

Uvodne informacije

□ **Statistika**

- Nastavnik: doc. dr Julija Cerović Smolović
- Saradnik: dr Nemanja Batrićević
- Udžbenik: Prem Mann, **Uvod u statistiku**, šesto izdanje, 2009. John Wiley & Sons.
- Mail: julija@ucg.ac.me
- Ispit: kolokvijum + završni ispit

POGLAVLJE 1



UVOD

1.1 ŠTA JE STATISTIKA? OBLASTI STATISTIKE

Definicija

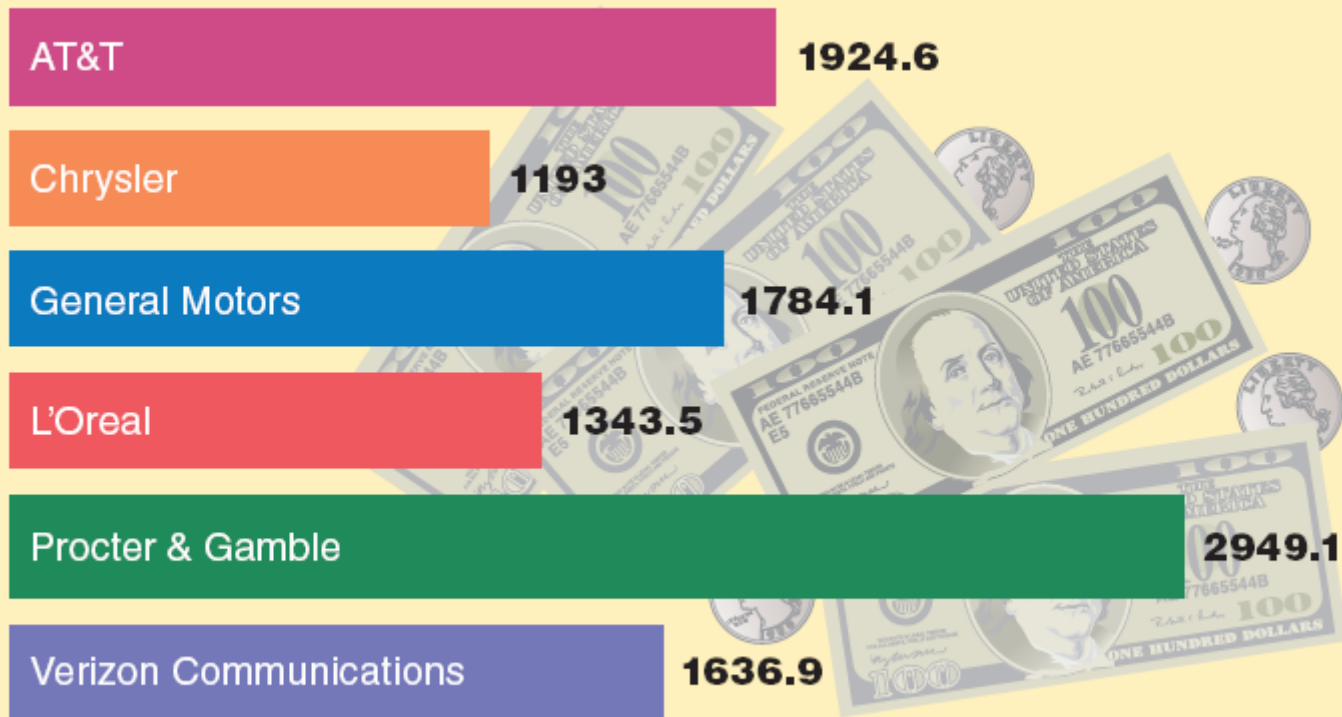
Statistika je naučni metod koji se koristi za prikupljanje, prikazivanje, analizu i interpretaciju podataka, kao i za donošenje statističkih zaključaka na osnovu tih analiza.

Deskriptivna statistika se sastoji od metoda prikupljanja, sređivanja, prikazivanja i opisivanja podataka pomoću tabela, grafikona i sumarnih pokazatelja.

Inferencijalna statistika obuhvata statističke metode koje primjenjujemo da bismo na osnovu rezultata iz dijela osnovnog skupa (uzorka) došli do zaključaka o karakteristikama osnovnog skupa.

Studija slučaja 1-1 Koliko su kompanije potrošile na oglase u 2011?

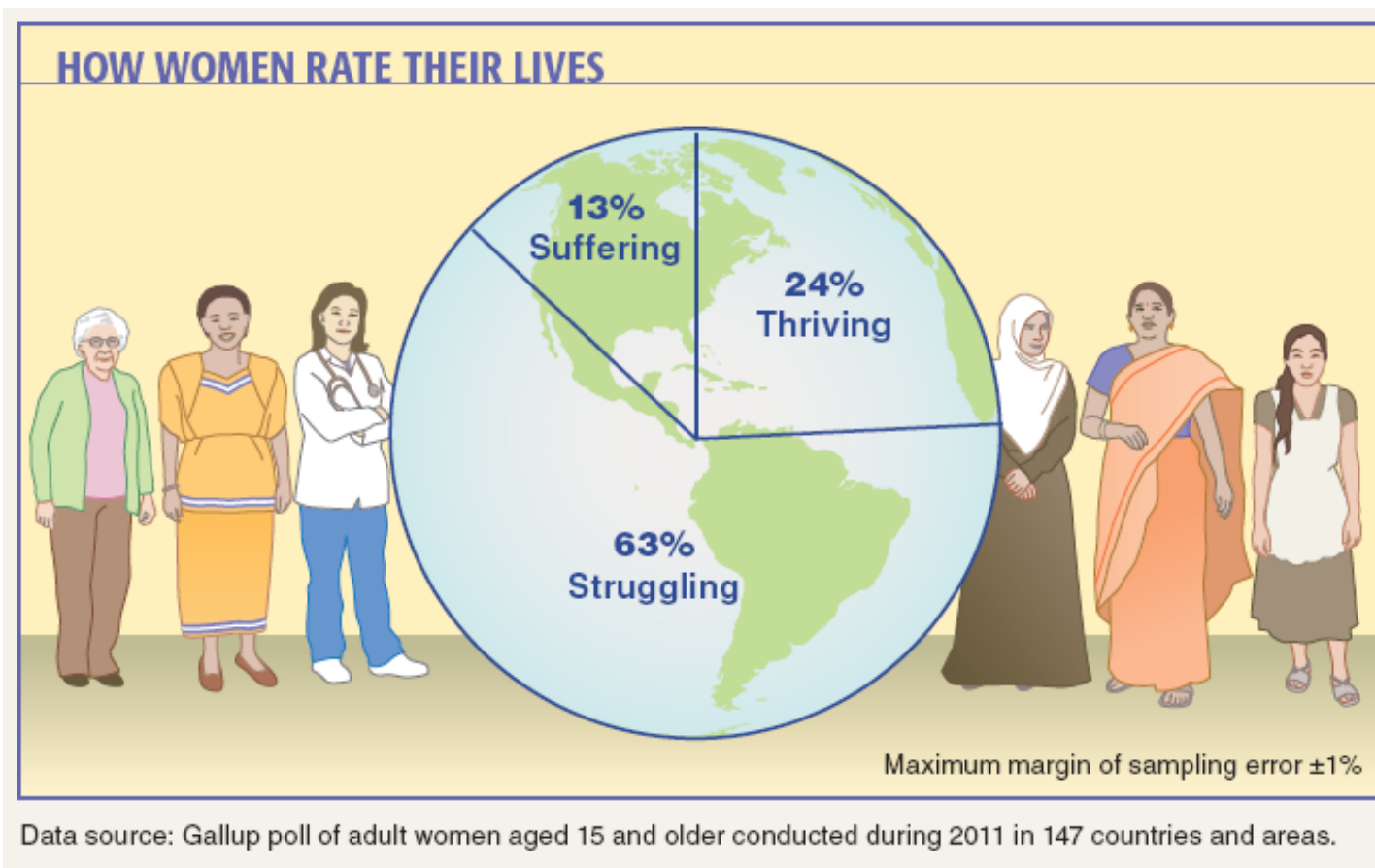
HOW MUCH DID COMPANIES SPEND ON ADS IN 2011?



Note: Amounts are in millions of dollars.

Data source: WPP Kantar Media.

Studija slučaja 1-2 Kako žene ocjenjuju svoj život



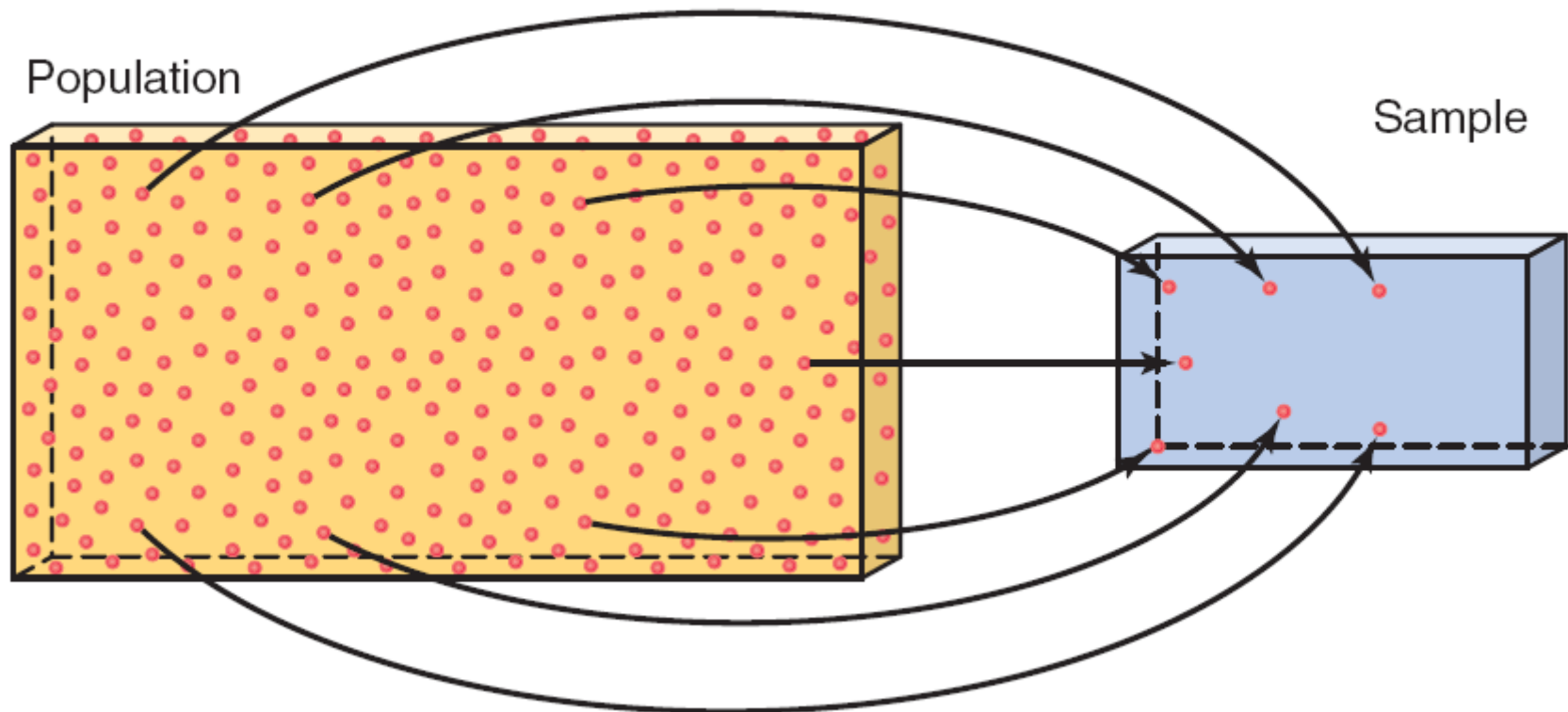
1.2 OSNOVNI SKUP I UZORAK

Definicija

Osnovni skup (populacija) se sastoji od svih elemenata ili jedinica posmatranja – pojedinaca (bića), stvari ili predmeta – čije karakteristike ispituјemo. Osnovni skup koji se proučava naziva se još i **ciljnom populacijom**.

Dio osnovnog skupa koji je izabran u svrhe statističke analize naziva se **uzorkom**.

Slika 1.1 Populacija i uzorak



POPULACIJA NASPRAM UZORKA

Definicija

Anketa kojom prikupljamo podatke o svim jedinicama osnovnog skupa nazivamo **popisom**. Tehnika prikupljanja podataka iz dijela osnovnog skupa naziva se **uzoračkom anketom**.

Uzorak koji u najvećoj mjeri odražava karakteristike osnovnog skupa jeste **reprezentativni uzorak**.

Uzorak dobijen tako da svaki element osnovnog skupa ima unaprijed poznatu vjerovatnoću da bude izabran naziva se **slučajnim uzorkom**. Ako svi uzorci iste veličine koje biramo iz osnovnog skupa imaju istu vjerovatnoću da budu izabrani, takav postupak se naziva **izborom prostog slučajnog uzorka**. Takav uzorak predstavlja **prost slučajan uzorak**.

POPULACIJA NASPRAM UZORKA

Uzorak može biti izabran sa ili bez ponavljanja.

Prilikom izbora uzorka (uzorkovanja) **sa ponavljanjem**, element koji smo iz osnovnog skupa izabrali u uzorak vraćamo u osnovni skup prije sledećeg izvlačenja.

Prilikom izbora uzorka **bez ponavljanja** jednom izabrani element u uzorak se ne vraća u osnovni skup.

1.3 OSNOVNI POJMOVI

Definicija

Jedinica posmatranja ili **element** uzorka ili osnovnog skupa jeste određeni subjekat ili objekat (na primjer, osoba, firma, predmet, država, ili zemlja) o kojem se prikupljaju podaci, odnosno, na kojem se određena pojava statistički posmatra.

Promjenljiva (obilježje ili varijabla) je osobina (karakteristika) koja se proučava ili istražuje, koja podrazumijeva različite vrijednosti po jedinicama posmatranja. Za razliku od promjenljive, vrijednost **konstante** je uvijek fiksna.

OSNOVNI POJMOVI

Definicija

Vrijednost promjenljive koja se odnosi na jednu jedinicu posmatranja naziva se opservacijom ili podatkom.

Seriya podataka je skup podataka koji se odnosi na jednu ili više promjenljivih.

Tabela 1.1 Ukupni prihodi šest američkih kompanija, 2010

Kompanija	Ukupni prihodi 2010 (milioni \$)	← promjenljiva
Wal-Mart Stores	421,849	
Royal Dutch Shell	378,152	
Exxon Mobil	354,674	← opservacija ili podatak
BP	308,928	
Sinopec Group	273,422	
China National Petroleum	240,192	

Izvor: *Fortune Magazine*, 25. jul, 2011.

Jedinica posmatranja ili element

1.4 VRSTE PROMJENLJIVIH

- Kvantitativne promjenljive
 - Prekidne promjenljive
 - Neprekidne promjenljive

- Kvalitativne ili kategorijske promjenljive

Kvantitativne promjenljive: prekidne i neprekidne

Definicija

Promjenljiva koja se može brojčano izraziti naziva se **kvantitativnom (numeričkom) promjenljivom**. Podaci prikupljeni o kvantitativnoj promjenljivoj nazivaju se **kvantitativnim podacima**.

Promjenljiva čije vrijednosti možemo da brojimo naziva se **prekidnom (diskretnom) promjenljivom**. Drugim riječima, prekidna promjenljiva može da ima samo izolovane vrijednosti, najčešće cijele brojeve, a ne međuvrijednosti.

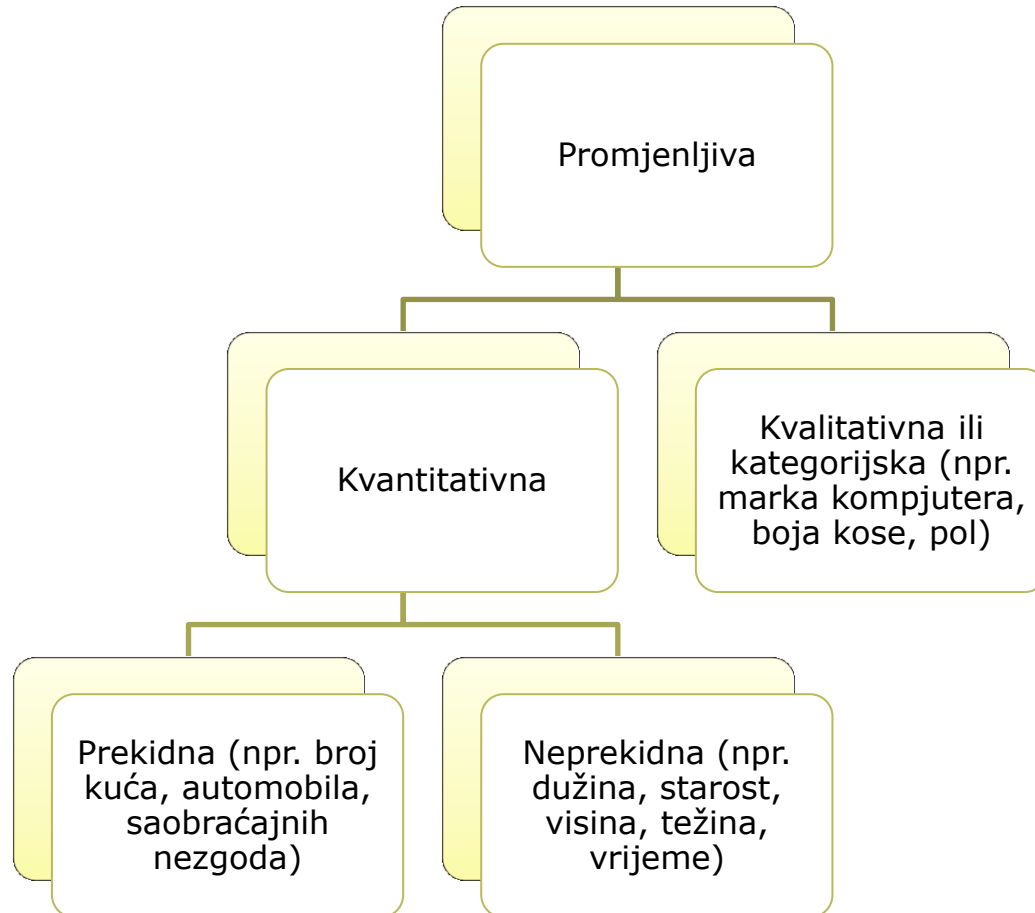
Promjenljiva koja može uzeti bilo koju numeričku vrijednost u određenom intervalu ili intervalima jeste **neprekidna (kontinuirana) promjenljiva**.

Kvalitativne ili kategorijske promjenljive

Definicija

Promjenljiva koja ne može uzeti numeričke vrijednosti, ali može da se klasifikuje u dvije ili više kategorija jeste **kvalitativna (atributivna)** ili **kategorijska promjenljiva**. Podaci prikupljeni o ovakvoj promjenljivoj nazivaju se **kvalitativnim podacima**.

Slika 1.2 Vrste promjenljivih



1.5 STRUKTURNE SERIJE I VREMENSKE SERIJE

- Strukturne serije
- Vremenske serije

Strukturne serije

Definicija

Podaci prikupljeni o različitim elementima osnovnog skupa ili uzorka u istom vremenskom trenutku, ili u istom vremenskom periodu nazivaju se **strukturnim serijama podataka (podacima presjeka ili uporednim podacima)**.

Tabela 1.2 Ukupni prihodi šest američkih kompanija, 2010

Kompanija	Ukupni prihodi 2010 (milioni \$)
Wal-Mart Stores	421,849
Royal Dutch Shell	378,152
Exxon Mobil	354,674
BP	308,928
Sinopec Group	273,422
China National Petroleum	240,192

Izvor: *Fortune Magazine*, 25. jul, 2011.

Vremenske serije

Definicija

Seriya podataka prikupljenih za istu jedinicu posmatranja o jednoj istoj promjenljivoj u različitim vremenskim trenucima ili vremenskim periodima, naziva se **vremenskom serijom**.

Tabela 1.3 Novac vraćen od presuda o prevarama u zdravstvu

Godina	Vraćen novac (milijarde \$)
2006	2.2
2007	1.8
2008	1.0
2009	1.6
2010	2.5

1.6 IZVORI PODATAKA

- Podaci se mogu dobiti iz
 - Internih izvora
 - Eksternih izvora
 - Anketa i eksperimenata

POGLAVLJE 2

SREĐIVANJE I GRAFIČKO PRIKAZIVANJE PODATAKA

2.1 Negrupisani podaci

Definicija

Podaci zapisani redosljedom kojim se prikupljaju, prije nego što se urede po veličini ili grupišu nazivaju se **negrupisani podaci**.

Tabela 2.1 Starost 50 studenata

21	19	24	25	29	34	26	27	37	33
18	20	19	22	19	19	25	22	25	23
25	19	31	19	23	18	23	19	23	26
22	28	21	20	22	22	21	20	19	21
25	23	18	37	27	23	21	25	21	24

SREĐIVANJE I GRAFIČKO PRIKAZIVANJE PODATAKA

- ❑ Raspodjele frekvencija
- ❑ Raspodjele relativnih frekvencija i procentualne raspodjele
- ❑ Grafičko prikazivanje kvalitativnih podataka

Tabela 2.2 Vrsta zaposlenja za koju će se studenti opredijeliti

Promjenljiva

Vrsta zaposlenja	Broj studenata
Privatne kompanije/firme	44
Federalni organi vlasti	16
Državni/lokalni organi vlasti	23
Sopstvena firma	17
	Ukupno = 100

Kolona frekvencija

Frekvencija

Kategorija

Raspodjele frekvencija

Definicija

Raspodjela frekvencija kvalitativnih podataka sadrži dva niza informacija: kategorije (modalitete) i odgovarajući broj jedinica posmatranja koji pripada svakoj kategoriji.

Primjer 2-1

Izabran je uzorak od 30 osoba koje često konzumiraju krofne i postavljeno im je pitanje koja vrsta krofne im je omiljena. Odgovori ovih 30 osoba su sledeći:

Primjer 2-1

glazirana	punjena	drugo	obična	glazirana	drugo
posuta	punjena	punjena	glazirana	drugo	posuta
glazirana	obična	drugo	glazirana	glazirana	punjena
posuta	obična	drugo	drugo	posuta	punjena
punjena	drugo	posuta	glazirana	glazirana	punjena

Formirati tabelu raspodjele frekvencija na osnovu ovih podataka.

Primjer 2-1: Rješenje

Tabela 2.3 Raspodjela frekvencija omiljene vrste krofne

Vrsta krofne	Oznaka	Frekvencija (f)
Glazirana		8
Punjena		7
Posuta		5
Obična		3
Drugo		7
		Ukupno = 30

Raspodjele relativnih frekvencija i procentualne raspodjele

Izračunavanje relativne frekvencije jedne kategorije

$$\textit{Relativna frekvencija jedne kategorije} = \frac{\textit{Frekvencija kategorije}}{\textit{Zbir svih frekvencija}}$$

Raspodjele relativnih frekvencija i procentualne raspodjele

Izračunavanje procenta (učešća)

$$\text{Učešće} = (\text{relativna frekvencija}) \cdot 100\%$$

Primjer 2-2

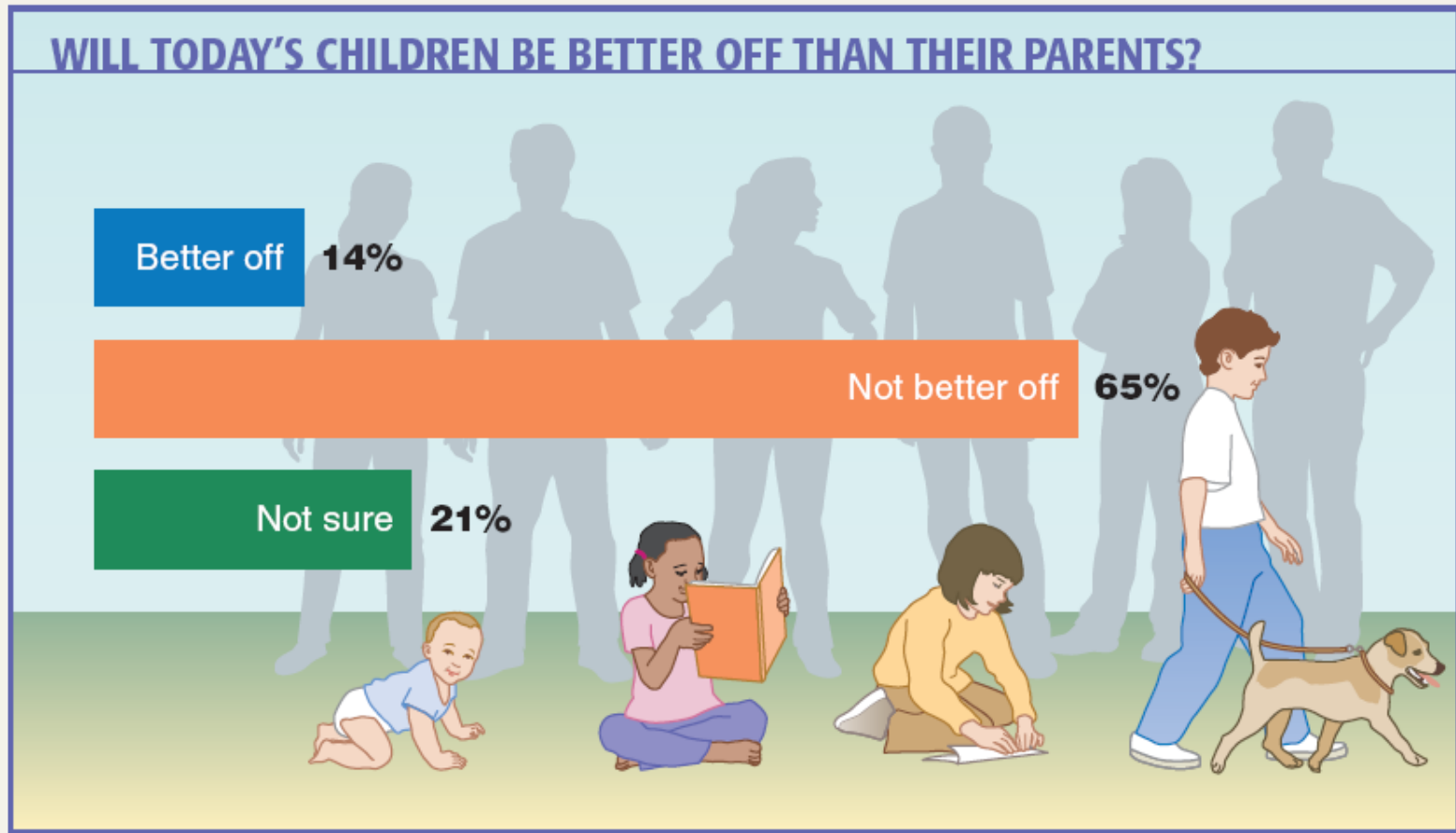
Odrediti relativne frekvencije i procenat za podatke u Tabeli 2.3.

Primjer 2-2: Rješenje

Tabela 2.4 Raspodjela relativne frekvencije i procentualna raspodjela omiljene vrste krofne

Vrsta krofne	Relativna Frekvencija	Procenat
Glazirana	$8/30= 0.267$	$0.267(100)=26.7$
Punjena	$7/30= 0.233$	$0.233(100)=23.3$
Posuta	$5/30= 0.167$	$0.167(100)=16.7$
Obična	$3/30= 0.100$	$0.100(100)=10.0$
Drugo	$7/30= 0.233$	$0.233(100)=23.3$
	Ukupno=1.000	Ukupno=100%

Studija slučaja 2-1 Da li će današnja djeca biti bolja od svojih roditelja?



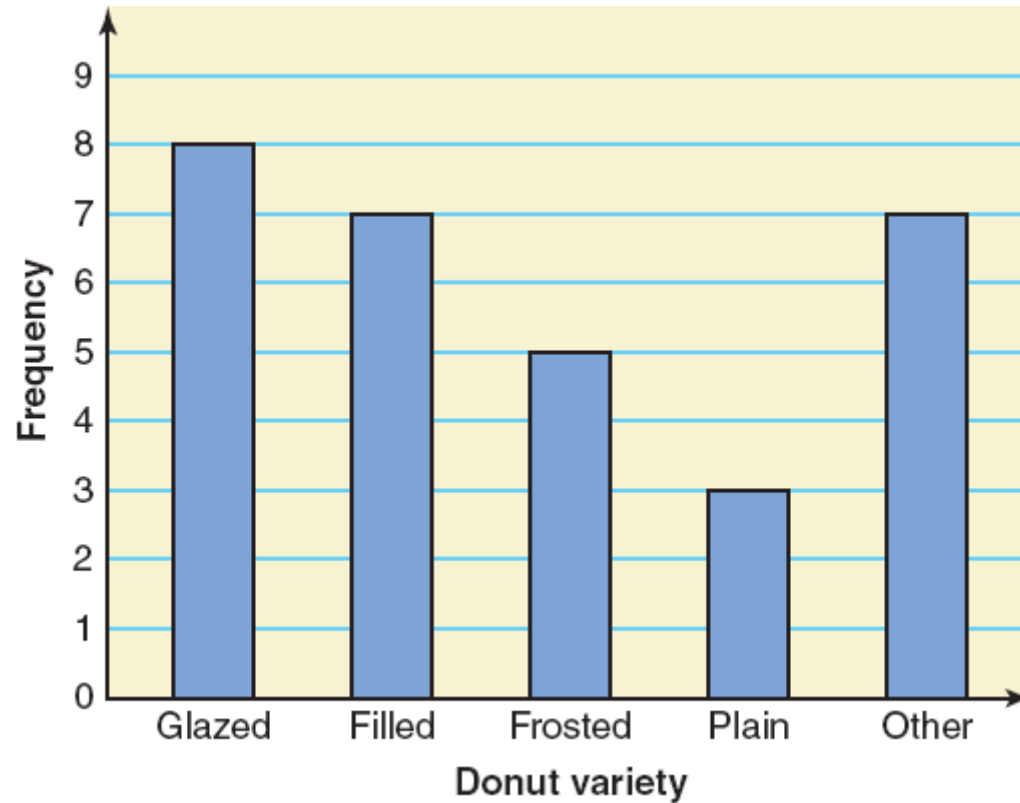
Data source: Rasmussen Reports national telephone survey of American adults.

Grafičko prikazivanje kvalitativnih podataka

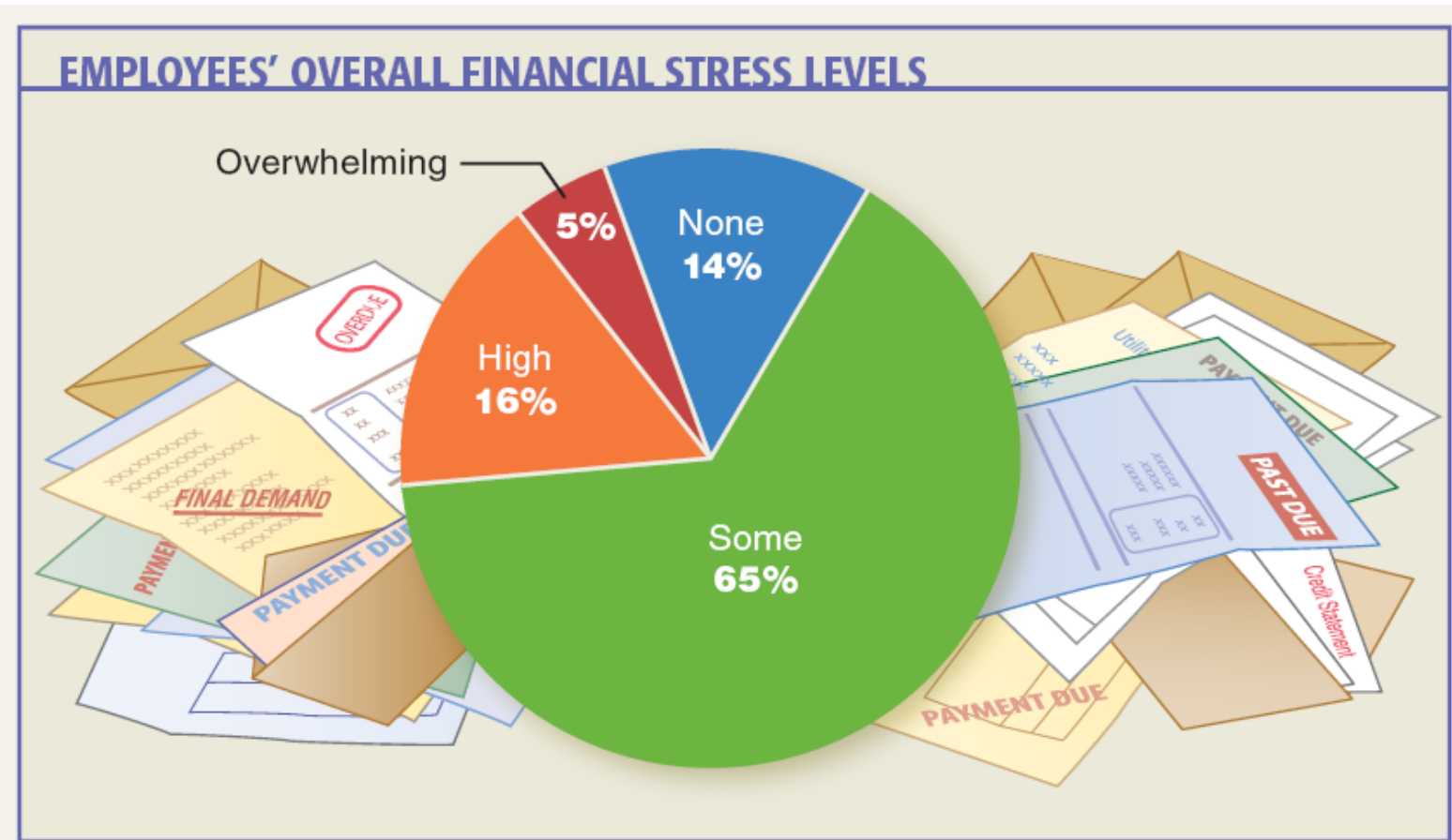
Definicija

Grafikon koji se sastoji od stubića čija visina predstavlja frekvenciju različitih kategorija naziva se **štapićastim dijagramom**.

Slika 2.1 Štapićasti dijagram za raspodjelu frekvencija podataka iz Tabele 2.3



Studija slučaja 2-2 Ukupni nivoi finansijskog stresa zaposlenih



Data source: Financial Finesse, Inc.

Grafičko prikazivanje kvalitativnih podataka

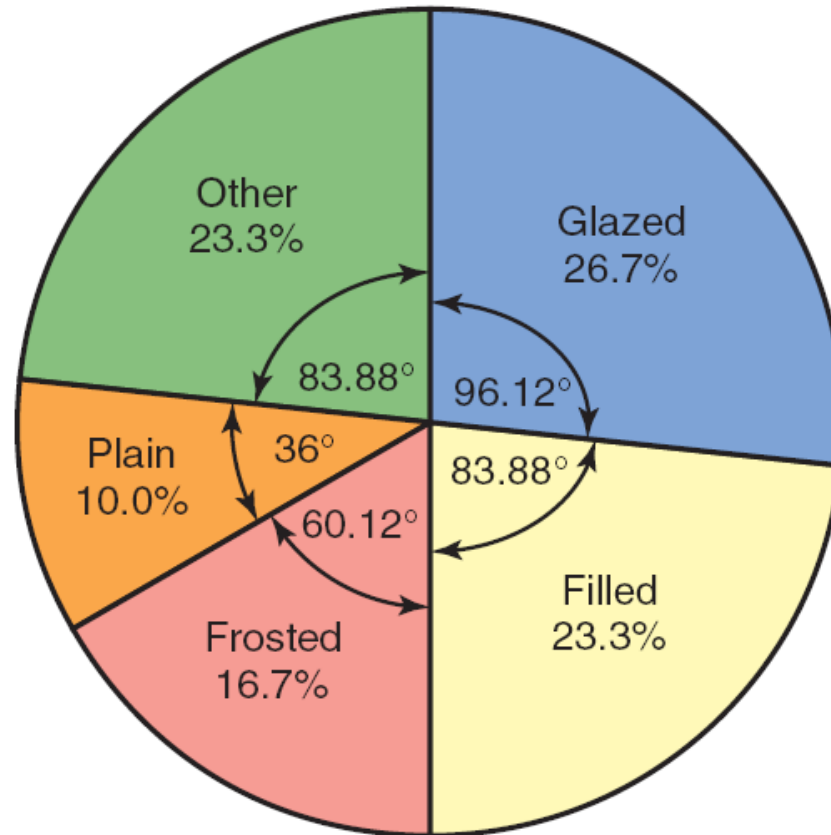
Definicija

Krug podijeljen na djelove, od kojih svaki dio predstavlja relativnu frekvenciju ili učešće odgovarajuće kategorije u osnovnom skupu ili uzorku, naziva se **pita dijagramom (strukturnim krugom)**.

Tabela 2.5 Izračunavanje veličine uglova za strukturni krug

Vrsta krofne	Relativna Frekvencija	Veličina ugla (stepeni)
Glazirana	0.267	$360(0.267)=96.12$
Punjena	0.233	$360(0.233)=83.88$
Posuta	0.167	$360(0.167)= 60.12$
Obična	0.100	$360(0.100)=36.00$
Drugo	0.233	$360(0.233)= 83.88$
	Ukupno=1.000	Ukupno=360

Slika 2.2 Strukturni krug za procentualnu raspodjelu na osnovu Tabele 2.5.



2.2 SREĐIVANJE I GRAFIČKO PRIKAZIVANJE NUMERIČKIH (KVANTITATIVNIH) PODATAKA

- ❑ Raspodjele frekvencija
- ❑ Tabelarno prikazivanje raspodjele frekvencija
- ❑ Raspodjela relativnih frekvencija i procentualna raspodjela
- ❑ Grafičko prikazivanje grupisanih podataka

Tabela 2.6 Nedjeljne zarade 100 zaposlenih u jednoj kompaniji

Treći grupni interval Promjenljiva

Nedjeljna zarada (u dolarima)	Broj zaposlenih f	Frekvencije
401 - 600	9	
601 - 800	22	
801 - 1000	39	Frekvencija trećeg grupnog intervala
1001 - 1200	15	
1201 - 1400	9	
1401 - 1600	6	

Donja granica šestog grupnog intervala

Gornja granica šestog grupnog intervala

Raspodjele frekvencija

Definicija

Raspodjela frekvencija numeričkih podataka sadrži dva niza informacija: vrijednosti promjenljive prikazane grupnim intervalima i njima odgovarajuće brojeve jedinica posmatranja. Podaci prikazani raspodjelom frekvencija nazivaju se **grupisanim podacima**.

Raspodjele frekvencija

Definicija

Granična vrijednost grupnog intervala predstavlja sredinu između gornje granice jednog grupnog intervala i donje granice narednog grupnog intervala.

Raspodjele frekvencija

Određivanje širine grupnog intervala

Širina grupnog intervala = Gornja granična vrijednost –
Donja granična vrijednost

Raspodjele frekvencija

Izračunavanje sredine grupnog intervala

$$\textit{Sredina grupnog intervala} = \frac{\textit{donja granica} + \textit{gornja granica}}{2}$$

Raspodjele frekvencija

Izračunavanje širine grupnog intervala

$$\begin{aligned} & \textit{Aproksimativna širina grupnih intervala} \\ & = \frac{\textit{Najveća vrijednost} - \textit{najmanja vrijednost}}{\textit{Broj grupnih intervala}} \end{aligned}$$

Tabela 2.7 Granice, granične vrijednosti, širine i sredine grupnih intervala za podatke iz Tabele 2.6

Granica	Granična vrijednost intervala	Širina	Sredina grupnog intervala
801 do 1000	800.5 do manje od 1000.5	200	900.5
1001 do 1200	1000.5 do manje od 1200.5	200	1100.5
1201 do 1400	1200.5 do manje od 1400.5	200	1300.5
1401 do 1600	1400.5 do manje od 1600.5	200	1500.5
1601 do 1800	1600.5 do manje od 1800.5	200	1700.5
1801 do 2000	1800.5 do manje od 2000.5	200	1900.5

Primjer 2-3

Sledeći podaci odnose se na broj prodatih iPods® od strane jedne kompanije u 30 dana. Prikazati tabelarno raspodjelu frekvencije.

8	25	11	15	29	22	10	5	17	21
22	13	26	16	18	12	9	26	20	16
23	14	19	23	20	16	27	16	21	14

Primjer 2-3: Rješenje

Minimalna vrijednost je 5, a maksimalna je 29. Pretpostavimo da smo odlučili da te podatke grupišemo u pet grupnih intervala jednake širine. Tada,

$$\text{Aproksimativna širina grupnih intervala} = \frac{29 - 5}{5} = 4.8$$

Dobijenu aproksimativnu širinu ćemo zaokružiti na neki podesan broj, recimo na 5. Kao donju granicu prvog grupnog intervala možemo uzeti 5 ili bilo koji drugi manji broj od 5. Pretpostavimo da smo uzeli 5 kao donju granicu prvog grupnog intervala. Grupni intervali će tada biti

$$5 - 9, \quad 10 - 14, \quad 15 - 19, \quad 20 - 24, \quad \text{i} \quad 25 - 29$$

Tabela 2.8 Raspodjela frekvencija za podatke o iPod prodajama

Prodati Ipods	Oznaka	f
5-9		3
10-14		6
15-19		8
20-24		8
25-29		5
		$\Sigma f = 30$

Raspodjela relativnih frekvencija i procentualna raspodjela

Izračunavanje relativne frekvencije i učešća

$$\text{Relativna frekvencija grupnog intervala} = \frac{\text{Frekvencija grupnog intervala}}{\text{Zbir svih frekvencija}} = \frac{f}{\sum f}$$

$$\text{Učešće} = (\text{relativna frekvencija}) \times 100$$

Primjer 2-4

Izračunati relativne frekvencije i učešće za podatke iz Tabele 2.8.

Primjer 2-4: Rješenje

Tabela 2.9 Raspodjela relativnih frekvencija i procentualna raspodjela za podatke iz Tabele 2.8

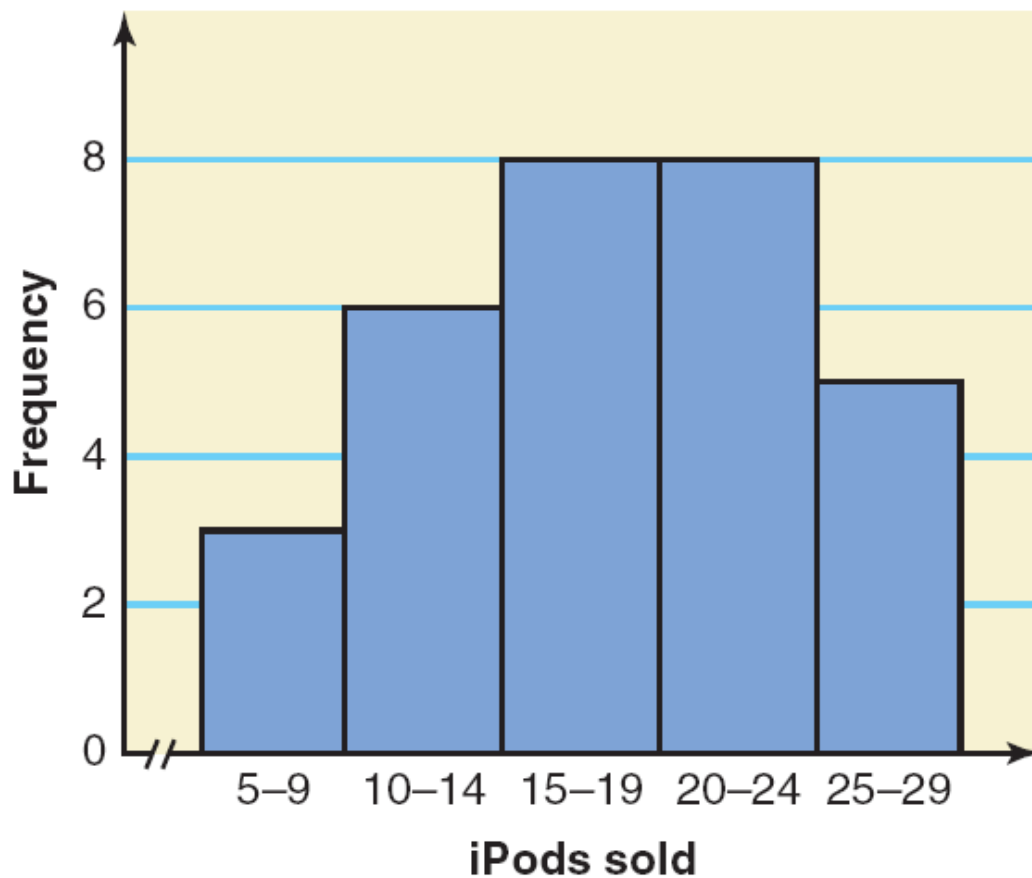
Prodati Ipods	Granična vrijednost intervala	Relativna frekvencija	Procenat
5-9	4.5 do manje od 9.5	$3/30=0.100$	10.0
10-14	9.5 do manje od 14.5	$6/30=0.200$	20.0
15-19	14.5 do manje od 19.5	$8/30=0.267$	26.7
20-24	19.5 do manje od 24.5	$8/30=0.267$	26.7
25-29	24.5 do manje od 29.5	$5/30=0.167$	16.7
		Zbir=1.001	Zbir=100.1

Grafičko prikazivanje grupisanih podataka

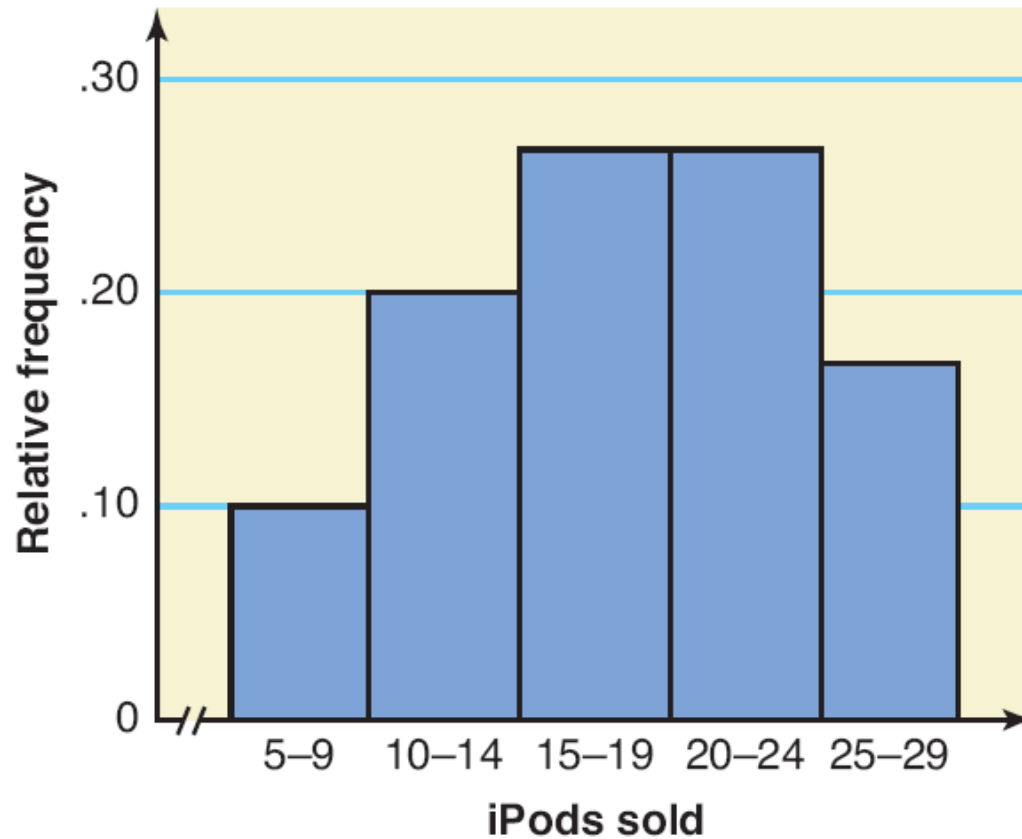
Definicija

Histogram je dijagram koji se sastoji od niza spojenih pravougaonika, čije su osnovice jednake grupnim intervalima prikazanim na x -osi, a visine odgovaraju frekvencijama, relativnim frekvencijama ili učešćima prikazanim na y -osi. Pravougaonici na histogramu crtaju se jedan do drugog, bez razmaka.

Slika 2.3 Histogram frekvencija za podatke iz Tabele 2.8.



Slika 2.4 Histogram relativnih frekvencija za podatke iz Tabele 2.9.

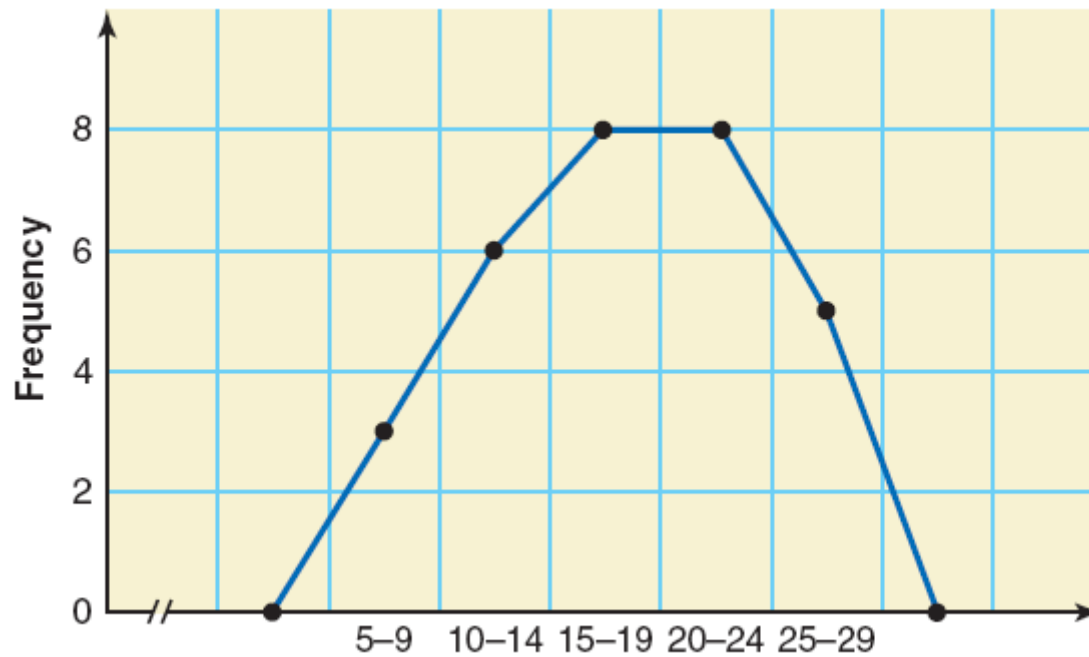


Grafičko prikazivanje grupisanih podataka

Definicija

Dijagram koji se dobija spajanjem tačaka čije su koordinate sredine grupnih intervala i frekvencije intervala (spajanjem sredina vrhova susjednih pravougaonika histograma frekvencija) zove se **poligon**.

Slika 2.5 Poligon frekvencija za podatke iz Tabele 2.8.



2.3 RASPODJELA KUMULATIVNIH FREKVENCIJA

Definicija

Raspodjela ***kumulativnih frekvencija*** pokazuje ukupan broj jedinica posmatranja koje imaju vrijednost ispod gornje granice svakog grupnog intervala.

Primjer 2-5

Koristeći raspodjelu frekvencija iz Tabele 2.9, koja je navedena dolje, formirati raspodjelu kumulativnih frekvencija za broj prodatih iPoda od strane date kompanije.

Prodati iPods	f
5-9	3
10-14	6
15-19	8
20-24	8
25-29	5

Primjer 2-5: Rješenje

Tabela 2.10 Raspodjela kumulativnih frekvencija prodaje iPoda

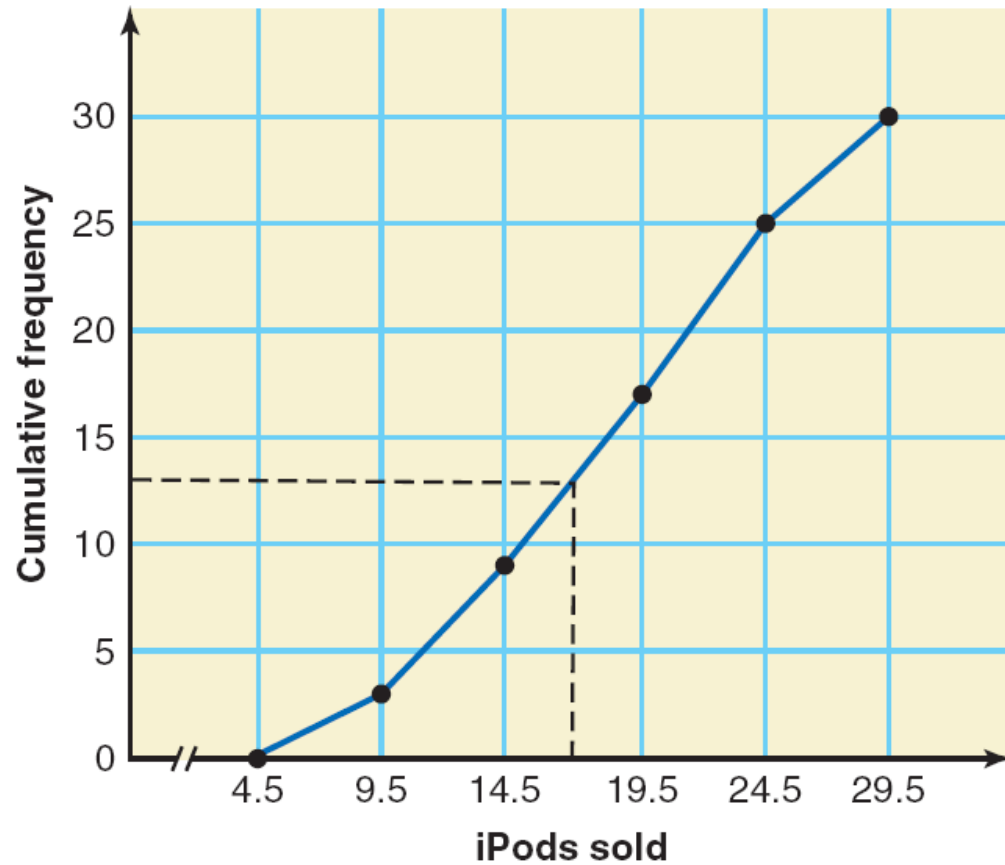
Granice grupnih intervala	Granične vrijednosti grupnih intervala	Kumulativna frekvencija
5-9	4.5 do manje od 9.5	3
10-14	9.5 do manje od 14.5	$3 + 6 = 9$
15-19	14.5 do manje od 19.5	$3 + 6 + 8 = 17$
20-24	19.5 do manje od 24.5	$3 + 6 + 8 + 8 = 25$
25-29	24.5 do manje od 29.5	$3 + 6 + 8 + 8 + 5 = 30$

RASPODJELE KUMULATIVNIH FREKVENCIJA

Definicija

Kumulanta je kriva koja pokazuje raspodjelu kumulativnih frekvencija i dobijena je kada se pravim linijama spoje tačke koje odgovaraju gornjim granicama grupnih intervala i kumulativnim frekvencijama odgovarajućih grupnih intervala.

Slika 2.11 Kumulanta za raspodjelu kumulativnih frekvencija za podatke iz Tabele 2.11.



2.4 PRIKAZ U OBLIKU STABLJIKE I LISTA

Definicija

Prikaz u obliku stabljike i lista koristi se za kvantitativne podatke, tako što se svaka vrijednost dijeli na dva dijela – stabljiku i list. Listovi svakog stabla se prikazuju posebno.

Slika 2.14 Rangirani stablo-list dijagram za broj poena na ispitu.

5		0	2	7						
6		1	4	5	8	9				
7		1	1	2	2	5	6	7	9	9
8		0	1	3	4	6	7	7		
9		2	2	3	5	6	8			

Jedna od prednosti stablo-list dijagrama jeste to što ne gubimo informacije o pojedinačnim opservacijama.

2.5 TAČKASTI DIJAGRAMI

Definicija

Vrijednosti koje su u odnosu na većinu vrijednosti u skupu podataka izuzetno male ili izuzetno velike nazivaju se **ekstremnim vrijednostima**.

