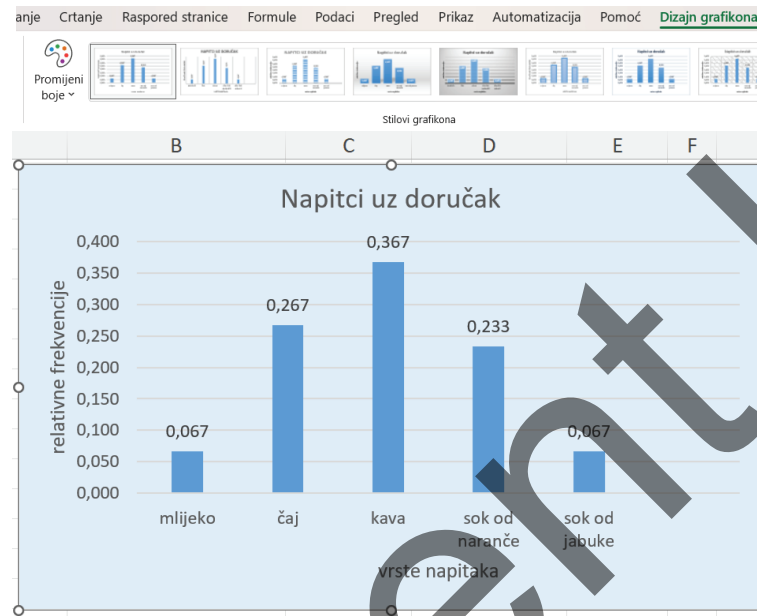
Dobili smo tablicu:

	A	B	C	D	E	F
1	vrsta napitka	broj čaša (frekvencija)	relativna frekvencija	relativna frekvencija u %		
2	mlijeko	2	0,067	6,7%	ukupno	30
3	čaj	8	0,267	26,7%		
4	kava	11	0,367	36,7%		
5	sok od naranče	7	0,233	23,3%		
6	sok od jabuke	2	0,067	6,7%		

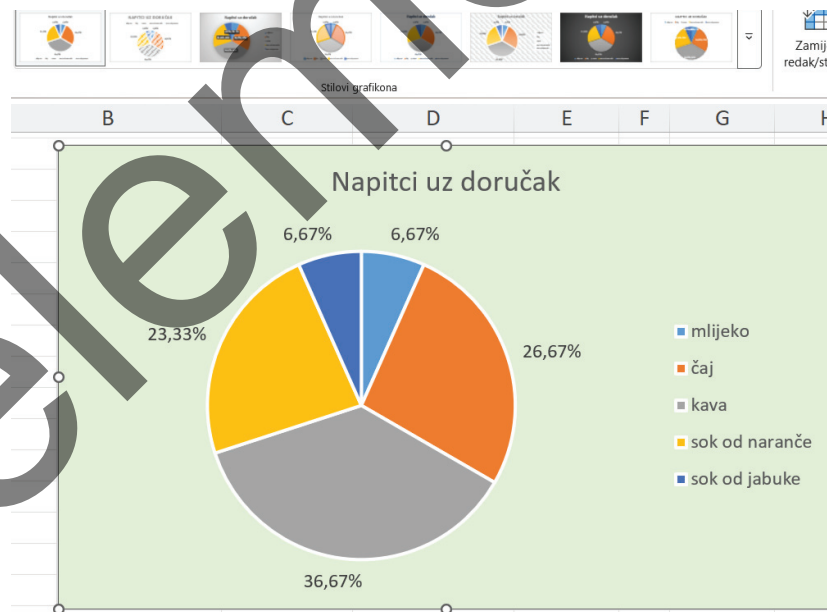
Sad možemo pristupiti crtanju stupčastog dijagrama frekvencija. Označimo prva dva stupca A i B te u alatnoj traci u kartici “Umetanje” odaberimo stupčasti dijagram. Pojavljuje se grafikon koji još treba urediti tako da mu se upiše naslov, naslovi na horizontalnoj i vertikalnoj osi. Ovdje smo dodatno oblikovali oznake stupaca tako da frekvencije pišu iznad pojedinog stupca.



Za crtanje stupčastog dijagrama relativnih frekvencija označavamo stupce A i C te biramo izradu stupčastog dijagrama.



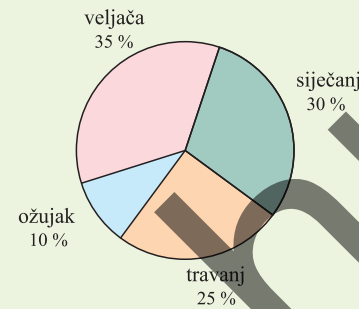
Za crtanje kružnog dijagrama označavamo stupce A i D te biramo izradu kružnog dijagrama.



Svaki dijagram može se još dodatno urediti. Samostalno otkrij koje mogućnosti uređivanja dijagrama nudi program.

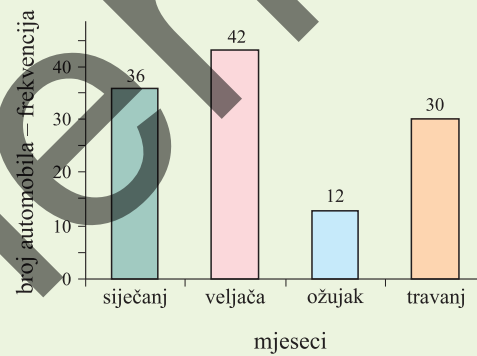
Primjer 2.

Kružnim su dijagramom prikazani podatci o broju iznajmljenih automobila u jednoj *rent-a-car* tvrtki tijekom četiri mjeseca. Ukupno je iznajmljeno 120 automobila. Prikaži podatke s pomoću tablice frekvencija, stupčastim i linijskim dijagramom.

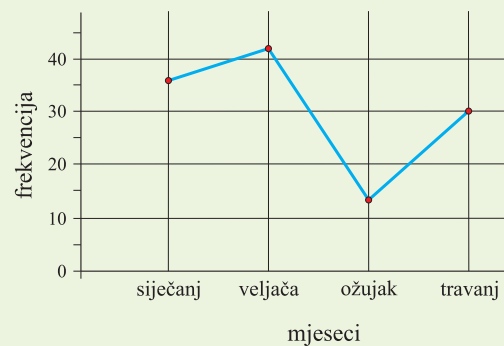


30 % od 120 je $0.3 \cdot 120 = 36$. U siječnju je iznajmljeno 36 automobila: to je frekvencija koju upisujemo u drugi stupac tablice. Ostale frekvencije su 35 % od 120 = $0.35 \cdot 120 = 42$, 10 % od 120 = $0.1 \cdot 120 = 12$, 25 % od 120 = $0.25 \cdot 120 = 30$.

mjesec	frekvencija (broj iznajmljenih automobila)
siječanj	36
veljača	42
ožujak	12
travanj	30



Kad podatke prikazujemo linijskim dijagramom, tada se na vodoravnoj osi nalaze vrijednosti obilježja, a na okomitoj osi frekvencije. Od tih se podataka stvaraju točke koje spajamo dužinama zbog čega se ovakav prikaz i zove **linijski dijagram**.



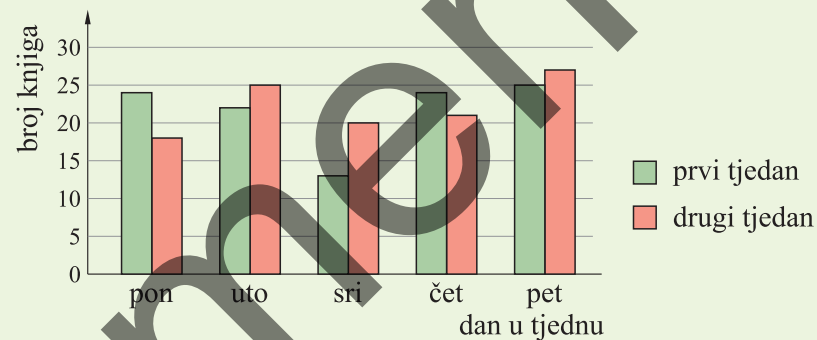
Primjer 3.

U tablici se nalazi broj posuđenih knjiga u školskoj knjižnici tijekom protekla dva tjedna. Prikažimo podatke stupčastim i linijskim dijagramom.

	pon	uto	sri	čet	pet
prvi tjedan	24	22	13	24	25
drugi tjedan	18	25	20	21	27

Budući da su za svaku vrijednost obilježja dana dva podatka u tablici, crtat ćemo **višestruki stupčasti dijagram** frekvencija.

Za svaki od dana u tjednu nacrtat ćemo dva pripadajuća stupca, a među njima nećemo stavljati razmak. Stupce koji se odnose na prvi tjedan crtat ćemo s lijeve, a one koji se odnose na drugi tjedan s desne strane. Obojit ćemo ih različitim bojama te sa strane predočiti legendu kako bi se znalo na koji se tjedan koji stupci odnose.



Nacrtajmo i linijski dijagram.

Nacrtat ćemo prvo linijski dijagram za broj posuđenih knjiga u prvom tjednu i označit ćemo ga jednom bojom. Potom ćemo na istom dijagramu nacrtati drugom bojom linijski dijagram za broj posuđenih knjiga u drugom tjednu.

