

Opisivanje skupova podataka

Veoma je važno da se .numerički pronalasci bilo kakvog istraživanja predstave jasno i precizno i da budu prikazani tako da omoguće brz uvid u esencijelni karakter podataka. Ovo je posebno potrebno kada je skup podataka ograničen, što je često i slučaj kada su u pitanju kontrolni eksperimenti. Zaista, efikasna prezentacija podataka često brzo otkriva važne osobine kao što su rang, stepen simetrije, na koji način je koncentracija raspoređena, gdje je koncentracija najgušća i slično.

Frekventne tabele i grafikon;

(#) Sljedeći podaci predstavljaju broj dana bolovanja koje su uzeli svaki od 50 radnika neke firme u zadnjih 6 sedmica:

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

- Konstruisati frekventnu tabelu za ove podatke.
- Koliko radnika je imalo najmanje jedan dan bolovanja?
- Koliko radnika je imalo između 3 i 5 dana bolovanja?
- Koliko radnika je imalo više od 5 dana bolovanja?
- Predstaviti podatke pomoću linijskog grafa.
- Predstaviti podatke pomoću bar grafikona.
- Predstaviti podatke pomoću poligona frekvencija.

Rj.

- S obzirom da oni podaci sadrže samo relativno malo različitih brojeva, ili različitih vrijednosti, to najpogodniji način za prikaz podataka je u frekventnoj tabeli, koja sadrži svaku od različite vrijednosti kao i frekvenciju pojavljivanja. Sljedeća tabela je frekventna tabela datih podataka. U njoj kolona frekvencije predstavlja broj pojavljivanja svake od različite vrijednosti datog skupa podataka. Primjetimo da je suma svih frekvencija 50, ukupan broj posmatranih podataka.

Frekventna tabela podataka o bolovanju

Vrijednost	Frekvencija	Vrijednost	Frekvencija
0	12	5	8
1	8	6	0
2	5	7	5
3	4	8	2
4	5	9	1

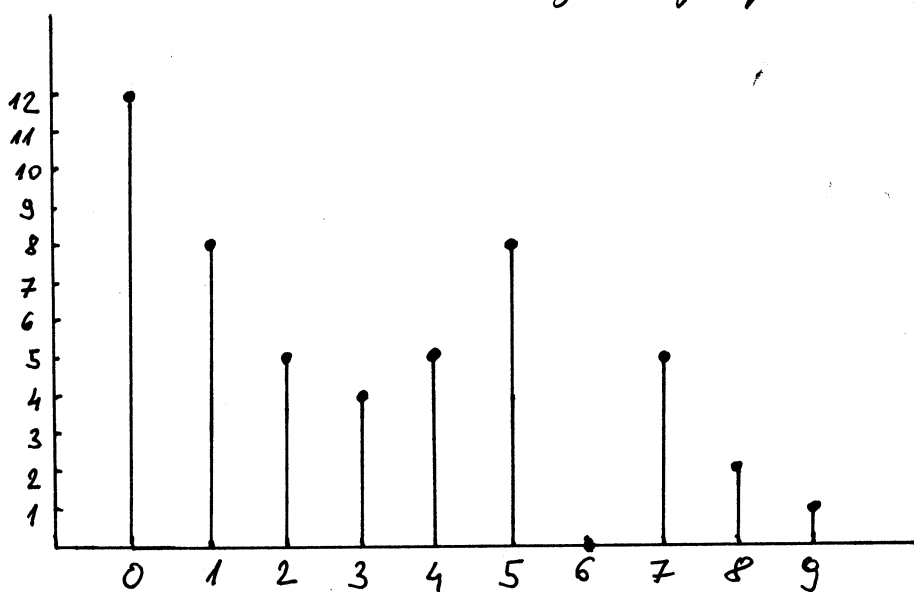
b) Kako 12 od 50 radnika nije imalo ni jedan dan bolovanja, odgovor je $50 - 12 = 38$

c) Odgovor je suma frekvencija za vrijednosti 3, 4 i 5;
tj. $4 + 5 + 8 = 17$

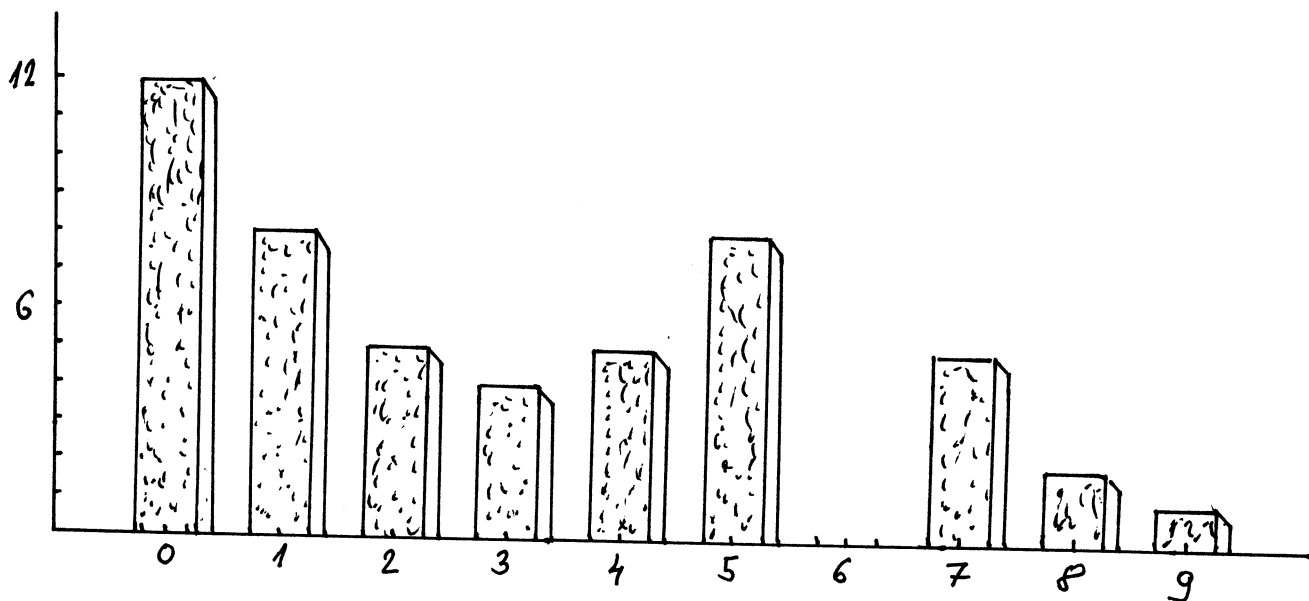
d) Odgovor je suma frekvencija za vrijednosti 6, 7, 8 i 9;
tj. $0 + 5 + 2 + 1 = 8$.

e) Podaci iz frekventne tabele se mogu grafički prikazati pomoću linijskog grafa, koji prikazuje uzastopne vrijednosti na horizontalnoj osi kao i odgovarajuće frekvencije pomoću visine vertikalne linije. Linijski graf je prikazan na sljedećoj slici.

Linijski graf
za broj dana uzetih
za bolovanje

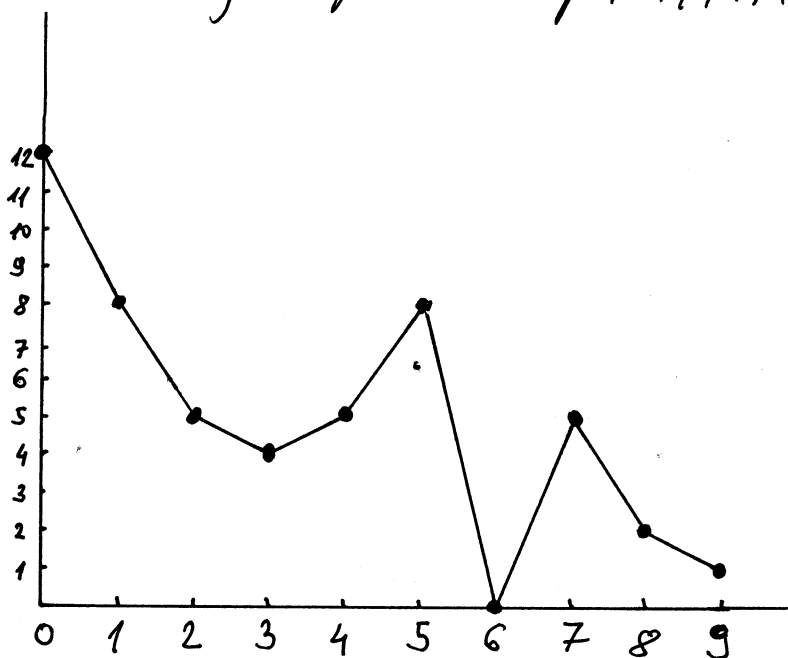


f) Ponekad se frekvencije ne predstavljaju pomoću linija nego pomoću pomoću pravougaonika sa nekom sjerom. Ove grafove nazivamo bar grafikoni, i primjer ovog grafikona za date podatke je prikazan na slici.



Bar grafikoni za broj dana bolovanja

g) Druga vrsta grafika koji se koristi za prikaz frekvencije tabele je poligon frekvencija, koji prikazuje frekvencije podataka različitih vrijednosti i onda spaja nacrtane tačke pravim linijama. Sljedeća slika prikazuje poligon frekvencija podataka prikazanih u frekventnoj tabeli:



Poligon frekvencije za broj dana uzetih za bolovanje

(#) Nacrtati relativnu frekventnu tabelu za podatke iz prethodnog zadatka. Grafički prikazati ove relativne frekvencije pomoću linijskog grafa, te nacrtati poligon relativnih frekvencija.

2. Frekventna tabela za broj dana uzetih za bolovanje iz prethodnog zadatka je

Vrijednost	Frekvencija	Vrijednost	Frekvencija
0	12	5	8
1	8	6	0
2	5	7	5
3	4	8	2
4	5	9	1

Ponekad je pogodnije razmatrati i prikazati relativne umjesto apsolutnih frekvencija vrijednosti podataka. Ako f predstavlja frekvenciju pojavljivanja nekog podatka vrijednosti x , tada relativna frekvencija $\frac{f}{n}$ se može nacrtati na mjesto x , gdje n predstavlja ukupan broj posmatranja u skupu podataka. Za podatke iz tabele iznad, imamo da je $n=50$ i odgovarajuće relativne frekvencije su date u sljedećoj tabeli. Primjetimo da, dok je suma vrijednosti iz frekventne kolone jednaka ukupnom broju posmatranja u skupu podataka, suma relativnih frekvencija u koloni je uvijek 1.

Relativne frekvencije, $n=50$, broja dana korištenih za bolovanje.

Vrijednost x	Frekvencija f	Relativna frekvencija $\frac{f}{n}$
----------------	-----------------	-------------------------------------

0	12	$\frac{12}{50} = 0,24$
---	----	------------------------

1	8	$\frac{8}{50} = 0,16$
---	---	-----------------------

2	5	$\frac{5}{50} = 0,10$
---	---	-----------------------

3	4	$\frac{4}{50} = 0,08$
---	---	-----------------------

4	5	$\frac{5}{50} = 0,10$
---	---	-----------------------

5	8	$\frac{8}{50} = 0,16$
---	---	-----------------------

6	0	$\frac{0}{50} = 0,00$
---	---	-----------------------

7	5	$\frac{5}{50} = 0,10$
---	---	-----------------------

8	2	$\frac{2}{50} = 0,04$
---	---	-----------------------

9	1	$\frac{1}{50} = 0,02$
---	---	-----------------------

Kako konstruisati relativnu tabelu frekvencija iz datog skupa podataka?

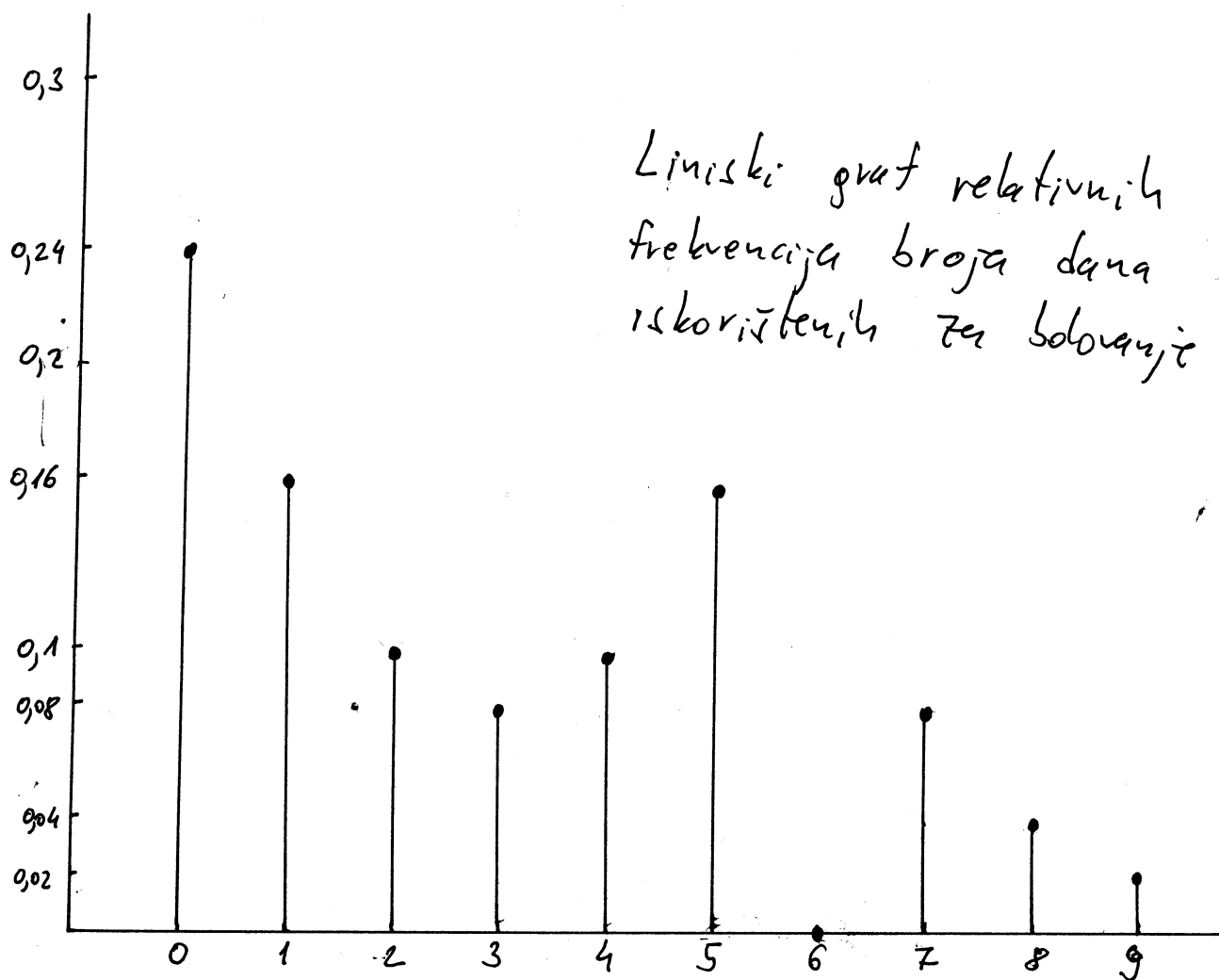
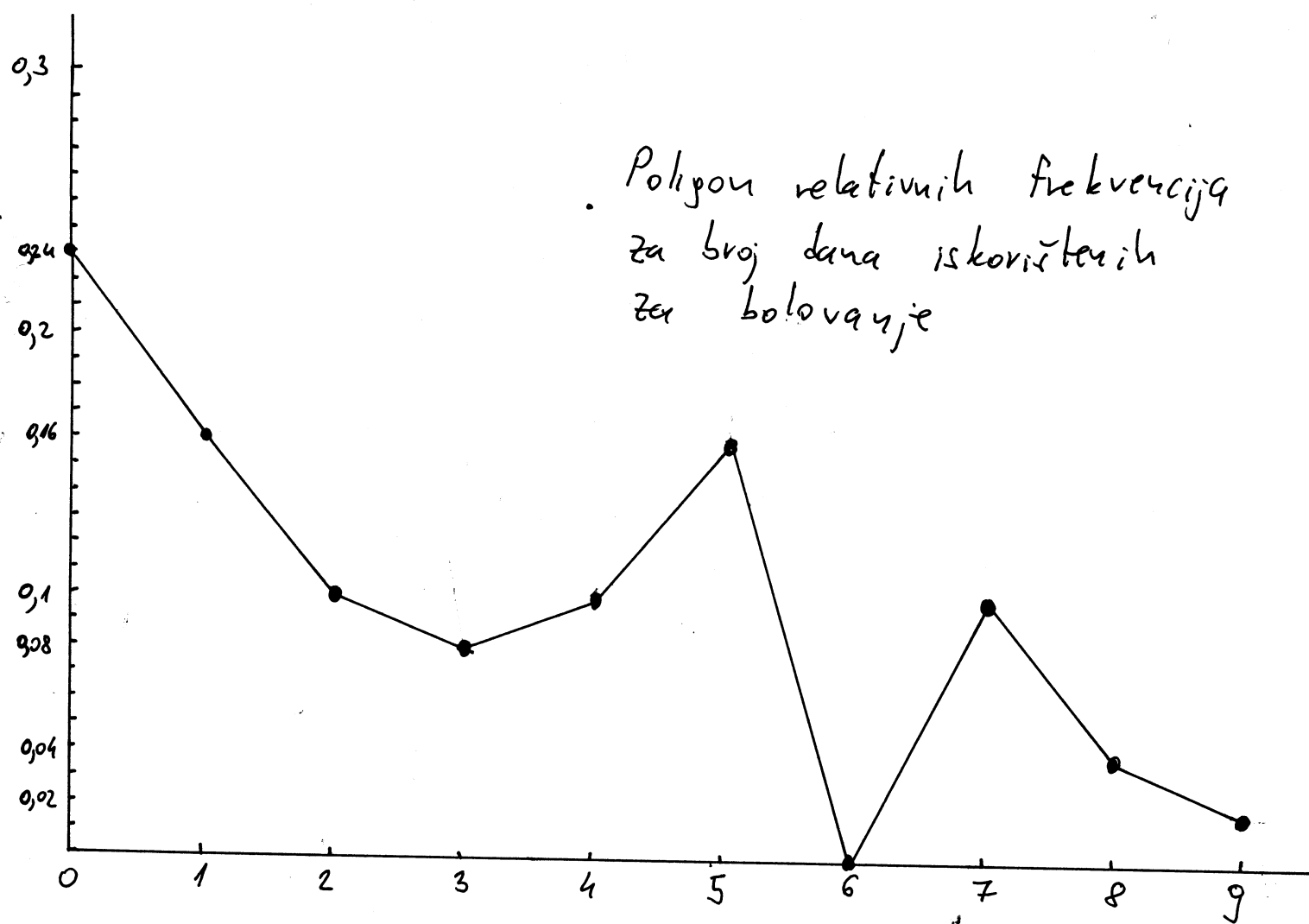
Poredati podatke iz skupa u rastućem poretku vrijednosti.

Odrediti različite vrijednosti i broj njihovih pojavljivanja.

Prikazati ove različite vrijednosti pored njihovih frekvencija

f i njihovih relativnih frekvencija $\frac{f}{n}$, gdje je n

ukupan broj posmatranja u skupu podataka.

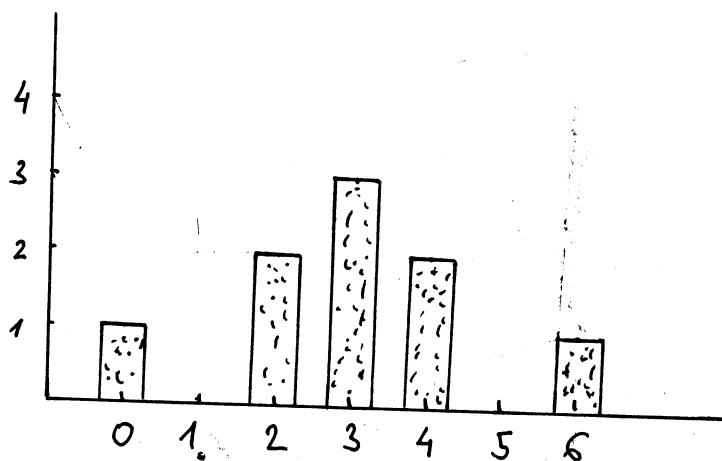


Ⓝ Za skup podataka kažemo da je simetričan oko vrijednosti x_0 ako su frekvencije vrijednosti $x_0 - c$ i $x_0 + c$ iste za sve c . Proveriti da li podaci iz sljedeće tabele imaju simetriju

Vrijednost	Frekvencija	Vrijednost	Frekvencija
0	1	4	2
2	2	6	1
3	3	—	—

Rj.

Drugim riječima, podaci su simetrični oko vrijednosti x_0 ako za svaku konstantu c , postoji tačno onoliko tačaka koje su c manje od x_0 , kao i onih koje su c veće od x_0 . Ovo je najlakše pročitati iz grafikona, pa nacrtajmo bar grafikon za datu tabelu



Podaci dati u frekventnoj tabeli su simetrični oko vrijednosti $x_0 = 3$.

(#) Golf Turnir Piraka se igra svake godine u Piramida Nacionalnom Golf Klubu u Visokom, B.H. Da bi otkrili koliko bodova treba ostvariti da bi pobijedili na ovom turniru, sakupili smo sve pobjedničke rezultate od 1968 do 2004.

Godina	Pobjednik	Broj bodova	Godina	Pobjednik	Broj bodova
1968	B. G.	277	1987	P. S.	285
69	G. A.	281	88	M. M.	281
70	A. B.	279	89	M. E.	283
71	B. D.	279	90	F. B.	278
72	O. E.	286	91	N. C.	277
73	E. F.	283	92	P. M.	275
74	F. G.	278	93	P. S.	277
75	G. Z.	276	94	S. M.	279
76	Z. S.	271	95	B. Dž.	274
77	S. T.	276	96	Z. Dž.	276
78	T. U.	277	97	H. S.	270
79	U. V.	280	98	B. N.	279
80	V. D.	275	99	A. T.	280
81	D. H.	280	2000	M. N.	278
82	H. P.	284	01	N. N.	272
83	P. S.	280	02	B. B.	276
84	T. S.	277	03	M. M.	281
85	T. T.	282	2004	P. S.	279
1986	T. E.	279			

- (a) Podatke iz tabele pobjedničkih rezultata predstaviti pomoću relativne frekventne tabele.
- (b) Nacrtati ove podatke pomoću relativnog frekventnog bar grafikona.

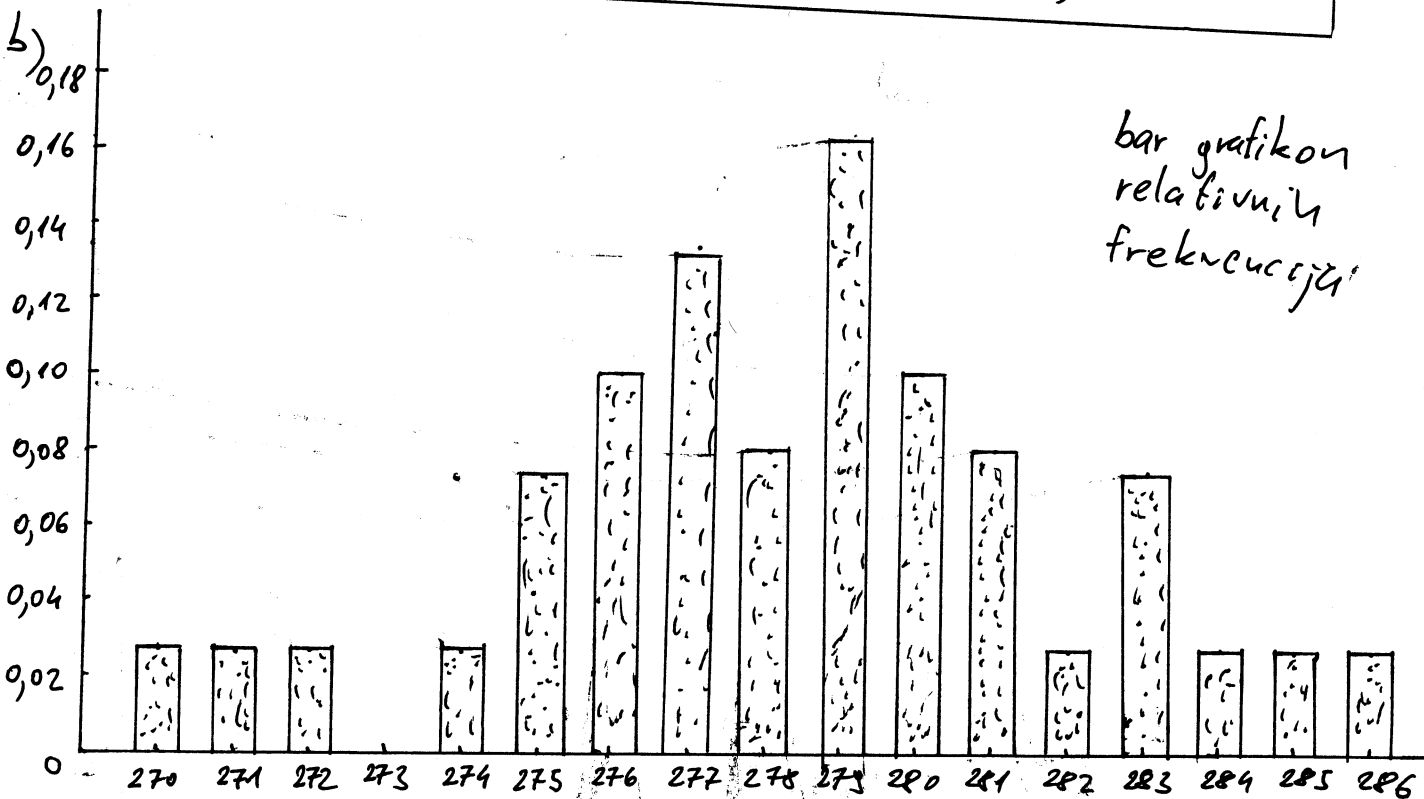
Rj. 37 datih pobjedničkih rezultata uzimaju vrijednosti od 270 do 286. Odgovarajuća relativna frekventna tabela je:

Pobjednički rezultat

Frekvencije f

Relativna
frekvencije
 $f/37$

270	1	0,027
271	1	0,027
272	1	0,027
274	1	0,027
275	2	0,054
276	4	0,108
277	5	0,135
278	3	0,081
279	6	0,162
280	4	0,108
281	3	0,081
282	1	0,027
283	2	0,054
284	1	0,027
285	1	0,027
286	1	0,027

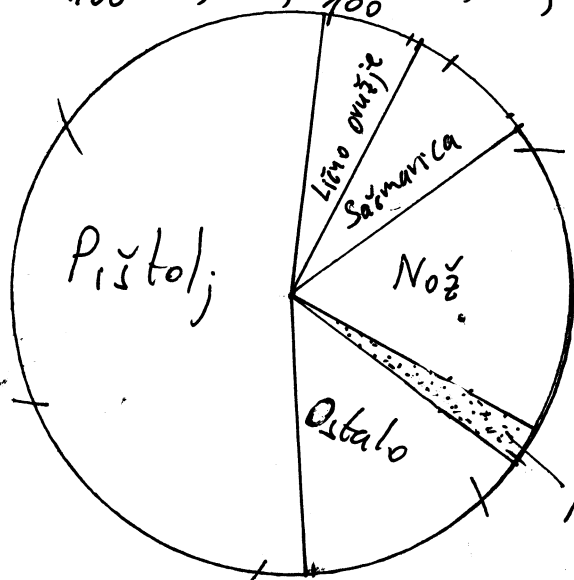


Podatke iz sljedeće tabele predstaviti pomoću zvrk grafikona

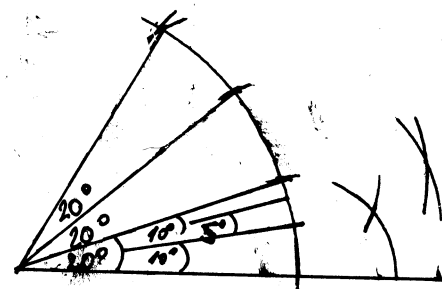
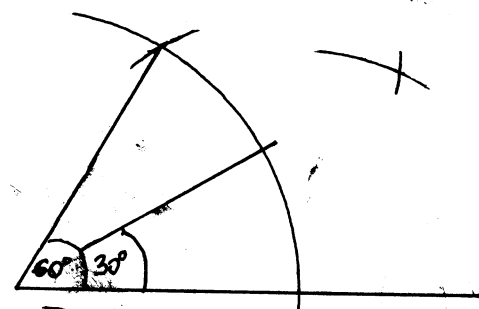
Tip oružja	Procenat ubistava u vedem zapadnom gradu uzrokovana ovim oružjem
Pištoli	52
Nož	18
Sačmarica	7
Puška	4
Lično oružje	6
Ostalo	13

zj. Zvrk grafikon se često koristi za prikaz relativnih frekvencija kada su podaci nenumerički. Nacrta se krug i onda se podijeli na različite odsječke (sektore), svaki koji predstavlja različitu vrijednost podatka. Površina svakog sektora, koja bi trebala da predstavlja relativnu frekvenciju vrijednosti koju taj sektor predstavlja, se određuje na sljedeći način.

$$\frac{52}{100} = 0,52; \quad \frac{18}{100} = 0,18; \quad \frac{7}{100} = 0,07; \quad \frac{4}{100} = 0,04;$$



$$\begin{aligned} 0,52 \cdot 360^\circ &= 187,2 \\ 0,18 \cdot 360^\circ &= 64,8 \\ 0,07 \cdot 360^\circ &= 25,2 \\ 0,04 \cdot 360^\circ &= 14,4 \\ 0,06 \cdot 360^\circ &= 21,6 \\ 0,13 \cdot 360^\circ &= 46,8 \end{aligned}$$



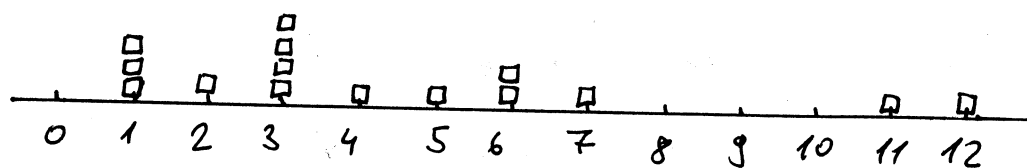
Zadaci za vježbu

1. Sljedeći podaci predstavljaju veličine 30 porodica koje su odrasle u malom gradu u Gvatemali:

5, 13, 9, 12, 7, 4, 8, 6, 6, 10, 7, 11, 10, 8, 15,
8, 6, 9, 12, 10, 7, 11, 10, 8, 12, 9, 7, 10, 7, 8.

- (a) Konstruirati frekventnu tabelu za ove podatke
- (b) Prikazati podatke koristeći linijski grafik
- (c) Nacrtati podatke pomoću poligona frekvencija.

2. Petnaestorici srednjoškolaca iz četvrtog razreda je postavljeno pitanje koliko kilometara ^{im je mjesto gdje žive} udaljeno od škole. Rezultati su prikazani u sljedećem grafu.



- (a) Koliki je najveći broj kilometara nekog studenta mjesto gdje živi udaljeno od škole?
- (b) Koliki je najmanji broj kilometara?
- (c) Koliko učenika živi manje od 5 kilometara od škole?
- (d) Koliko učenika živi više od 4 kilometara od škole?

3. Tabela desno prikazuje ^{sve vrijednosti ali} samo neke frekvencije za simetričan skup podataka. Popuniti brojeve koji nedostaju.

Vrijednost	Frekvencija
10	8
20	
30	7
40	
50	3
60	

④ Nacrtati relativnu tabelu frekvencija za podatke iz Problema 1. Ove relativne frekvencije prikazati pomoću linijskog grafa.

⑤ Sljedeća tabela relativnih frekvencija je dobijena iz skupa podataka o broju hitnih operacija slijepog crijeva koje se izvode svakog mjeseca u jednoj bolnici

Vrijednost	0	1	2	3	4	5	6	7
Relativna frekvencija	0,05	0,08	0,12	0,14	0,16	0,20	0,15	0,10

- (a) Koliki postotak mjeseci imaju manje od dvije hitne operacije slijepog crijeva?
- (b) Koliki postotak mjeseci imaju više od 15?
- (c) Da li je ovaj skup podataka simetričan?

⑥ Sljedeća tabela prikazuje brojke smrtnosti na Bosanskim cestama u 1987 za ljude u različitim klasifikacijama

Klasifikacija	Broj smrtnosti
Pješaci	1699
Biciklisti	280
Motoristi	650
Vozaci auta	1327

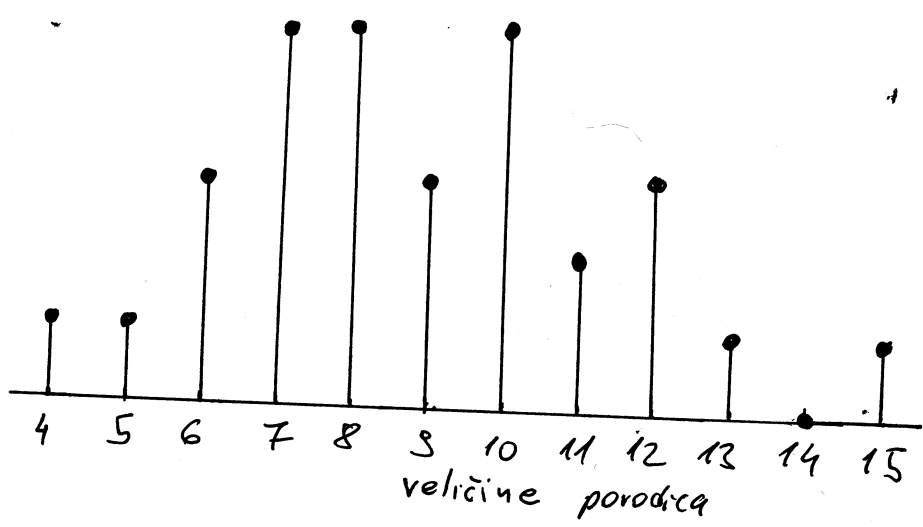
Prikazati ovaj skup podataka pomoću zvrk grafikona.

Odgovori:

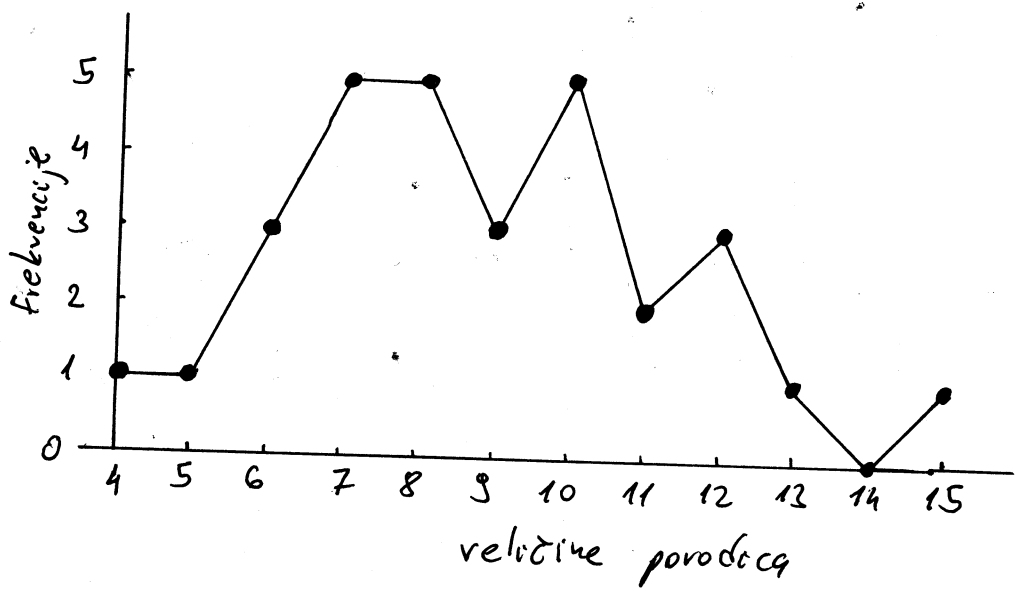
(b)

Velicina porodice	Frekvencija
4	1
5	1
6	3
7	5
8	5
9	3
10	5
11	2
12	3
13	1
14	0
15	1

(b)



(c)



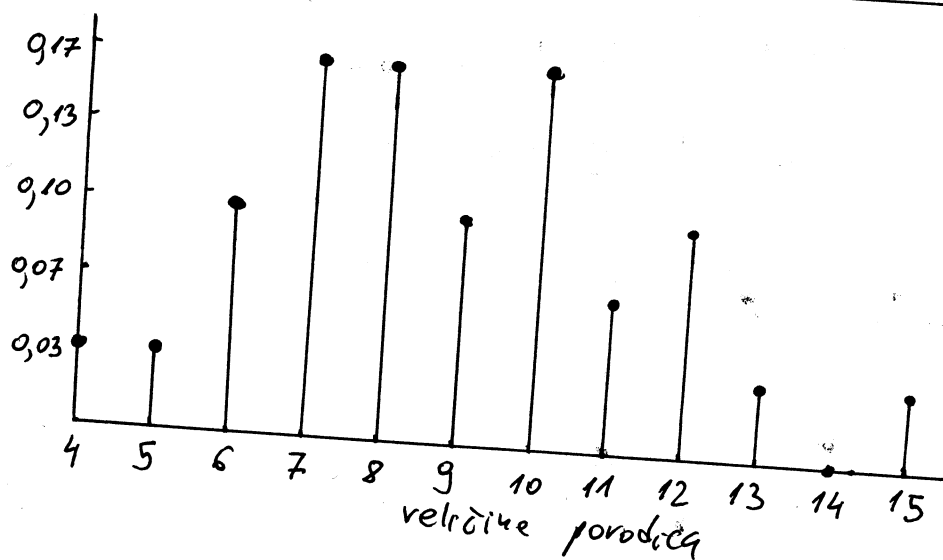
2. (a) 12
(b) 1
(c) 3
(e) 3

3.

Vrijednost	Frekvencije
10	8
20	3
30	7
40	7
50	3
60	8

4.

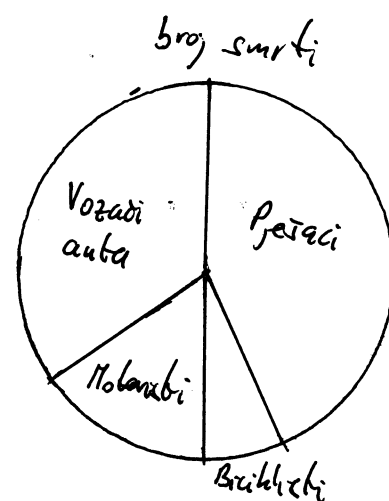
Velicina porodice	Frekvencija	Relativna frekvencija
4	1	0,03
5	1	0,03
6	3	0,10
7	5	0,17
8	5	0,17
9	3	0,10
10	5	0,17
11	2	0,07
12	3	0,10
13	1	0,03
14	0	0,00
15	1	0,03



5.

- (a) 0,13
(b) 0,25
(c) Ne

6.



Grupirani podaci i histogrami

Kao što smo vidjeli u prethodnoj lekciji koristeći linijski ili bar grafikon za prikaz frekvencija vrijednosti podataka je često efikasan način za portiranje skupa podataka. Bez obzira, za neke skupove podataka broj različitih vrijednosti je prevelik da primijenimo ovaj pristup. Umjesto toga, u tom slučaju, mi ćemo podijeliti vrijednosti u grupe, ili klase intervala, pa tek onda ćemo prikazati broj vrijednosti podataka koja pripada svakoj od ovih klasa intervala. Broj klasa intervala koje ćemo izabrati će imati dvije oskrice

- (1) ako izaberemo premalo klasa tada ćemo izgubiti previše informacija o stvarnim vrijednostima podataka u klasi;
- (2) ako izaberemo previše klasa tada ćemo, zbog premalenog broja frekvencija u svakoj klasi, kao rezultat dobiti uzorak iz kojeg je teško vidjeti strukturu.

Iako je od 5 do 10 klasa intervala uobičajeno, približan broj je subjektivan izbor, i naravno možete eksperimentirati sa različitim brojem klasa intervala i posmatrati koji od dobijenih grafikona najviše odgovara podacima. Vrlo je uobičajeno, iako ne esencijalno, da su izabrane klase intervala jednake dužine.

Sljedeća tabela predstavlja nivo holesterola u krvi 40 studenata druge godine Politehničkog fakulteta.

Tabela 1 - Nivo holesterola u krvi							
213	174	193	196	220	183	194	200
192	200	200	199	178	183	188	193
187	181	193	205	196	211	202	213
216	206	195	191	171	194	184	191
221	212	221	204	204	191	183	227

- (a) Predstaviti ovaj skup podataka pomoću histograma frekvencija
 (b) Navesti glavne korake za konstrukciju histograma iz datog skupa podataka
 (c) Navesti nekoliko stvari koje mogu biti pročitane iz histograma (u opštem slučaju).

Rj. (a) Prije nego što odredimo veličine klasa frekvencija, konvencija je da poredamo ^{date} podatke u rastućem redu. Time dobijemo sljedeću tabelu

Tabela 2 - Nivo holesterola u rastućem poretka										
171,	174,	178,	181,	183,	183,	183,	184,	187,	188,	191,
191,	191,	192,	193,	193,	193,	194,	194,	195,	196,	196
199,	200,	200,	200,	202,	204,	205,	206,	211,	212,	213,
213,	216,	220,	221,	221,	227					

S obzirom da rang podataka ima najmanju vrijednost 171, a najveću vrijednost 227, krajnja-lijeva granica prve klase intervala mora biti manja ili jednaka 171, a desna-krajnja

granica zadnje klase intervala mora biti veća od 227. Jedan izbor bi bio da za prvu klasu uzmemo interval $[170, 180)$. Time ćemo dobiti šest klasa intervala. Tabela frekvencija (frekventna tabela) datih frekvencija (kao i relativnih frekvencija) vrijednosti podataka koje pripadaju svakoj klasi intervala je prikazana u sljedećoj tabeli.

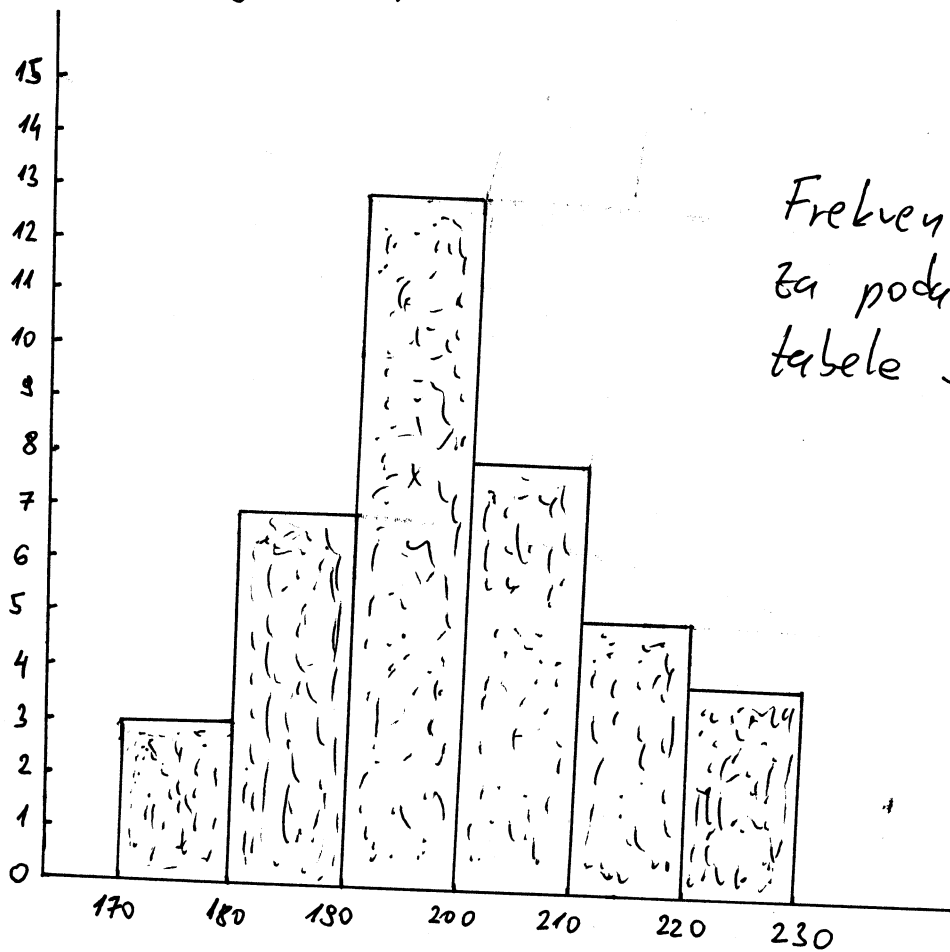
Tabela 3 - Frekventna tabela nivoa holesterola u krvi

Klasa intervala	Frekvencija	Relativna frekvencija
$[170, 180)$	3	$\frac{3}{40} = 0,075$
$[180, 190)$	7	$\frac{7}{40} = 0,175$
$[190, 200)$	13	$\frac{13}{40} = 0,325$
$[200, 210)$	8	$\frac{8}{40} = 0,20$
$[210, 220)$	5	$\frac{5}{40} = 0,125$
$[220, 230)$	4	$\frac{4}{40} = 0,10$

Primetimo da $200 \in [200, 210)$ a da $200 \notin [190, 200)$.

Bar grafikom podataka, sa pravougaonima nacrtaanih jedan pored drugog, se naziva histogram. Vertikalna osa histograma može predstavljati ili klasu frekvencija ili relativnu klasu frekvencija. U prvom slučaju histogram

se naziva frekventni histogram a u drugom relativni; frekventni; histogram. Sljedeća slika predstavlja frekventni; histogram podataka iz Tabele 3.



Frekventni; histogram
za podatke iz
tabele 3

Važno je prepoznati da klase frekventne tabele, ili histograma baziranoj na toj tabeli, ne sadrži sve informacije iz originalnog skupa podataka. Ove dvije reprezentacije bilježe samo broj vrijednosti podataka svake klase, a ne stvarne vrijednosti podataka između sebe. Time, bez obzira što su takve tabele i grafikon; korisni za prikaz podataka, originalni red skupa podataka se uvijek mora sačuvati.

b) Kako konstruirati histogram iz skupa podataka?

1. Složiti podatke u rastućem poretku
2. Izabrati klasu intervala tako da su sve tačke podataka pokrivena

3. Konstruisati frekventna tabelu

4. Nacrtati susjedne pravougaonike čija je visina određena sa frekvencijama iz kolona 3

c) Važnost histograma je ta da omogućava organizovanje i prezentaciju podataka grafički tako da nam skrene pažnju na određene važne osobine podataka. Na primjer, histogram često može uputiti na

1. Koliko simetrični su podaci

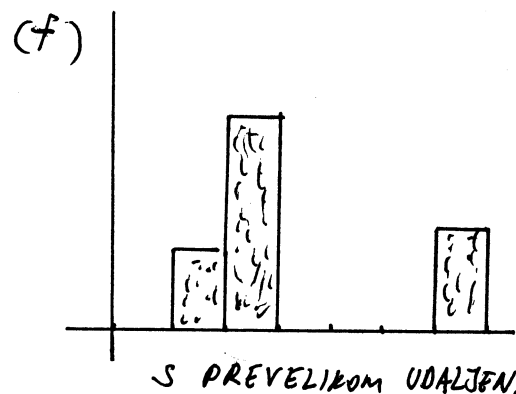
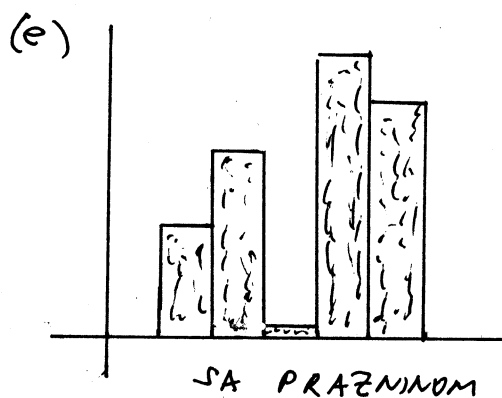
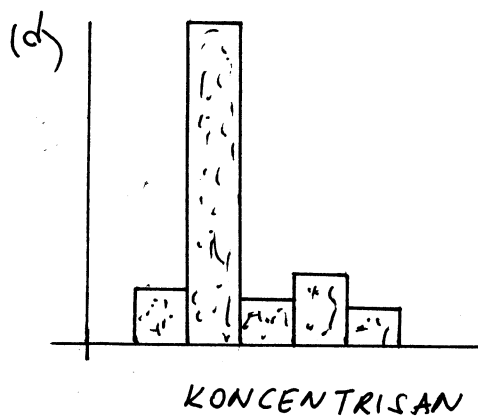
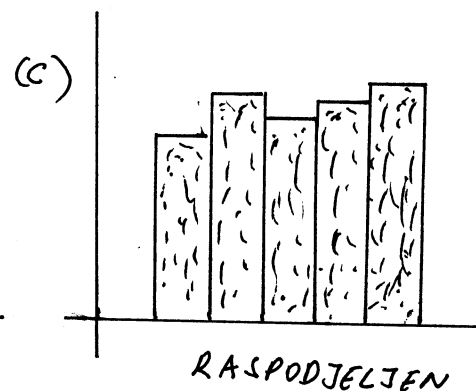
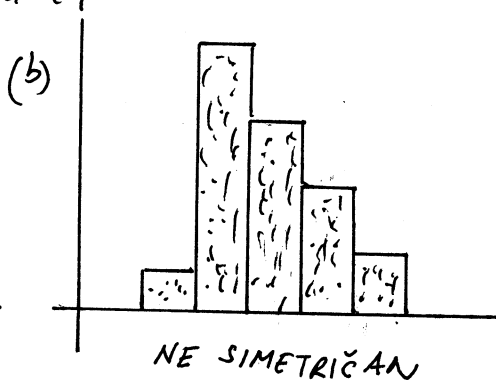
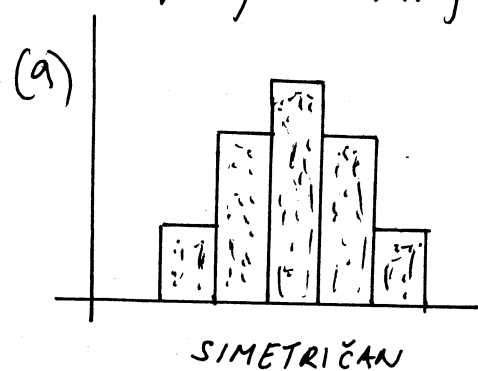
2. Na koji način su raspodjeljeni podaci

3. Da li postoje intervali koji imaju visok nivo koncentracije podataka

4. Da li postoje praznine u podacima

5. Da li su neke vrijednosti podataka previše udaljene od drugih.

Neki primjeri histograma



Data tabela prikazuje stopu nataliteta (po 1000 stanovnika) u 50 glavnih gradova država većinom iz Evrope.

Tabela 1 - Stopa nataliteta po 1000 stanovnika

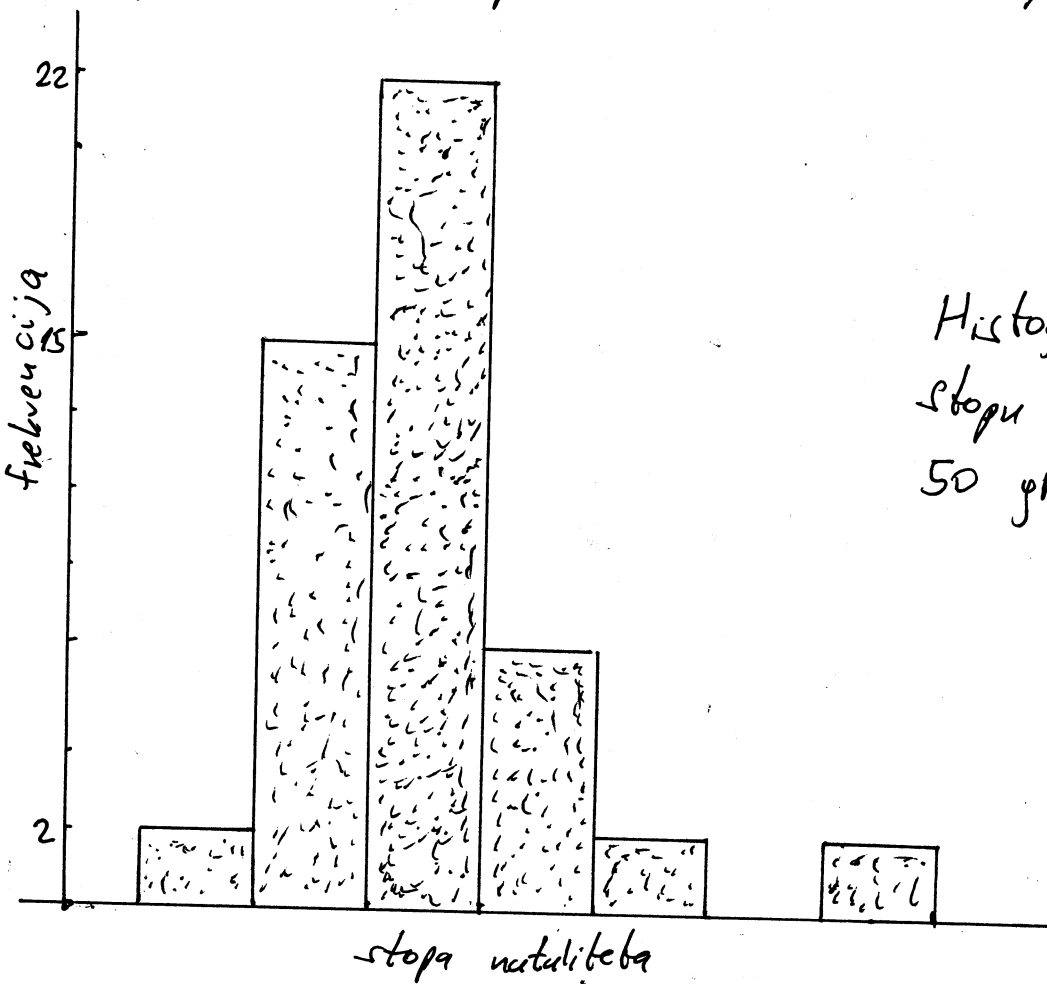
Grad	Stopa nataliteta	Grad	Stopa nataliteta	Grad	Stopa nataliteta
Tirana	14,2	Toršavn	15,7	Vilnius	14,8
Andora la Vela	21,9	Helsinki	13,8	Luksemburg	14,4
Beč	13,0	Pariz	14,4	Skoplje	15,5
Minsk	14,5	Berlin	16,3	Valetta	14,1
Brisel	13,2	Gibraltar	15,4	Čisina	15,3
Sarajevo	15,9	Atena	15,3	Monako	15,7
Sofija	14,7	Budimpešta	16,1	Podgorica	15,4
Saint Helier	17,1	Dablin	15,5	Amsterdam	15,5
Zagreb	15,2	Daylas	14,1	Oslø	17,7
Nikosia	17,1	Rim	15,1	Varšava	21,2
Prag	17,6	Priština	16,5	Lisabon	14,0
Kopenhagen	15,2	Riga	16,2	Bukurešt	15,3
Talin	16,0	Vaduz	15,1	San Marino	15,4
Beograd	14,8	Bratislava	17,9	Ljubljana	13,4
Madrid	13,1	Longerbin	16,2	Stokholm	14,8
Bern	14,2	Kijev	15,6	London	13,7
Moskva	14,1	Ankara	16,5		

Prikazati ove podatke pomoću histograma frekvencija.

Rj. S obzirom da je rang podataka od najmanje vrijednosti 13,4 do najveće vrijednosti 21,9, za klasu intervala ćemo uzeti dužinu 1,5, polazeći od vrijednosti 12. Sa ovom klasom intervala dobit ćemo sledeću tabelu frekvencija.

Interval	Frekvencija	Interval	Frekvencija
$[12,0; 13,5)$	2	$[18,0; 19,5)$	2
$[13,5; 15,0)$	15	$[19,5; 21,0)$	0
$[15,0; 16,5)$	22	$[21,0; 22,5)$	2
$[16,5; 18,0)$	7		

Histogram za ove podatke je dat na sljedećoj slici:



U biti, histogram je bar grafikon koji prikazuje frekvencije ili relativne frekvencije podataka koje pripadaju različitim klasama intervala. One klase frekvencija se također mogu grafički prikazati pomoću frekventnog (ili relativno frekventnog) poligona. Tada svaka klasa intervala je predstavljena pomoću vrijednosti, obično uzeta kao sredina tog intervala. Crtamo točke koje predstavljaju frekvenciju klase intervala i onda spajamo točke pravim linijama i dobijemo frekventni poligon.

Ⓝ) Sljedeća tabela predstavlja klase frekvencija za sistolički krvni pritisak dvije grupe industrijskih radnika muškog spola: dvije starosne grupe, prva od 30 do 40 godina i druga od 50 do 60 godina.

Tabela 1- Klase frekvencija sistoličkog krvnog pritiska dvije grupe radnika muškog spola

Krvni pritisak	Broj radnika	
	Godine od 30 do 40	Godine od 50 do 60
Manji od 90	3	1
90-99	17	2
100-109	118	23
110-119	460	57
120-129	768	122
130-139	675	149
140-149	312	167
150-159	120	73
160-169	45	62
170-179	18	35
180-189	3	20
190-199	1	9
200-209		3
210-219		5
220-229		2
230-240		1
Ukupno	2540	731

Grafčki usporediti dvije date grupe radnika.

Rj. S obzirom da je ukupan broj radnika u obe grupe različit, teško je direktno napraviti poređenje krvnog pritiska za ove dvije grupe.

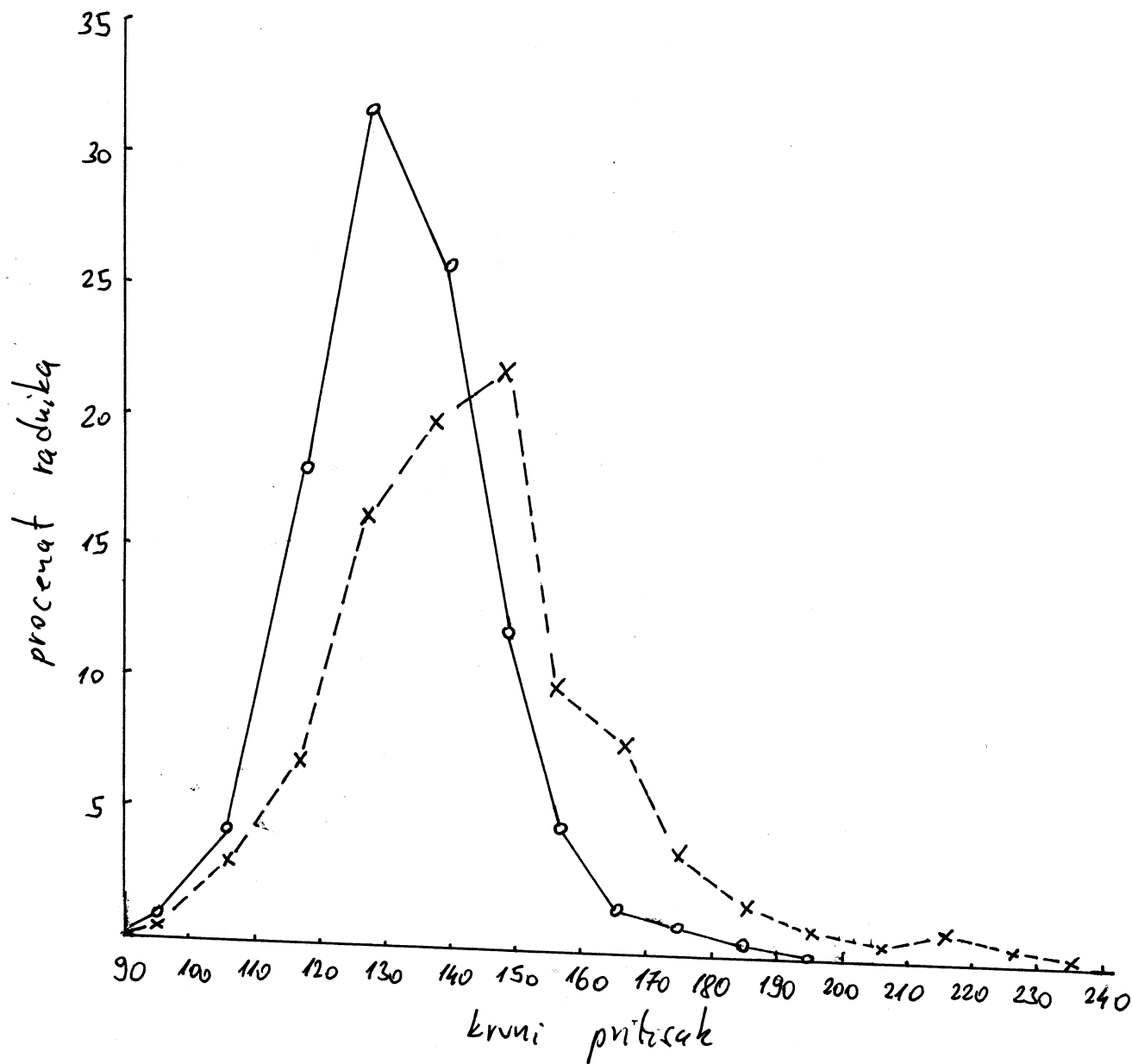
Da bi uklonili ovu teškoću možemo izračunati i grafički prikazati relativne frekvencije za svaku od klasa. To jest, podijelimo sve frekvencije koje se odnose na radnike od 30 do 39 godina sa 2540 (ukupan broj takvih radnika) i sve frekvencije koje se odnose na radnike od 50 do 59 godina sa 731. Kao rezultat imamo sljedeću tabelu.

Tabela 2- Relativne klase frekvencija krvnog pritiska

Krvni pritisak	Procenat radnika	
	Godine 30-40	Godine 50-60
<90	0,12	0,14
[90, 100)	0,67	0,27
[100, 110)	4,65	3,15
[110, 120)	18,11	7,80
[120, 130)	30,24	16,69
[130, 140)	26,57	20,38
[140, 150)	12,28	22,84
[150, 160)	4,72	9,99
[160, 170)	1,77	8,48
[170, 180)	0,71	4,79
[180, 190)	0,12	2,74
[190, 200)	0,04	1,23
[200, 210)		0,41
[210, 220)		0,68
[220, 230)		0,27
[230, 240)		0,14
Ukupno:	100.00	100.00

Sljedeća figura grafički predstavlja poligon relativnih frekvencija za obe starosne grupe. Imajući oba frekventna poligona

na istom grafu lako možemo napraviti poređenje dva skupa podataka. Na primer, čini se da je krvni pritisak starije grupe mnogo rašireniji među većim vrijednostima nego kod mlađe starosne grupe.



Relativni poligon frekvencija za podatke iz Tabele 2.

Zadaci za vježbu

1. Sljedeći skup podataka predstavljaju rezultate testa inteligencije (IQ) od 40 učenika šestih razreda neke osnovne škole:

114, 122, 103, 118, 99, 105, 134, 125, 117, 106, 109, 104,
111, 127, 133, 111, 117, 103, 120, 98, 100, 130, 141, 119,
128, 106, 109, 115, 113, 121, 100, 130, 125, 117, 119, 113,
104, 108, 110, 102

- Predstaviti ove podatke pomoću histograma frekvencija.
- Koje klase intervala sadrže najveće broj vrijednosti podataka?
- Da li ^{približno} približno jednak broj podataka u svakoj klasi intervala?
- Da li je histogram približno simetričan? Ako jest oko kojeg intervala se nalazi ova približna simetrija?

2. Sljedeći podaci predstavljaju visinu dnevne temperature (u stepenu celzijusa) na ^{obitku} 10. maju u Zenici u nizu od 30 godina

22,8; 26,2; 31,7; 31,1; 26,9; 28,0; 29,4; 28,8; 26,7;
27,4; 28,2; 30,3; 29,5; 28,9; 27,5; 28,3; 24,1; 25,3;
28,5; 27,7; 24,4; 29,2; 30,3; 33,7; 27,5; 29,3; 30,2;
28,5; 32,2; 33,7;

- Prikazati ovaj skup podataka pomoću histograma frekvencija.
- Da li biste mogli reći kolika je "tipična" temperatura prvog maja u Zenici?
- Kakvi drugačiji zaključci se mogu izvesti iz histograma?

3) Data je koncentracija ozona (mjerena po djelovima po 100 miliona) zraka u jednom dijelu ženice mjerena 25 uzastopnih dana u 2004:

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

- (a) Konstruirati histogram frekvencija za ove podatke tako da se interval $[3,5)$ nalazi među klasama intervala,
- (b) Konstruirati histogram frekvencija za ove podatke tako da se interval $[2,3)$ nalazi među klasama intervala,
- (c) Za koji frekventni histogram smatrate da sadrži više informacija.

4) Posmatrajmo nivo holesterola u krvi 100 studenata iz sljedeće tabele

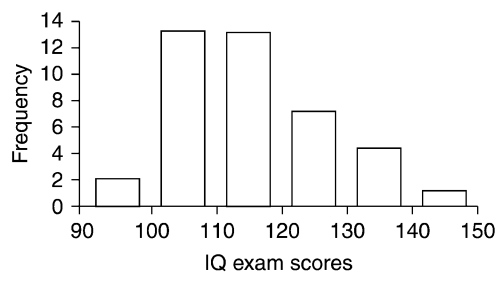
Student	Holesterol	Spol	Student	Holesterol	Spol	Student	Holesterol	Spol
1	213	ž	16	193	ž	31	184	ž
2	174	M	17	187	ž	32	191	M
3	193	ž	18	181	M	33	221	ž
4	196	ž	19	193	M	34	212	ž
5	220	ž	20	205	M	35	221	M
6	183	M	21	196	M	36	204	ž
7	194	ž	22	211	ž	37	204	ž
8	200	ž	23	202	M	38	191	M
9	192	M	24	213	ž	39	183	ž
10	200	ž	25	216	M	40	227	M
11	200	M	26	206	M	41	188	M
12	199	ž	27	195	ž	42	200	M
13	178	M	28	191	M	43	197	M
14	183	M	29	171	M	44	208	ž
15	188	ž	30	194	M	45	218	ž

Student	Holesterol	Spol	Student	Holesterol	Spol
46	194	ž	71	186	M
47	212	ž	72	186	M
48	207	ž	73	196	M
49	219	M	74	195	ž
50	201	ž	75	205	ž
51	208	M	76	223	ž
52	192	M	77	195	M
53	192	M	78	205	ž
54	191	M	79	198	M
55	196	M	80	215	M
56	193	M	81	202	M
57	185	M	82	196	M
58	201	M	83	196	M
59	185	M	84	190	ž
60	203	M	85	185	ž
61	172	M	86	188	M
62	213	ž	87	197	ž
63	192	ž	88	196	ž
64	198	ž	89	227	M
65	212	M	90	211	ž
66	188	M	91	219	M
67	201	ž	92	202	ž
68	219	M	93	204	M
69	189	M	94	213	ž
70	203	ž	95	202	ž
			96	205	M
			97	184	M
			98	198	M
			99	180	M
			100	204	ž

Podijeliti ove studente po spolu i konstruisati tabelu klasa relativnih frekvencija za svaku grupu. Nacrtati, na istom grafiku, odvojeno, poligon klasa relativnih frekvencija za muške i ženske studente. Može li se ikakav zaključak izvući o relaciji između spola i nivoa holesterola,

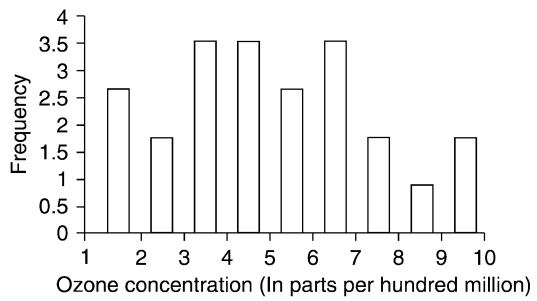
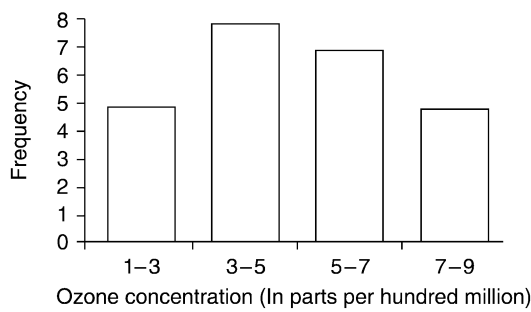
Odysseus:

10



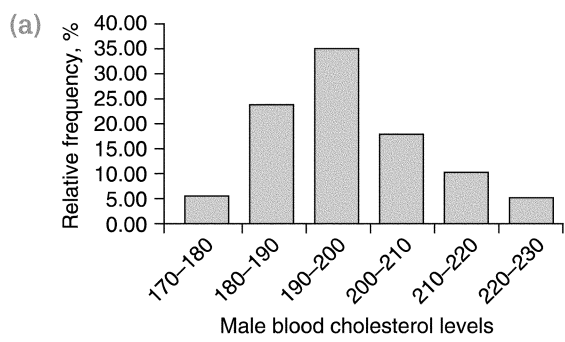
Class intervals 100–110 and 110–120
No
No

30

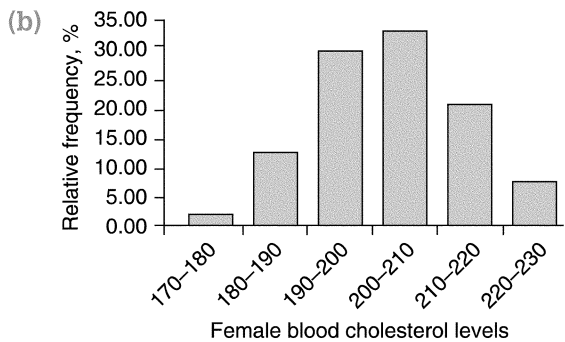


The chart in part (a) seems more informative since it shows a clearer pattern.

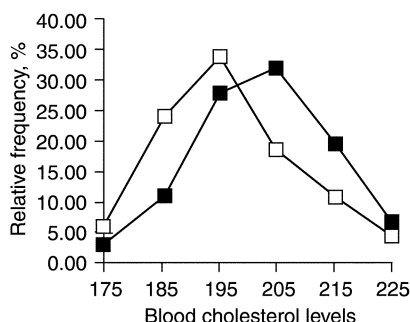
40



Female cholesterol	Frequency	Relative frequency
170–180	1	$1/46 = 0.02$
180–190	5	$5/46 = 0.11$
190–200	13	$13/46 = 0.28$
200–210	15	$15/46 = 0.33$
210–220	9	$9/46 = 0.20$
220–230	3	$3/46 = 0.07$



Male cholesterol	Frequency	Relative frequency
170–180	3	$3/54 = 0.06$
180–190	13	$13/54 = 0.24$
190–200	19	$19/54 = 0.35$
200–210	10	$10/54 = 0.19$
210–220	6	$6/54 = 0.11$
220–230	3	$3/54 = 0.06$



Female Male

Female students appear to have higher cholesterol levels.

Prikaz pomoću stabljika-i-listova

Veoma efikasan način da prikazemo malu-do-umjerenu veličinu podataka je da koristimo prikaz pomoću stabljike-i-listova. Takav prikaz dobijemo dijeleći svaku od vrijednosti podatka na dva dijela - na njegovu stabljiku i njegov list. Na primjer, ako je podatak dvocifren broj, tada možemo pustiti da je stabljika podatka vrijednost desetice a list vrijednost jedinice. To jest, vrijednost 84 možemo izraziti kao

stabljika		list
8		4

a podaci sa dvije vrijednosti 84 i 87 mogu izraziti kao

stabljika		list
8		4; 7

Data je tabela koja predstavlja prihod po glavi stanovnika 48 država Evrope.

Tabela 1- Prihod po glavi stanovnika (KM-ova po osobi)

Ime države		Ime države		Ime države	
Albanija	30 941	Farski otoci	25 057	Litvanija	23 941
Andora	25 128	Finska	33 404	Luksemburg	36 043
Austrija	32 151	Francuska	28 240	Makedonije	27 711
Bjelorusija	26 183	Njemačka	28 280	Malta	26 382
Belgija	23 512	Gibraltar	29 141	Moldavija	29 405
Bos. i Herc.	32 896	Grčka	25 578	Monako	25 575
Bugarska	33 278	Mađarska	25 446	Crna Gora	28 731
Šved. ostrva	42 706	Irska	27 744	Holandija	31 727
Hrvatska	32 778	Manx ostrvo	36 298	Norveška	31 318
Kipar	42 120	Italija	39 244	Poljska	25 400
Češka rep.	29 596	Kosovo	30 286	Portugal	26 894
Danska	28 821	Latvia	34 071	Rumunija	27 671
Estonija	30 001	Liktenštajn	22 372	San Marino	28 551
Srbija	28 836	Slovačka	30 180	Slovenija	24 306
Španija	25 020	Svalbard	34 334	Švedska	29 567
Švicarska	29 771	Ukrajina	39 453	Velika Britanija	32 922

Nacrtati prikaz pomoću stabljika-i-listova.

Rj. Podaci dati u tabeli se mogu predstaviti pomoću sledećeg stabljika-i-list prikaza. Primjetimo da su vrijednosti za listove postavljeni u prikazu u rastućem poretku

22	372
23	512; 841
24	706
25	020; 057; 128; 400; 446; 575; 579
26	183; 894; 982
27	671; 711; 744
28	240; 280; 551; 731; 821; 936
29	141; 405; 567; 596; 771;
30	001; 180; 296
31	319; 727
32	151; 779; 922; 996
33	276; 404
34	071; 334
36	043; 298
39	244; 453
42	120; 706

Izbor za stabljike uvijek treba biti napravljen na takav način da rezultirajući stabljika-i-list prikaz sadrži što više informacija o podacima (npr. vidi sljedeći primjer).

⑧ Sljedeći podaci predstavljaju postotak učenika koji pohađaju javnu osnovnu školu i koji su klasificirani kao manjina u 18 gradova BiH.

55,2; 47,8; 44,6; 64,2; 61,4; 36,6;
 28,2; 57,4; 41,3; 44,6; 55,2; 39,6;
 40,9; 52,2; 63,3; 34,5; 30,8; 45,3

Nacrtati stabljika-i-list prikaz podataka.

lj.

Ako pustimo da stabljika označava cifru desetice a da list predstavlja ostatak vrijednosti, tada stabljika-i-list prikaz za dati skup podataka je sljedeći

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

Mogli smo uzeti da stabljika predstavlja cijeli broj, a list decimalni dio vrijednosti, tako da bi vrijednost 28,2 bila prikazana kao

28 | ,2

Bezobzira, ovo bi kao rezultat dalo do prevrte stabljika (su prenatlo listova) i time podaci uopšte nebi bili jasno prikazani.

Slijedeći stabljika-i-list prikaz predstavlja težine 80 učenika sportskih darova. Stabljika predstavlja brojke desetice, a listovi cifru jedinice.

10	2; 3; 3; 4; 7	(5)
11	0; 1; 2; 2; 3; 6; 9	(7)
12	1; 2; 4; 4; 6; 6; 6; 7; 9	(9)
13	1; 2; 2; 5; 5; 6; 6; 8; 9	(9)
14	0; 4; 6; 7; 7; 9; 9	(7)
15	1; 1; 5; 6; 6; 6; 7	(7)
16	0; 1; 1; 1; 2; 4; 5; 6; 8; 8	(10)
17	1; 1; 3; 5; 6; 6; 6	(7)
18	1; 2; 2; 5; 5; 6; 6; 9	(8)
19	0; 0; 1; 2; 4; 5	(6)
20	9; 9	(2)
21	7	(1)
22	1	(1)
23		(0)
24	9	(1)

Objasnite kakve informacije možemo pročitati iz datih podataka.

Rj. Prvo primjetimo da brojevi u zagradama na desnoj strani predstavljaju broj vrijednosti u svakoj klasi list. Ovi brojevi koji predstavljaju zbir su često korisni. Na primjer, oni nam govore da postoji 10 vrijednosti koje imaju stabljiku 16; tj. 10 osoba imaju težinu između 160 i 169. Primjetimo da stabljika bez i jednog lista (kao npr. vrijednost 23) nam

govori da nema ničega u toj klasi.

Iz datog prikaza jasno je da su skoro sve vrijednosti podataka između 100 i 200, i da su skoro uniformno (ravnomjerno) raspoređeni u ovom regionu, sa izuzetkom na manje vrijednosti u intervalima između 100 i 110 i između 190 i 200.

Napomena: Ako se desi da jedna stabljika ima previše listova npr.

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

tada ovaj prikaz "razbijamo" u dvije linije

6 | 0, 0, 1, 2, 2, 3, 4, 4, 4, 4

6 | 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 9

Zadaci za vježbu

- ① Za sljedeće podatke nacrtati stabljika-i-list prikaz koji ima
- (a) 4 stabljike
 - (b) 8 stabljika

124; 129; 118; 135; 114; 139; 127; 141; 111; 144; 133; 127;
122; 119; 132; 137; 146; 122; 119; 115; 125; 132; 118; 126;
134; 147; 122; 119; 116; 125; 128; 130; 127; 135; 122; 141.

- ② Sljedeći podaci predstavljaju broj godina, tačnije najbližu godinu, od 43 pacijenta koji su primljeni u odjel za hitne slučajeve neke bolnice za odrasle.

23; 18; 31; 78; 44; 51; 24; 19; 17; 25; 27; 19; 44; 61; 22;
18; 14; 17; 29; 31; 22; 17; 15; 40; 55; 16; 17; 13; 20; 32;
20; 45; 53; 27; 16; 13; 22; 20; 18; 30; 20; 33; 21

Nacrtati stabljiku i list prikaz za ove podatke. Koristiti ovaj nacrt da bi odredili 5-godišnji interval koji sadrži najveći broj tački.

- ③ Sljedeći podaci predstavljaju Zeničku dnevnu naplatu sa parking mjesta (jedinice su u 100 km) tokom 30 dana u 2013-oj.

83; 97; 108; 77; 58; 88; 65; 52; 104; 75; 80; 83; 74; 68; 94;
71; 78; 83; 80; 79; 84; 81; 68; 57; 59; 32; 75; 83; 100; 88

- (a) Predstaviti ove podatke pomoću stabljike-i-list prikaza

(b) Da li neki podatak izgleda "sumnjiv"? Zašto?

(4) Dat je sledeći stabljika-i-list prikaz nekih podataka

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

(a) Koliko mnogo vrijednosti podataka je u 40-tim?

(b) Koliki procenat vrijednosti je veći od 50?

(c) Koliki postotak vrijednosti imaju cifru jedinice jednak 1?

(5) Koristan način za poređenje dva skupa podataka je da stavimo njihov stabljika-i-list prikaz jedan pored drugog. Ono što slijedi predstavlja rezultate standardnog ispita učenika iz dvije različite škole. U obe škole 24 studenta su pristupili ispitu

Škola A		Škola B
Listovi	Stabljika	Listovi
0	5	3, 5, 7
8, 5	6	2, 5, 8, 9, 9
9, 7, 4, 2, 0	7	3, 6, 7, 8, 8, 9
9, 8, 8, 7, 7, 6, 5, 3	8	0, 2, 3, 5, 6, 6
8, 8, 6, 6, 5, 5, 3, 0	9	0, 1, 5
	10	0

(a) Koja škola je imala "veće rezultate"?

(b) Koja škola je imala "niže rezultate"?

(c) Koju škola je uradila bolje na ispitu?

(d) Kombinujte ove dvije škole i nacrtajte stabljika-i-list prikaz za ^{suk 48 vrijednosti}

Odgovori:

10

(a)

11	1, 4, 5, 6, 8, 8, 9, 9, 9
12	2, 2, 2, 2, 4, 5, 5, 6, 7, 7, 7, 8, 9
13	0, 2, 2, 3, 4, 5, 5, 7, 9
14	1, 1, 4, 6, 7

(b)

11	1, 4
11	5, 6, 8, 8, 9, 9, 9
12	2, 2, 2, 2, 4
12	5, 5, 6, 7, 7, 7, 8, 9
13	0, 2, 2, 3, 4
13	5, 5, 7, 9
14	1, 1, 4
14	6, 7

20

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

The interval 15–20 contains 14 data points.
The interval 16–21 contains 17 data points.

30

(a)

3	2
4	
5	2, 7, 8, 9
6	5, 8, 8
7	1, 4, 5, 5, 7, 8, 9
8	0, 1, 3, 3, 3, 4, 8, 8
9	0, 3, 4, 7
10	0, 4, 8

40

- (a) 6
- (b) 43.75%
- (c) 12.5%

(b) Yes. The value 32 seems suspicious since it is so much smaller than the others.

50

(a) School B
(b) School A
(c) School A
(d)

5	0, 3
5	5, 7
6	2
6	5, 5, 8, 8, 9, 9
7	0, 2, 3, 4
7	6, 7, 7, 8, 8, 9, 9
8	0, 2, 3, 3
8	5, 5, 6, 6, 6, 7, 7, 8, 8, 9
9	0, 0, 1, 3
9	5, 5, 5, 6, 6, 8, 8
10	0

Skupovi podataka uređenih parova

Ponekad je skup podataka sačinjen od uređenih parova vrijednosti koje imaju neku relaciju jedna sa drugom. Tada za svaki član podatka podrazumijevamo da ima x i y vrijednost. Za i -ti par ovakve vrste podatka često koristimo oznaku (x_i, y_i) , $i=1, 2, \dots, n$. Na primjer, kao što ćemo vidjeti u prvom zadatku x_i može predstavljati veličinu inteligencije ostvarenu na IQ testu, a y_i godišnju veličinu plate. U ovoj lekciji ćemo vidjeti kako na efikasan način prikazati skup podataka koji čine parovi vrijednosti.

#) Sljedeća tabela prikazuje godišnju platu 30 radnika neke kompanije, kao i veličine njihovih inteligencija ostvarenih na IQ testu.

Tabela 1- Veličine plata i IQ

Radnik i	IQ rezultat x_i	Godišnje plata y_i (u 1000 km)
1	110	
2	107	68
3	83	30
4	87	13
5	117	24
6	104	40
7	110	22
8	118	25
9	116	62
10	94	45
11	93	70
12	101	15
13	93	22
14	76	18
15	91	20
		14

nastavak tabele

i	x_i	y_i
16	84	19
17	83	16
18	112	52
19	80	11
20	91	13
21	113	29
22	124	71
23	79	19
24	116	43
25	113	44
26	94	17
27	95	15
28	104	30
29	115	63
30	90	16

- Nacrtati stabljika-i-list prikaz za IQ rezultate kao i za godišnju prosječnu platu;
- Nacrtati dijagram disperzije (raspršenosti)
- Šta možemo zaključiti iz dijagrama disperzije (raspršenosti)
- Da li možemo ^{na osnovu dobijenih podataka} predviđati koliku će platu imati radnik sa IQ rezultatom od 120.
- Koji par vrijednosti bi nazvali uljezom?

Rj.
(a) Stabljika-i-list prikaz za IQ rezultate

12	4	(1)
11	0, 0, 2, 3, 3, 5, 6, 6, 7, 8	(10)
10	1, 4, 4, 7	(4)
9	0, 1, 1, 3, 3, 4, 4, 5	(8)
8	0, 3, 3, 4, 7	(5)
7	6, 9	(2)

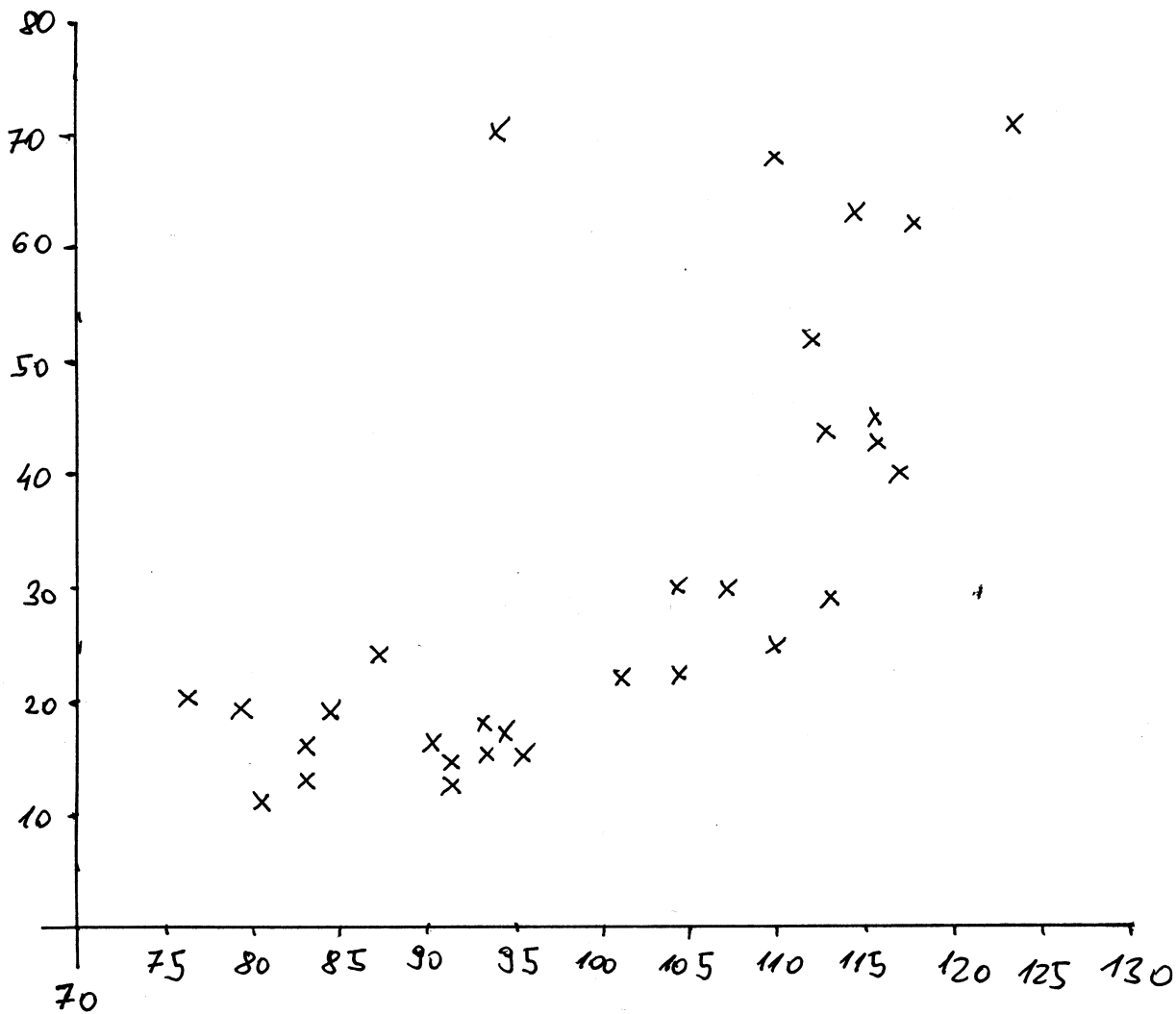
Stabljika-i-list prikaz za godišnje plate (u 1000 km)

7	0, 1	(2)
6	2, 3, 8	(3)
5	2	(1)
4	0, 3, 4, 5	(4)
3	0, 0	(2)
2	0, 2, 2, 4, 5, 9	(6)
1	1, 3, 3, 4, 5, 5, 6, 6, 7, 8, 9, 9	(12)

(b) Iako nam stabljika-i-list prikaz dobijen pod (a) govori mnogo o pojedinačnom IQ rezultatu i plati radnika, oni nam ne govore ništa o relaciji između ove dvije varijable. Time, na primjer, oni sami nisu nimalo korisni u određivanju da li veći IQ rezultat teži i većim primanjima u ovoj kompaniji. Da naučimo na koji način su podaci povezani sa ovim prikazom, potrebno je posmatrati uređenu vrijednost svake bačke podatka istovremeno.

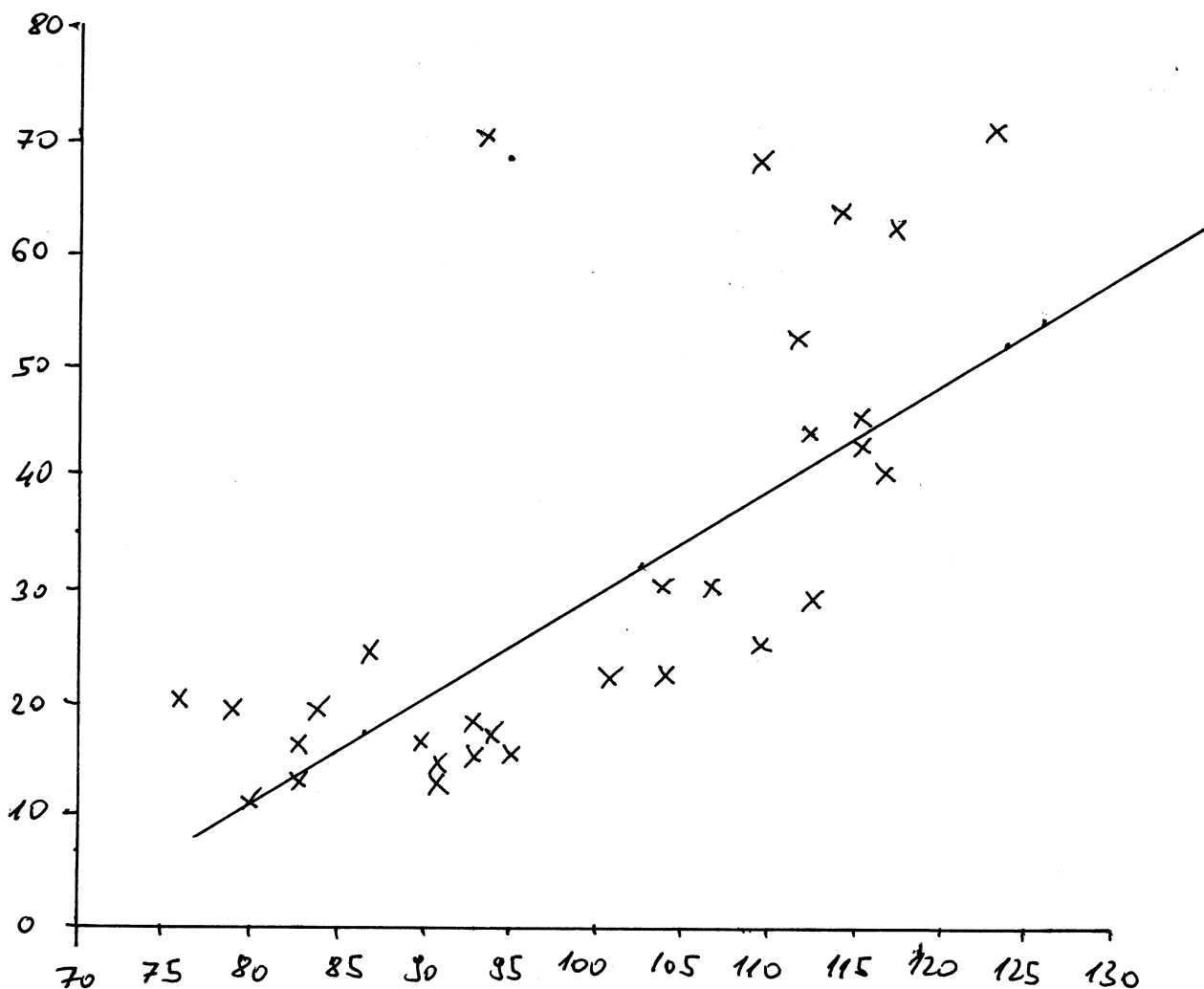
Koristan način za portriranje skupa podataka uređenih vrijednosti je da prikazemo podatke u dvodimenzionalnom pravougaonom

nacrta u kojem x-osa predstavlja vrijednost x podatka a y osa predstavlja vrijednost y podatka. Takav nacrt nazivamo dijagram disperzije. Sljedeća slika predstavlja dijagram disperzije za podatke iz postavke zadatka.



Dijagram disperzije za IQ protiv prihoda

- c) Iz ove slike jasno je da veći prihodi obično zajedno sa većim rezultatima ostvarenim na IQ testu.
- d) Dijagram disperzije prikazan iznad izgleda da također ima neke korisne osobine. Na primjer, pretpostavimo da želimo predviđeti platu radnika čiji je IQ test 120. Jedan način da ovo uradimo je da "odmjerimo od oka" pravu kroz dati skup podataka, kao što je urađeno na sljedećoj slici.



Kako vrijednost y na liniji odgovara vrijednosti x imamo da za 120 odgovara 45, ovo izlazi kao realna procjena za podizanje plate radnika čiji je IQ jednak 120.

e) Dijagram disperzije može biti koristan za uočavanje uljeza, koji su tačke podataka koje ne prate pravila pozicije ostalih tački (npr. tačka sa koordinatama (84; 70)). Kada uočimo uljeze, tada možemo odlučiti da li parovi podataka imaju značenje ili su uzrokovani greškom u podacima.

Zadaci za vježbu

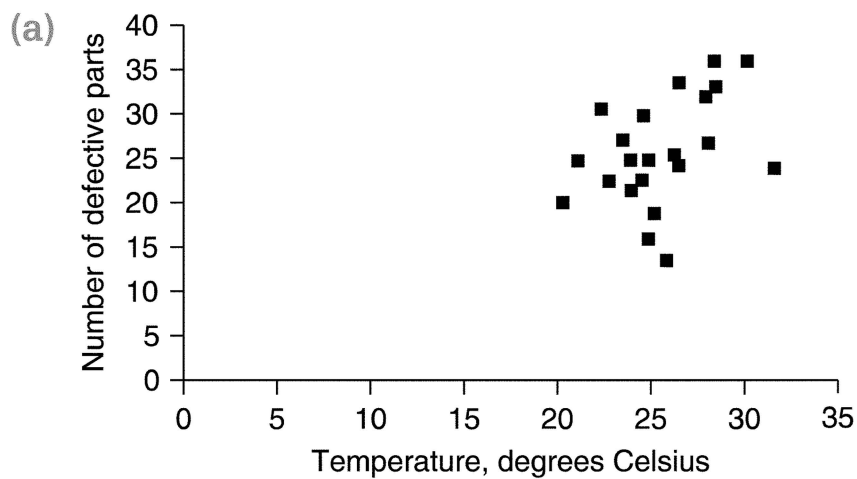
1) U pokušaju da odredi povezanost između dnevne podnevne temperature (mjerene u stepenima Celzijusa) i broja neispravnih dijelova proizvedenih tokom dana, neka kompanija je zabilježila sljedeće podatke nad 22 dana.

Temperatura	Broj neispravnih dijelova	Temperatura	Broj neispravnih dijelova
24,2	25	24,8	23
22,7	31	20,6	20
30,5	36	25,1	25
28,6	33	21,4	25
25,5	19	23,7	23
32,0	24	23,9	27
28,6	27	25,2	30
26,5	25	27,4	33
25,1	16	28,3	32
26,0	14	28,8	35
24,4	22	26,6	24

- (a) Nacrtati dijagram disperzije.
- (b) Šta možete zaključiti na osnovu dijagrama disperzije?
- (c) Ako sutrašnja podnevna temperatura bude 24,0; koja je vaša najbolja procjena da predvidite broj neispravnih dijelova koji se proizvedu.

Odgovori:

13



- (b) The number of defective parts tends to increase as the temperature increases.
- (c) About 23 or 24

Ključni pojmovi iz ove lekcije

Frekvencija: Broj puta koji se data vrijednost pojavi u skupu podataka.

Tabela frekvencija: Tabela koja predstavlja, za dati skup podataka, svaku različitu vrijednost kao i njezinu frekvenciju.

Linijski graf: Graf tabele frekvencija. Apcisa predstavlja vrijednost podatka, a frekvencija pojavljivanja te vrijednosti je prikazana visinom vertikalne linije.

Bar dijagram (ili bar grafikon): Slično linijskom grafu, osim što je sad vrijednost frekvencije podatka prikazano visinom prede (pravougaonika).

Polinom frekvencija: Nacrt različitih vrijednosti podataka i njihovih frekvencija koje spajaju nacrtane točke pravim linijama.

Simetrični skup podataka: Skup podataka je simetričan oko vrijednosti x_0 ako su frekvencije vrijednosti podatka $x_0 - c$ i $x_0 + c$ iste za sve vrijednosti c .

Relativna frekvencija: Frekvencija vrijednosti podataka podijeljena sa brojem dijelova podataka u skupu.

Zvrk grafikon: Dijagram koji prikazuje relativne frekvencije podjelom kruga na različite sektore.

Histogram: Grafikon u kome su podaci podijeljeni na klase intervala, čije su frekvencije prikazane u bar grafikonu.

Relativni histogram frekvencija: Histogram koji prikazuje relativne frekvencije za svaku vrijednost podatka u skupu.

Stabljika-i-list prikaz: Slično histogramu osim što su frekvencije prikazane povezujući zajedno zeduje brojke (listove) podataka.

Dijagram disperzije (raspršenosti): Dvo-dimenzionalni prikaz skupa podataka kao uređene vrijednosti.