

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

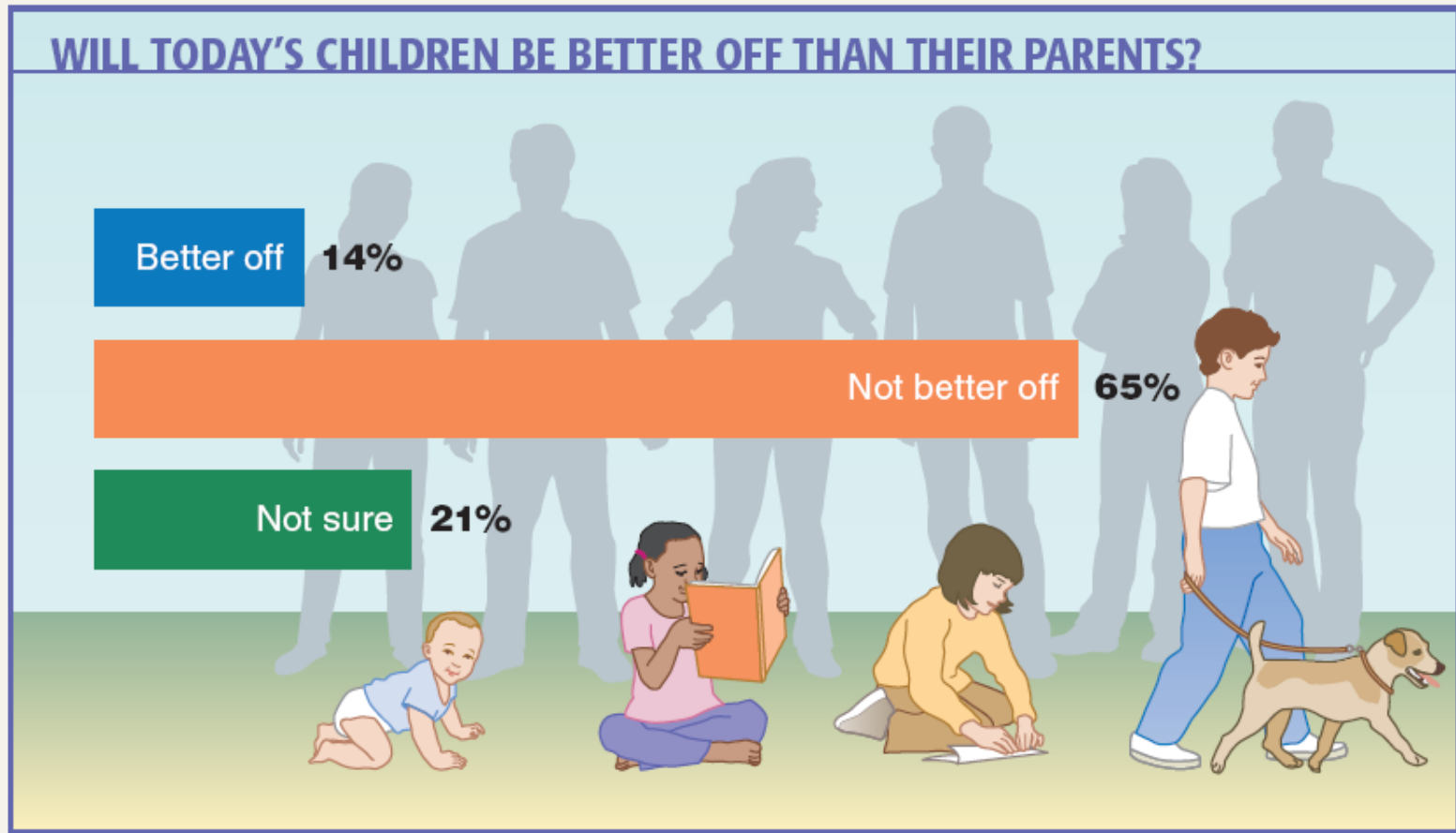
Odrediti relativne frekvencije i procenat za podatke u Tabeli 2.3.

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?



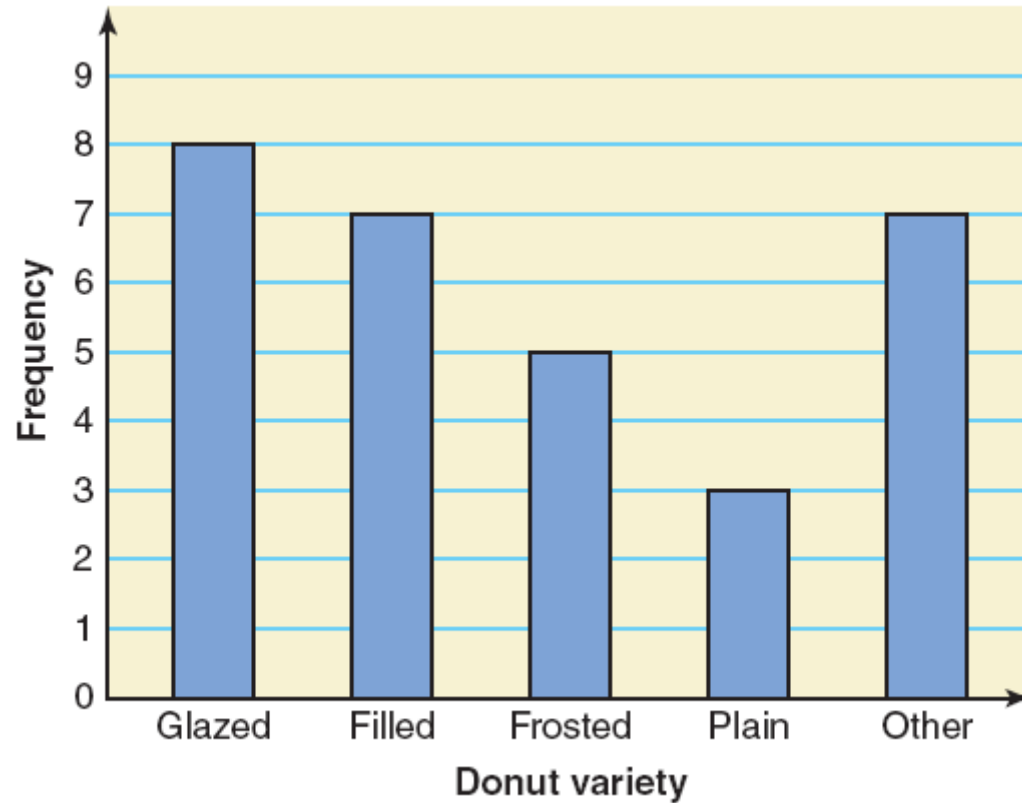
Data source: Rasmussen Reports national telephone survey of American adults.

Grafičko prikazivanje kvalitativnih podataka

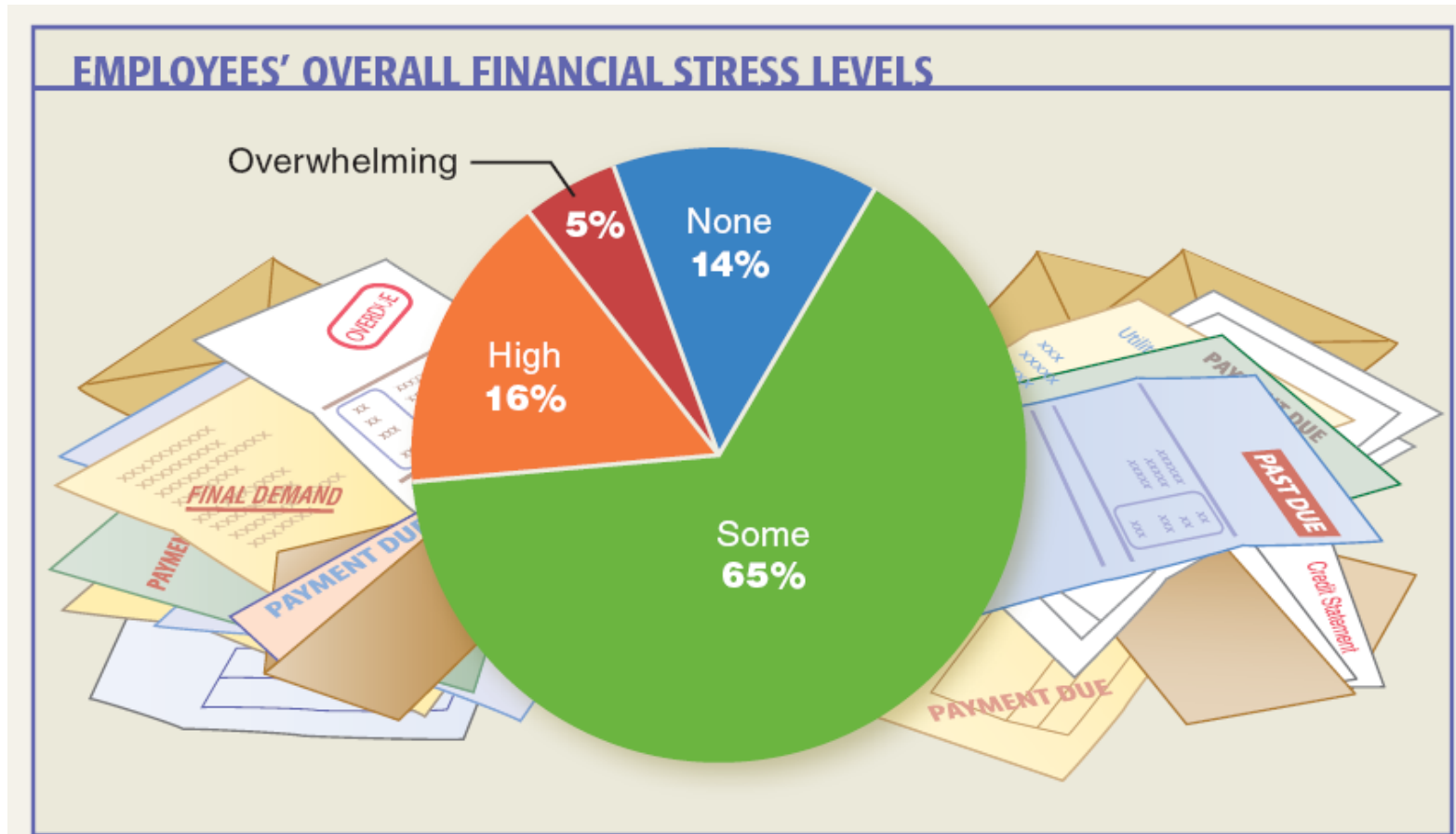
Definicija

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

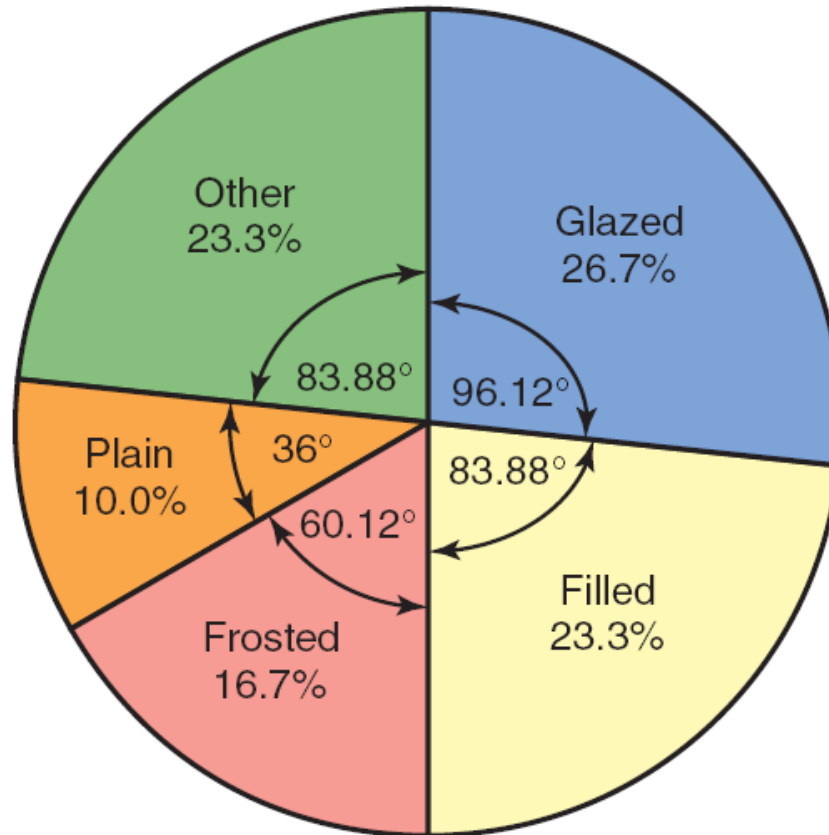
Definicija

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, 10 – 14, 15 – 19, 20 – 24, i 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

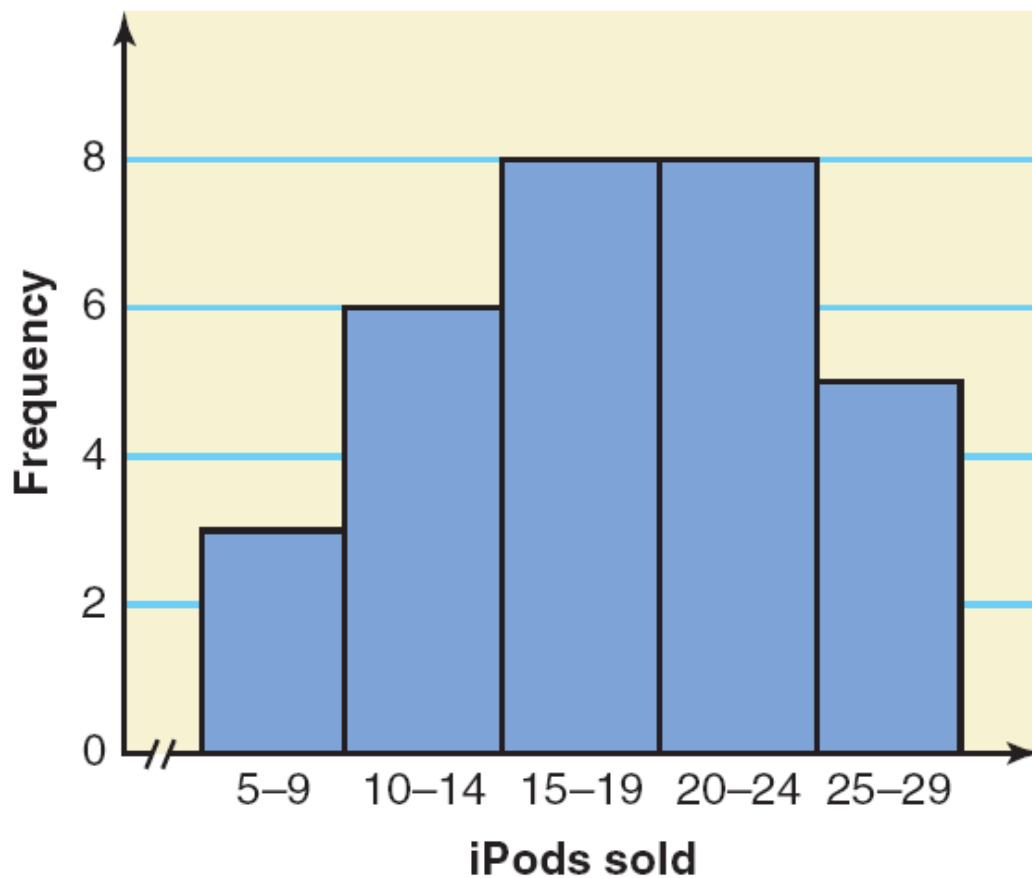
Prodati Ipods	Granična vrijednost intervala	Relativna frekvencija	Procenat
5-9	4.5 do manje od 9.5	$3/30=0.100$	10.0
10-14	9.5 do manje od 14.5	$6/30=0.200$	20.0
15-19	14.5 do manje od 19.5	$8/30=0.267$	26.7
20-24	19.5 do manje od 24.5	$8/30=0.267$	26.7
25-29	24.5 do manje od 29.5	$5/30=0.167$	16.7
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

Grafičko prikazivanje grupisanih podataka

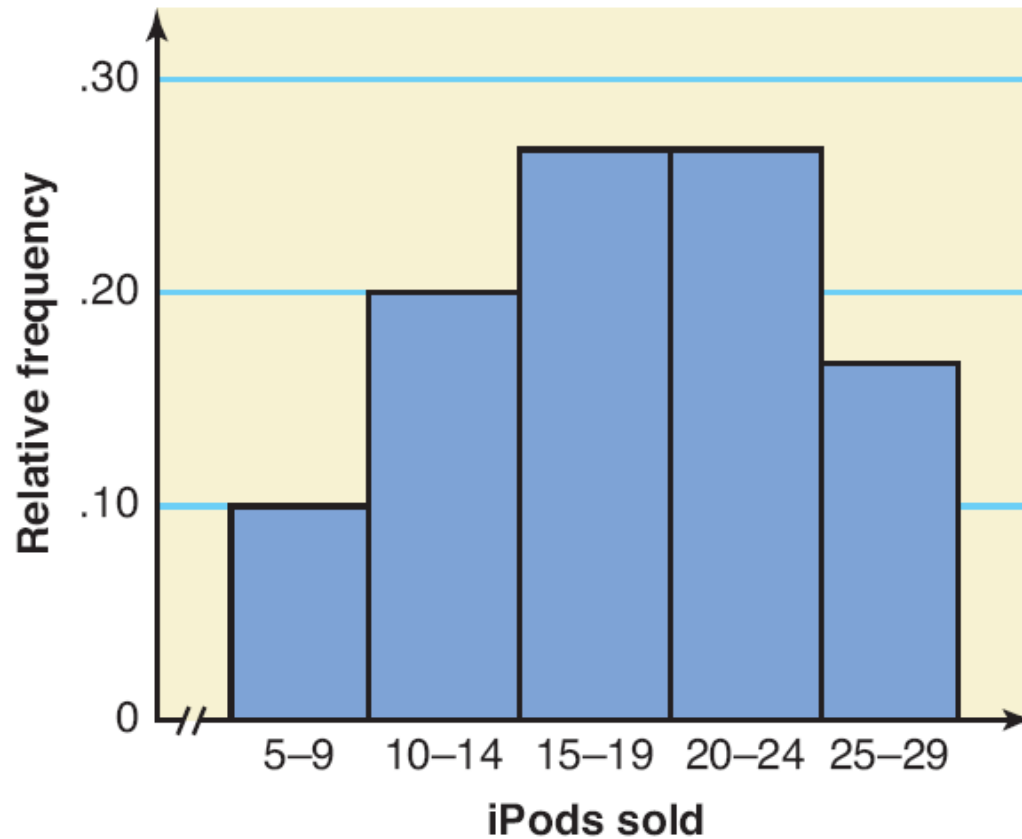
Definicija

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.

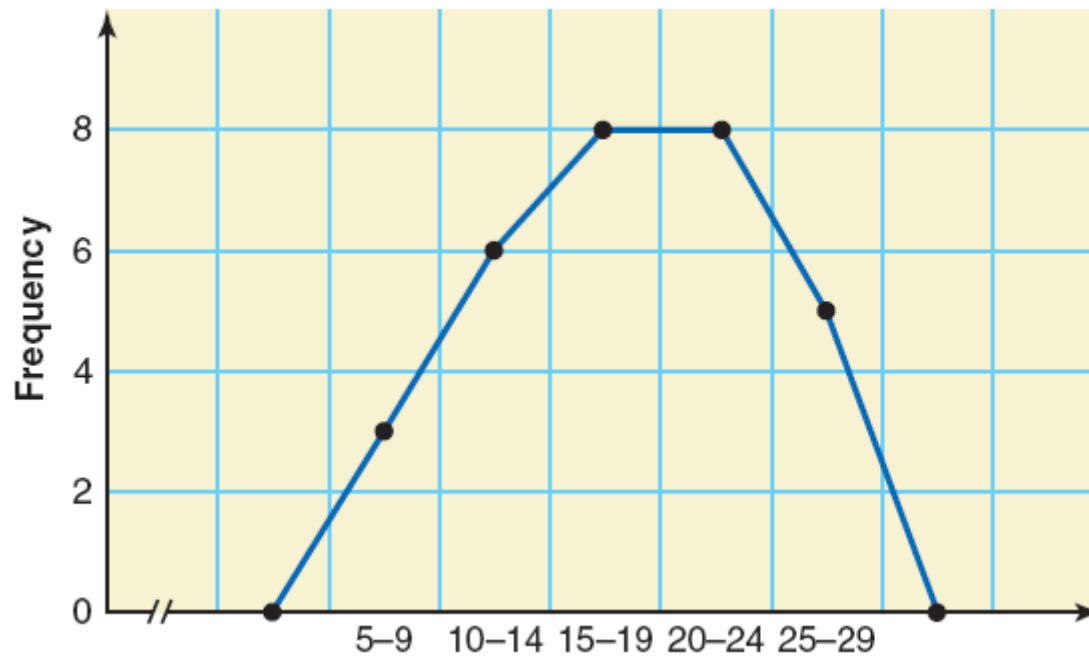


Grafičko prikazivanje grupisanih podataka

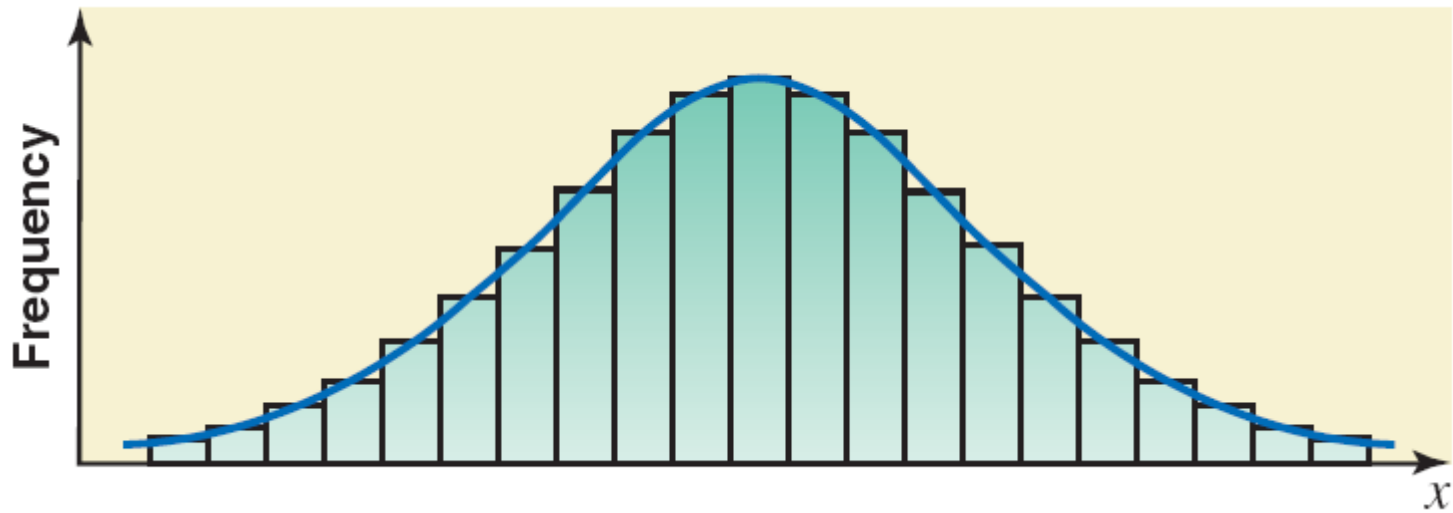
Definicija

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.



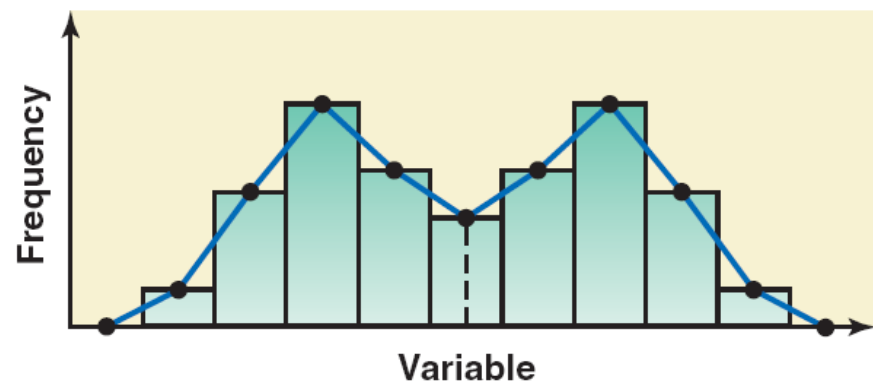
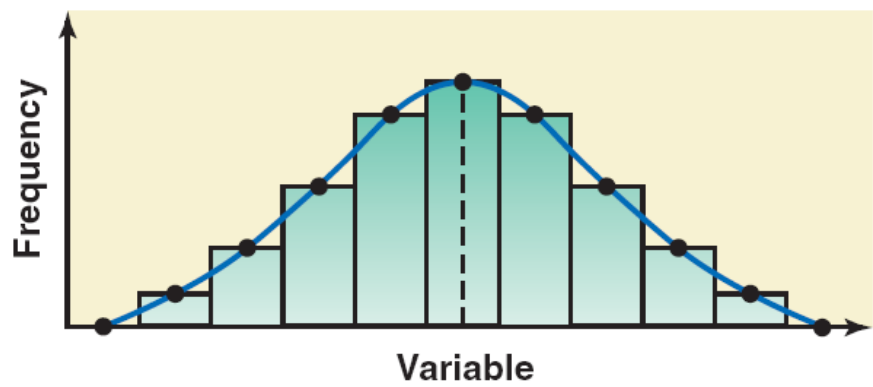
Slika 2.6 Kriva raspodjele frekvencija.



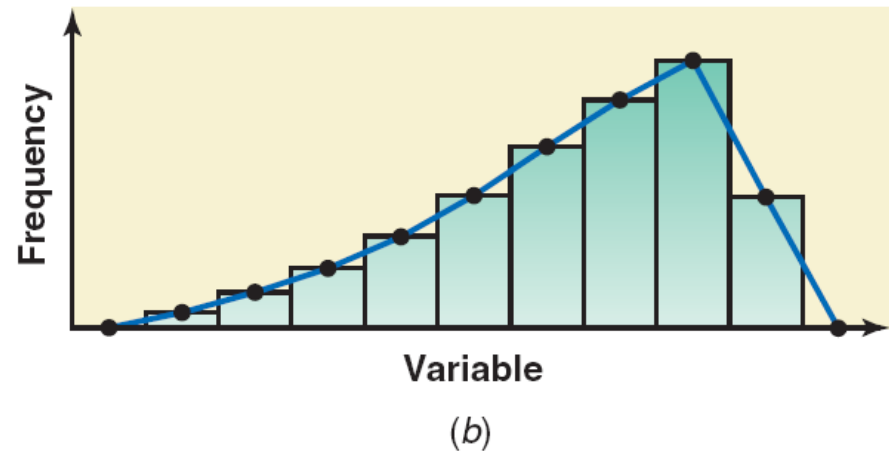
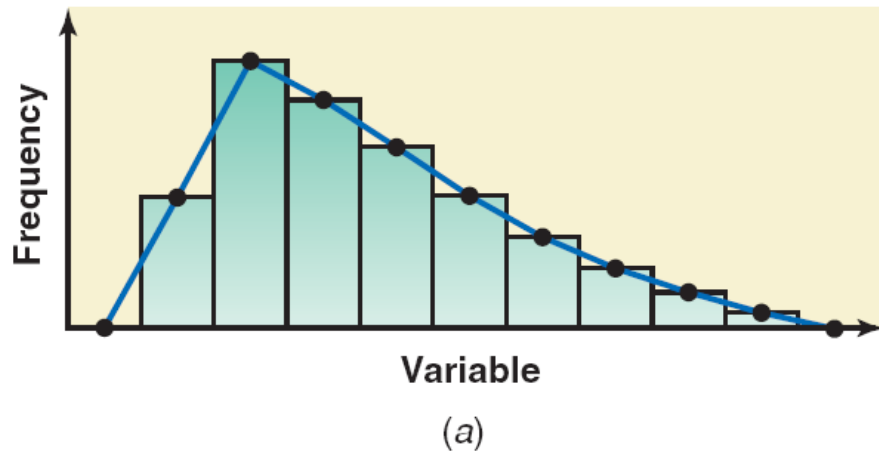
OBLICI HISTOGRAMA

1. Simetričan
2. Asimetričan
3. Uniformni ili pravougaoni

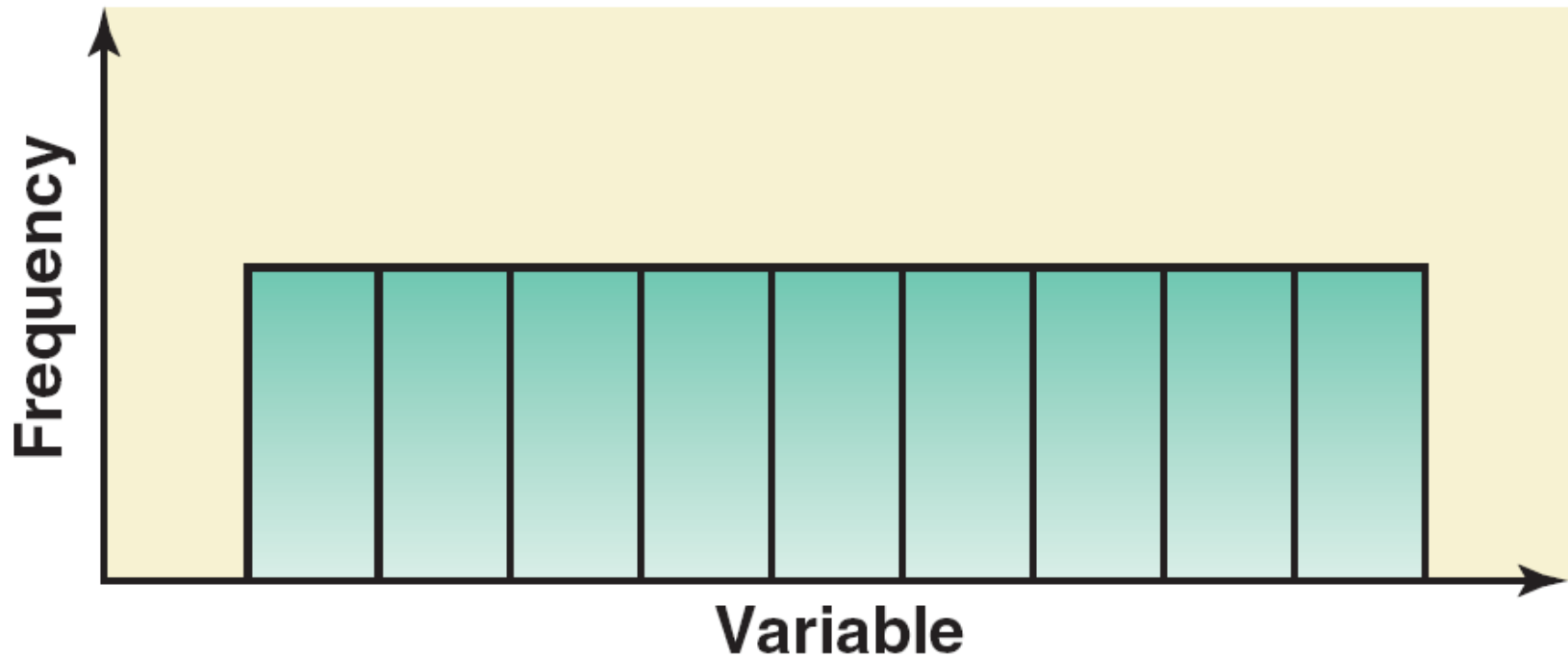
Slika 2.7 Simetrični histogrami.



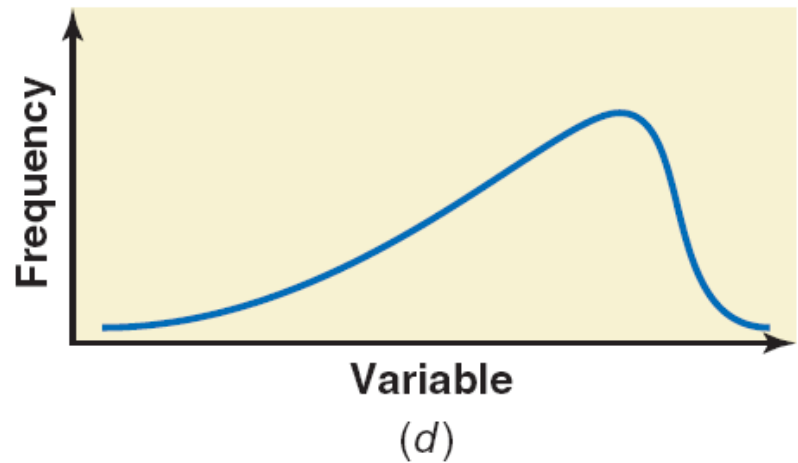
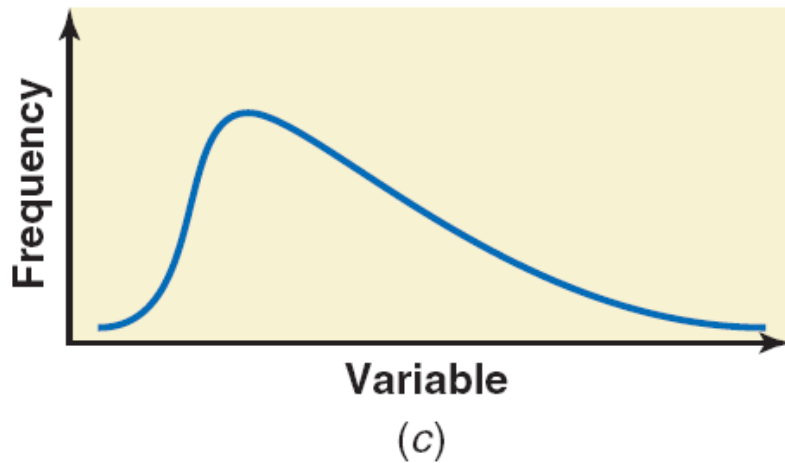
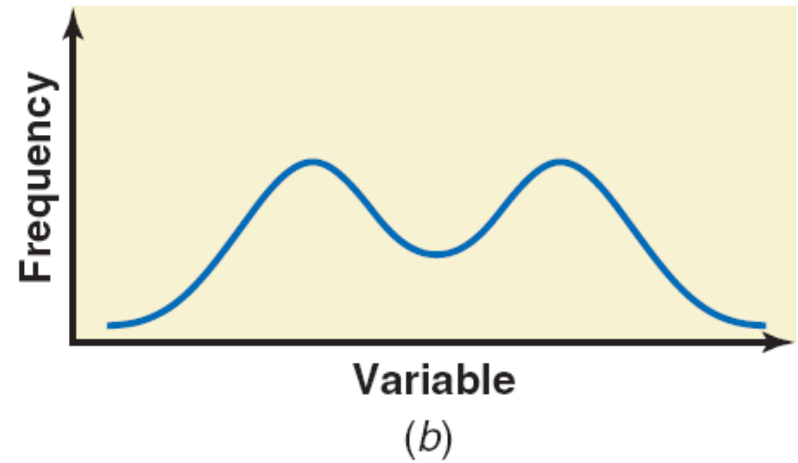
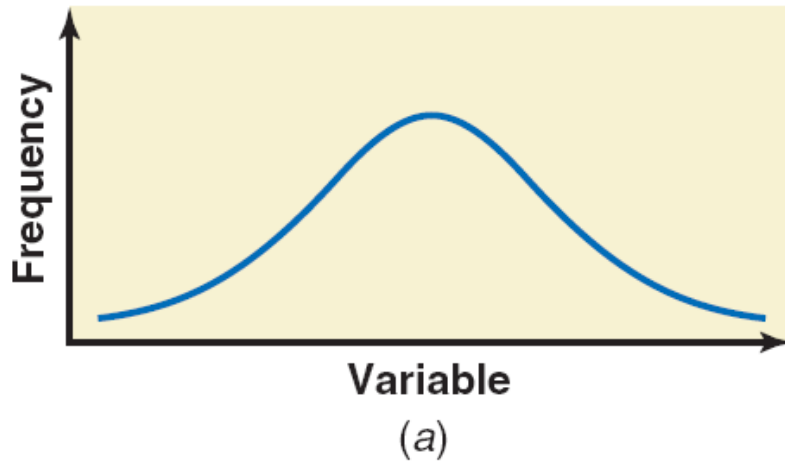
Slika 2.8 (a) Histogram asimetričan udesno. (b) Histogram asimetričan ulijevo.



Slika 2.9 Histogram za uniformnu raspodjelu.



Slika 2.10 (a) i (b) Simetrične krive frekvencija. (c) Kriva frekvencija asimetrična udesno. (d) Kriva frekvencija asimetrična ulijevo.



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

Izračunavanje kumulativne relativne frekvencije i kumulativnog učešća (procenta)

Kumulativna relativna frekvencija = $\frac{\text{Kumulativna frekvencija grupnog intervala}}{\text{Ukupan broj podataka}}$

Kumulativno učešće = (kumulativna relativna frekvencija) x 100

Tabela 2.11 Raspodjela kumulativnih frekvencija i kumulativna procentualna raspodjela prodaje iPoda

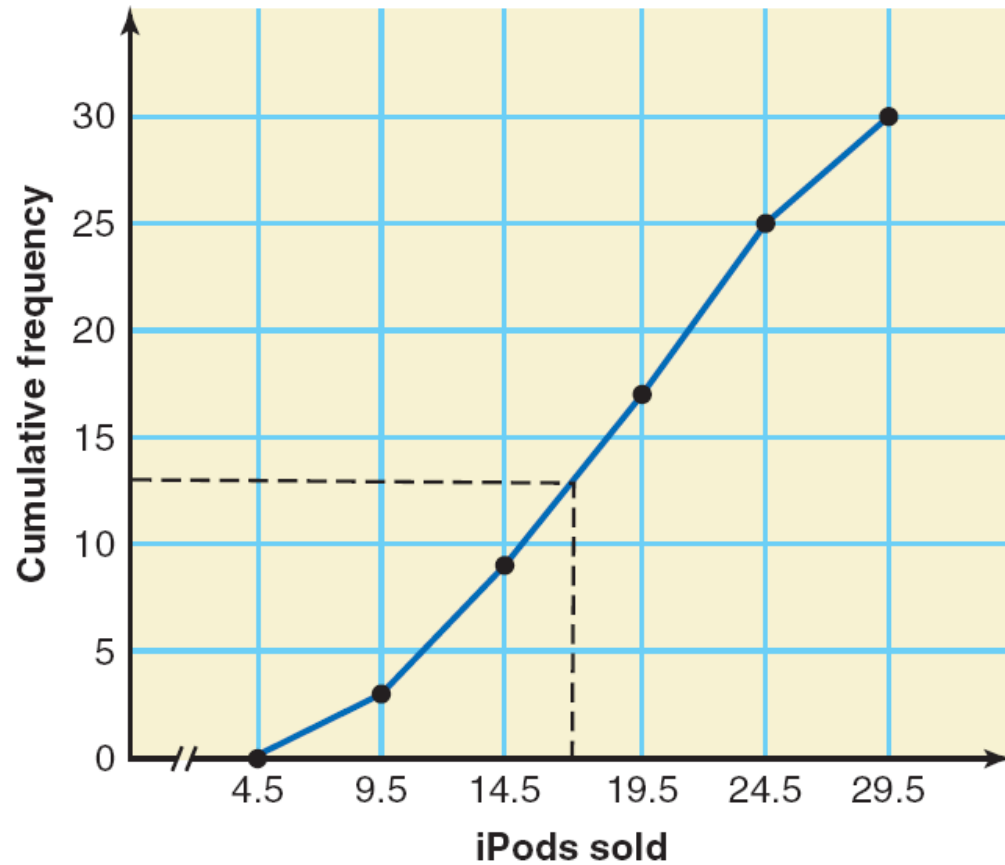
Grupni interval	Kumulativna relativna frekvencija	Kumulativni procenat
5-9	$3/30=0.100$	10.0
10-14	$9/30=0.300$	30.0
15-19	$17/30=0.567$	56.7
20-24	$25/30=0.833$	83.3
25-29	$30/30=1.000$	100.0

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.

Primjer 2-6

Dati su podaci o ostvarenim bodovima 30 studenata na ispitu iz statistike:

75	52	80	96	65	79	71	87	93	95
69	72	81	61	76	86	79	68	50	92
83	84	77	64	71	87	72	92	57	98

Grafički prikazati podatke u obliku stabljike i lista.

Primjer 2-6: Rješenje

Da bi se ovi podaci grafički prikazali pomoću stablo-list dijagrama, svaki podatak dijelimo na dva dijela. Prvi dio se odnosi na prvu cifru, koja se zove *stablo*. Drugi dio sadrži drugu cifru, koja predstavlja *list*. Na osnovu ovih podataka uočavamo da stabla za bodove uzimaju cifre 5, 6, 7, 8, i 9 jer se bodovi nalaze u rasponu od 50 do 98.

Slika 2.12 Stablo-list dijagram.



Primjer 2-6: Rješenje

Nakon što smo izlistali stabla, čitamo listove za sve bodove i bilježimo ih pored odgovarajućeg stabla sa desne strane odgovarajuće linije. Kompletan stablo-list dijagram bodova je prikazan na Slici 2.13.

Slika 2.13 Stablo-list dijagram za broj poena na ispitu.

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

Primjer 2-6: Rješenje

Listovi za svako stablo na dijagramu sa Slike 2.14 su *rangirani* (po rastućem redosljedu) i prikazani na Slici 2.14.

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**.

Primjer 2-7

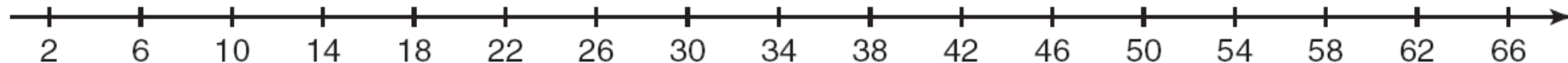
U Tabeli 2.16 dati su brojevi minuta za koje je svaki igrač Boston Bruins hokej tima kažnjen za vrijeme plejofa šampionata Stanley Cup u 2011. godini. Formirati tačkasti dijagram za ove podatke.

Tabela 2.12 Broj kaznenih minuta za igrače Boston Bruins hokej tima za vrijeme plejofa Stenli kupa u 2011.

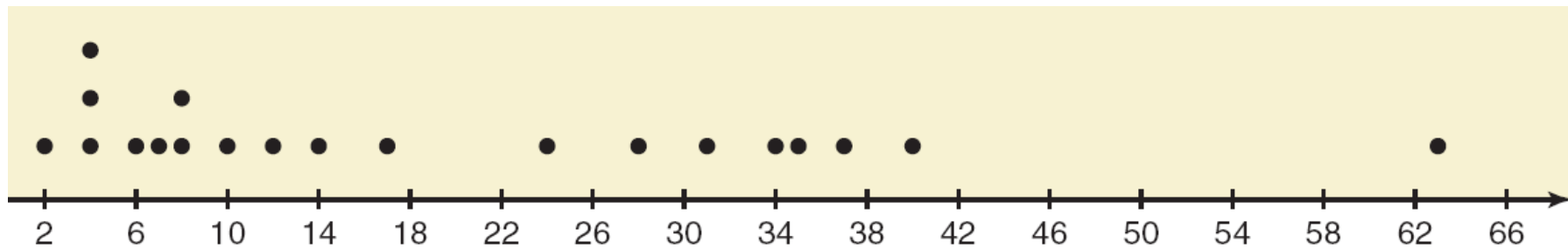
Name	Penalty Minutes	Name	Penalty Minutes
Adam McQuaid	14	Michael Ryder	8
Andrew Ference	37	Milan Lucic	63
Brad Marchand	40	Nathan Horton	35
Chris Kelly	6	Patrice Bergeron	28
Daniel Paille	4	Rich Peverley	17
David Krejci	10	Shane Hnidy	7
Dennis Seidenberg	31	Shawn Thornton	24
Gregory Campbell	4	Tomas Kaberle	4
Johnny Boychuk	12	Tyler Seguin	2
Mark Recchi	8	Zdeno Chara	34

Primjer 2-7: Rješenje

Korak 1. Nacrtati horizontalnu brojnu osu koja će biti označena brojevima kao na slici dolje.



Korak 2. Ucrtati tačke iznad vrijednosti na brojnoj osi, koje će predstavljati odgovarajuće kaznene minute iz date tabele. Kada se ucrtaju sve tačke, slika ispod prikazuje kompetan tačkasti dijagram.



Primjer 2-7: Rješenje

Na osnovu tačkastog dijagrama na slici, možemo uočiti da postoje dvije grupe (klastera) podataka. Šezdeset procenata igrača imalo je 17 ili manje kaznenih minuta za vrijeme plejofa, dok je ostalih 40% imalo 24 ili više kaznenih minuta.