



Deskriptivna statistička analiza

Predavač: Dr Mirko Savić

savicmirko@ef.uns.ac.rs

www.ef.uns.ac.rs

Deskriptivna statistička analiza predstavlja skup metoda kojima se vrši izračunavanje, prikazivanje i opisivanje osnovnih karakteristika statističkih serija.

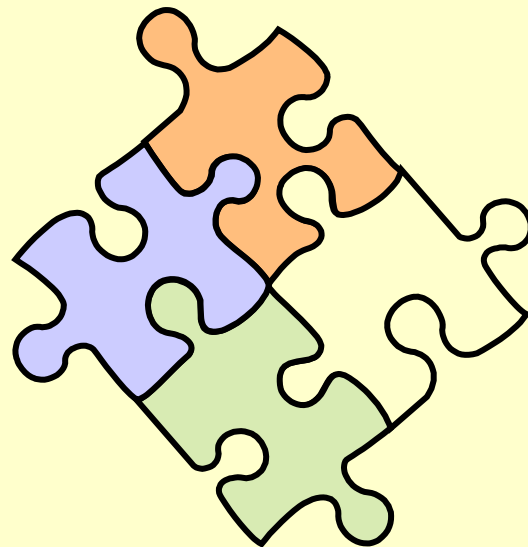
Deskriptivna statistička analiza ima sledeće zadatke:

1. Grupisanje i sređivanje statističkih podataka.
2. Prikazivanje statističkih podataka.
3. Određivanje osnovnih pokazatelja statističkih serija.

Grupisanje i sređivanje statističkih podataka

Grupisanje podataka se vrši prema vrednostima ili modalitetima posmatranih obeležja.

Kao krajnji rezultat grupisanja javlja se statistička serija.



Statistička serija predstavlja uređeni skup varijacija obeležja posmatrane statističke mase.

Statistička serija se prikazuje u obliku tabele.

Primer za negrupisane podatke (prekidna obeležja)

Raspored 45 studenata prema broju položenih ispita:

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

Primer za grupisane podatke, neintervalna serija

Tabela broj XX: Raspored studenata prema broju položenih ispita:

Broj ispita (x_i)	Broj studenata (f_i)
1	2
0	3
1	4
2	3
3	10
4	5
5	5
6	8
7	5
8	2
Ukupno	45

Primer za grupisane podatke, intervalna serija

Tabela broj 4: Raspored studenata prema broju položenih ispita

Broj ispita (X)	Broj studenata (f_i)
1	2
0-2	10
3-5	20
6-8	15
Ukupno	45

Obeležje (X) –
kvalitativna strana

Frekvencija (f_i) –
kvantitativna strana

Primer za negrupisane podatke (neprekidna obeležja)

Raspored cena benzina na 10 benzinskih pumpi u gradu:
95,2 98,0 95,4 96,0 99,1 97,5 99,8 99,8 95,1
100,3.

Primer za grupisane podatke u intervalnu seriju

Tabela broj XX: Raspored cena benzina:

Cena u dinarima (x_i)	Broj pumpi (f_i)
1	2
95,01-97	4
97,01-99	2
99,01-101	4
Ukupno	10

Određivanje donje granice prvog intervala

Prikazivanje statističkih podataka

Prikazivanje statističkih serija se može vršiti na dva načina:

- tabelarno,
- grafički.



Statistička tabela predstavlja uokvirenu površinu u koju se unosi statistička serija.

Broj i naziv tabele

Naziv obeležja (X)	Naziv frekvencije (f_i)	← Zaglavlje
1	2	← Redni broj kolone
Vrednost ili modalitet obeležja	f_1	
Vrednost ili modalitet obeležja	f_2	
·	·	
·	·	
·	·	
Vrednost ili modalitet obeležja	f_n	
Ukupno:	$\sum_{i=1}^n f_i$	← Zbirni red

Primedba:

Izvor:

↑
Predkolona

Kako studenti treba da izrađuju tabelu

Tabela 1.

Mesečna potrošnja u litrama (X)	Broj domaćinstava (f_i)	Grupna sredina (x_i)	$x_i f_i$	$x_i^2 f_i$
1	2	3	4	5
1,1-2	6	1,5		
2,1-3	20	2,5		
3,1-4	16	3,5		
4,1-5	14	4,5		
5,1-6	9	5,5		
6,1-7	5	6,5		
Ukupno:	70	-	260	1101,5

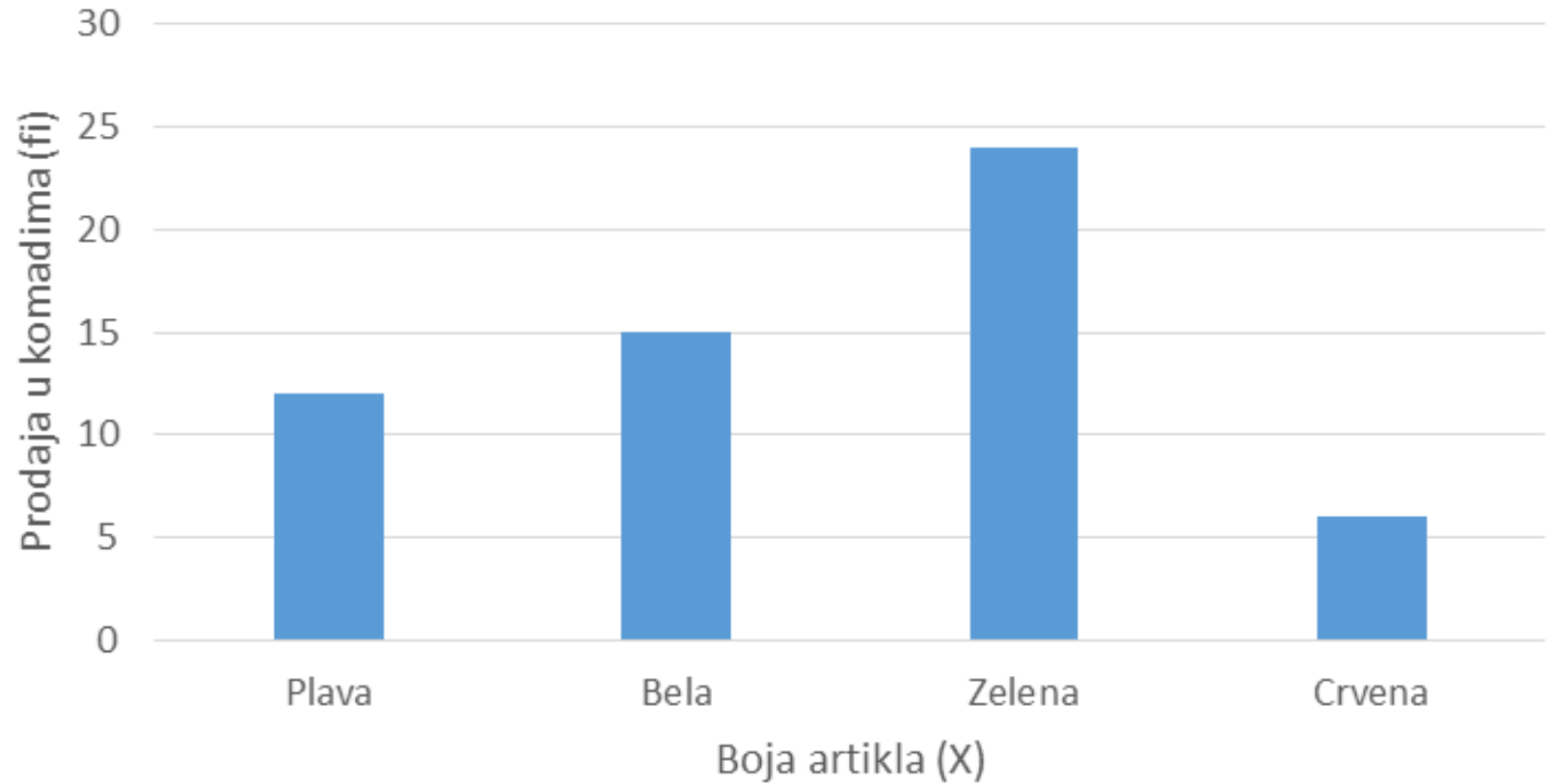
Grafičko prikazivanje

Grafikoni se najčešće dele na sledeći način:

- tačkasti (stigmogrami),
- površinski,
- prostorni,
- linijski,
- kartogrami.

Boja artikla	Prodaja u komadima
Plava	11
Bela	15
Zelena	24
Crvena	6

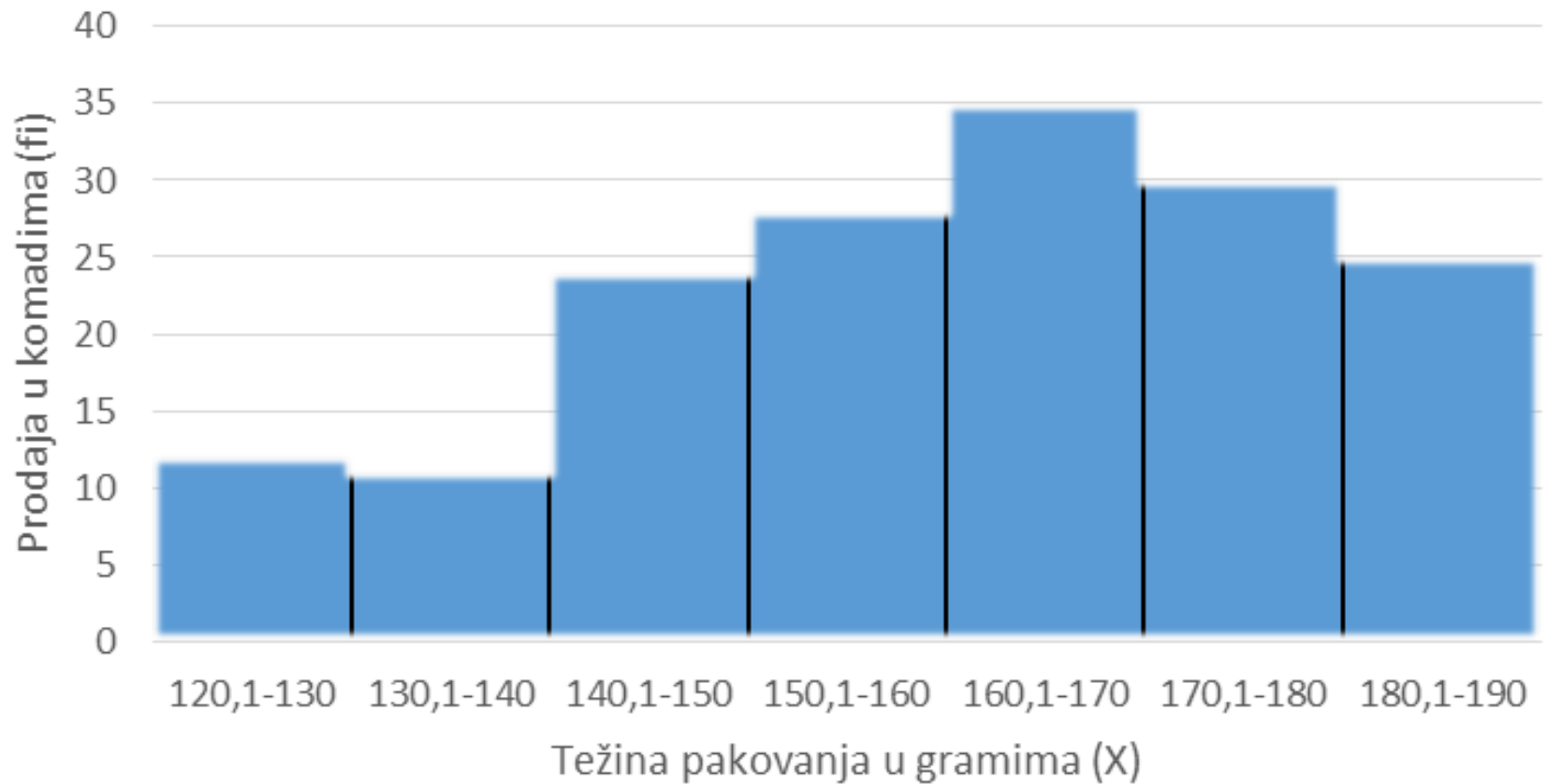
Prodaja artikla po bojama



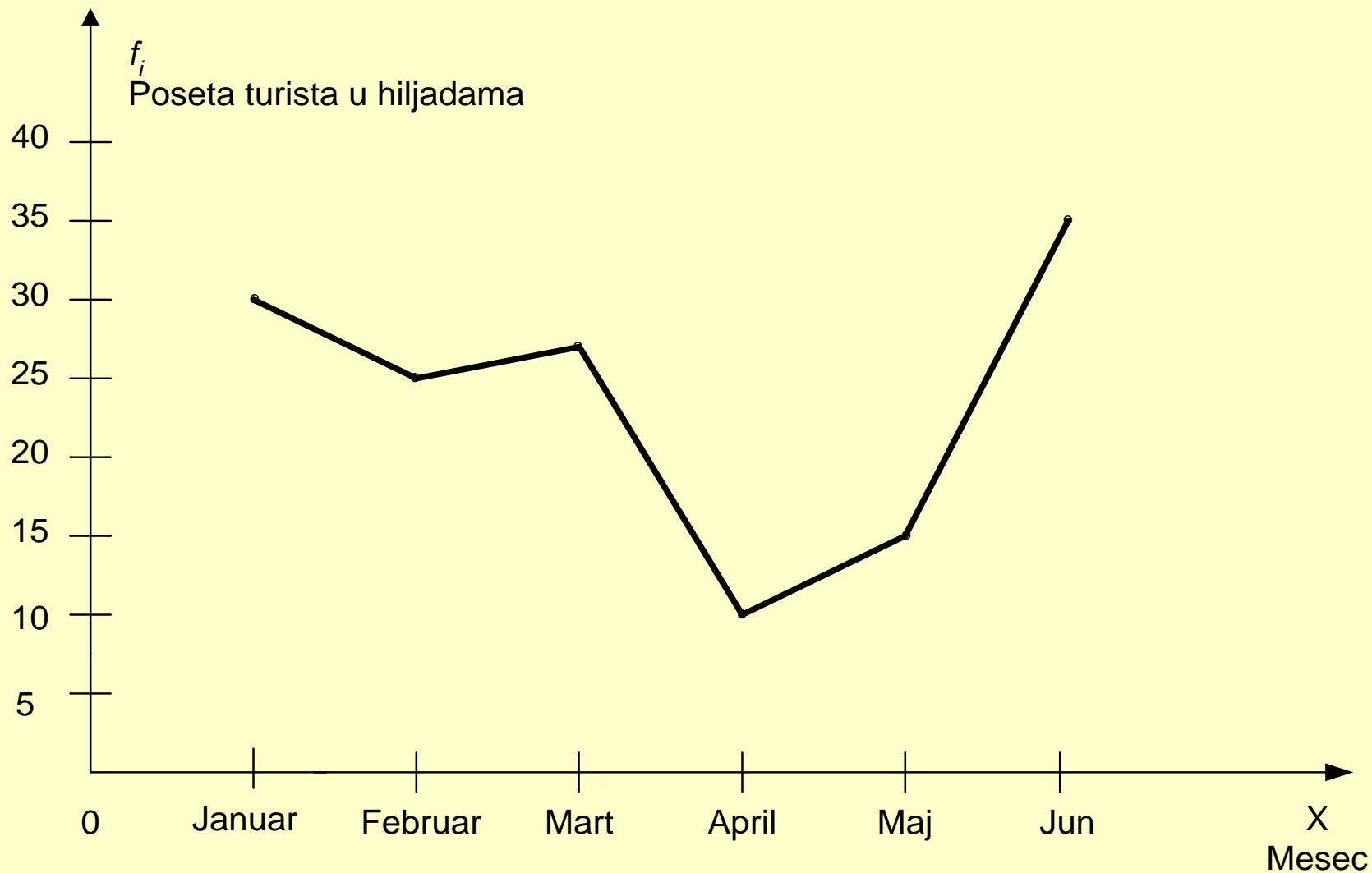
Primer za bar dijagram

Težina pakovanja u gramima	Prodaja u komadima
120,1-130	12
130,1-140	11
140,1-150	24
150,1-160	28
160,1-170	35
170,1-180	30
180,1-190	25

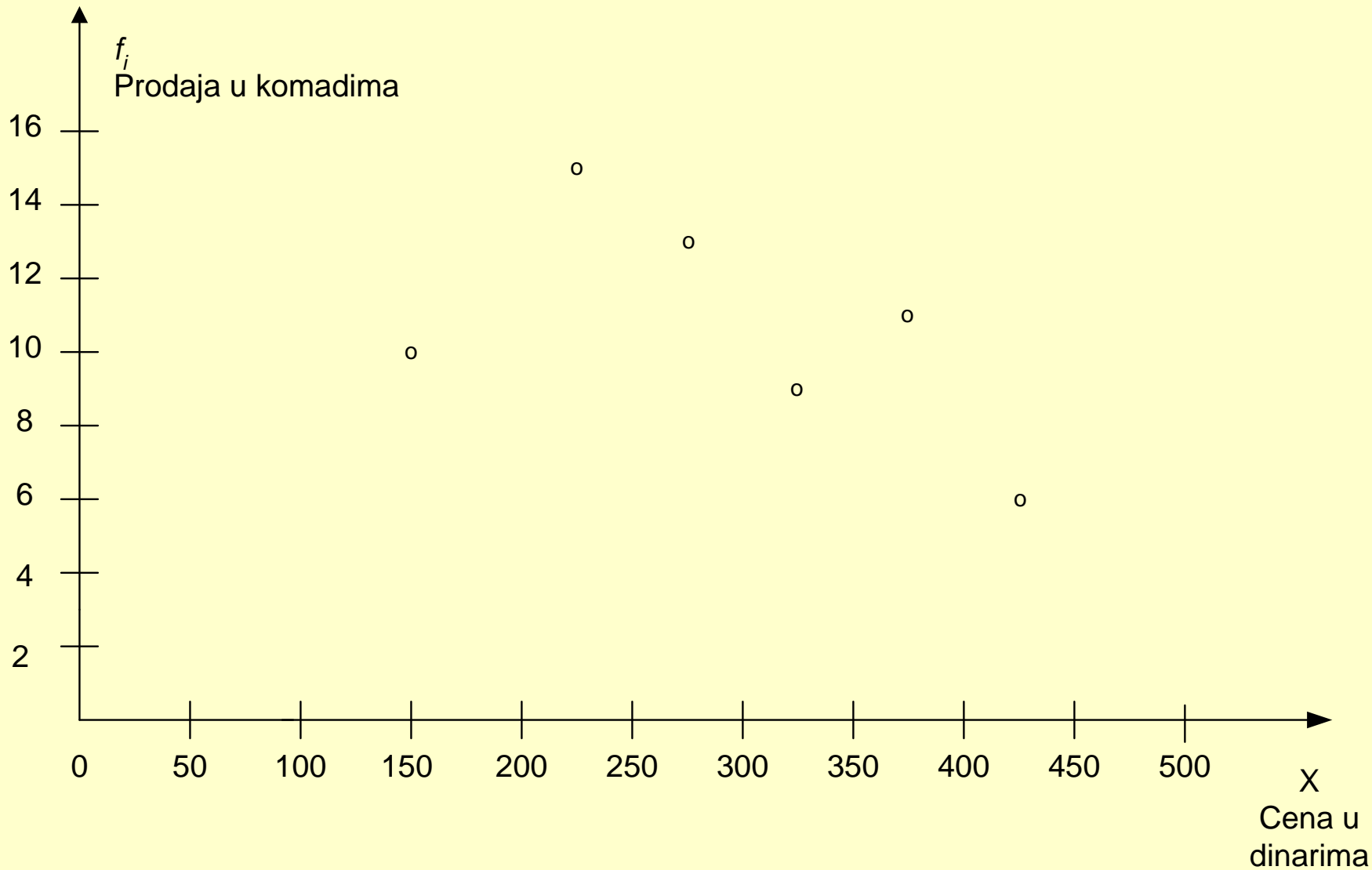
Prodaja prema težini pakovanja



Primer za histogram

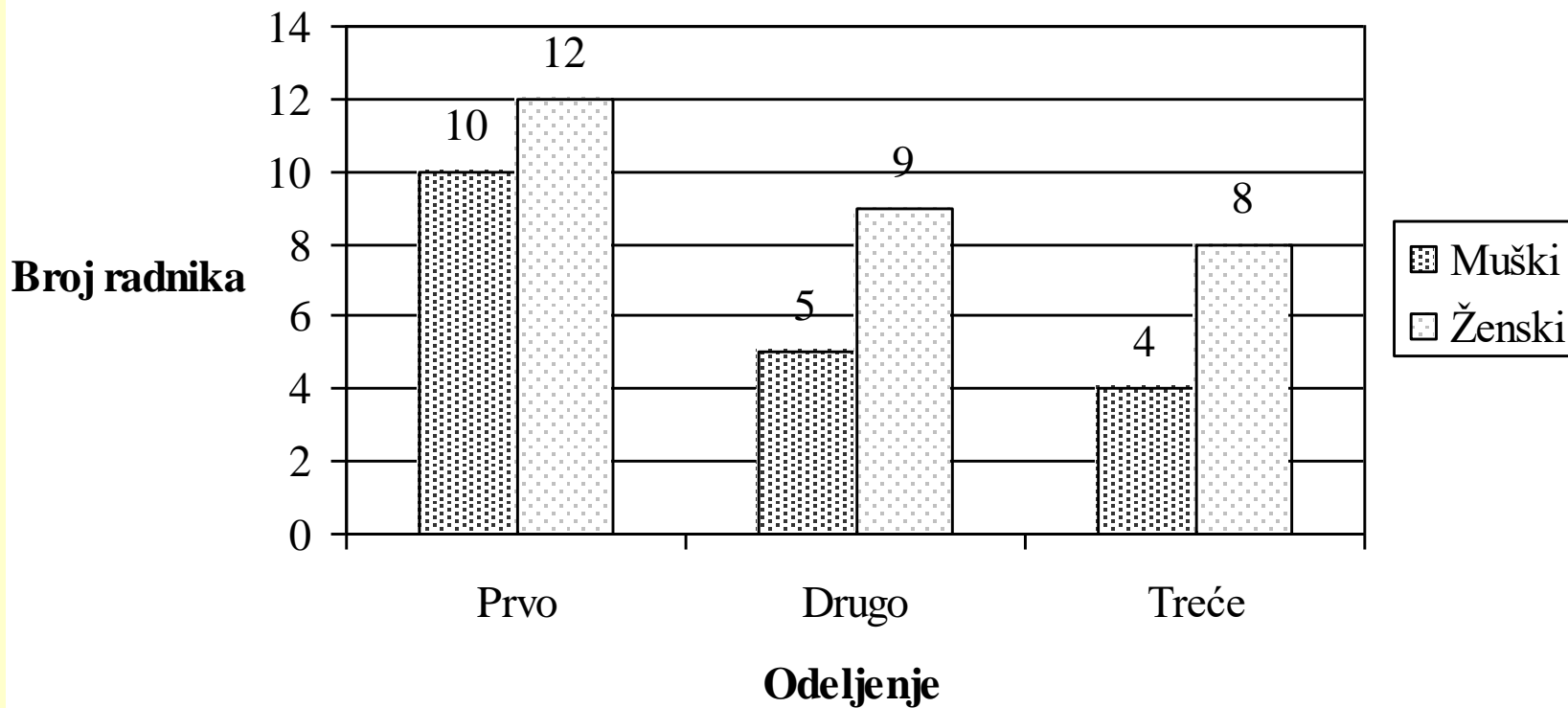


Primer za poligon frekvencija (linijski dijagram)



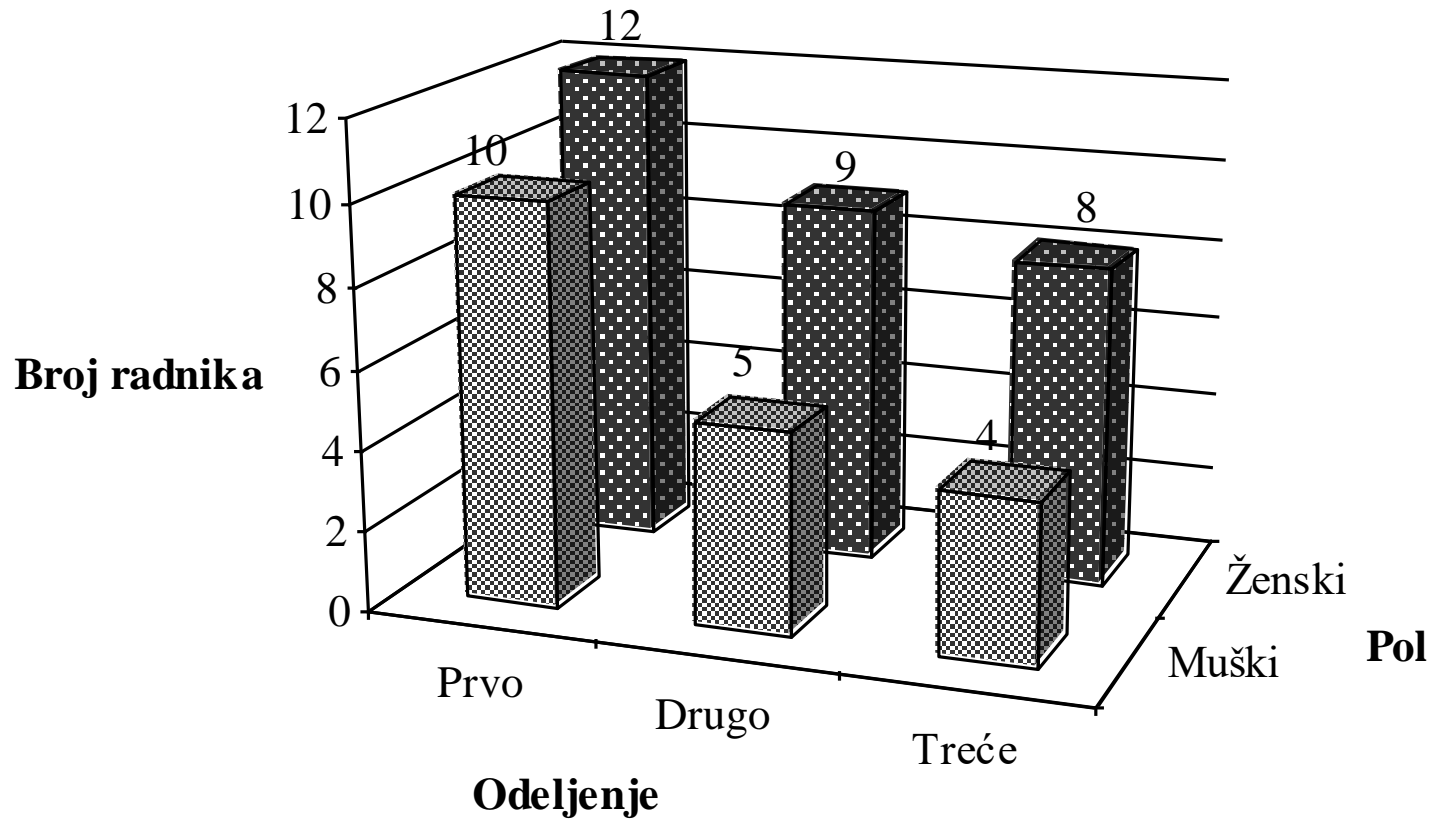
Primer za dijagram rasturanja

Broj radnika prema odeljenjima i polu



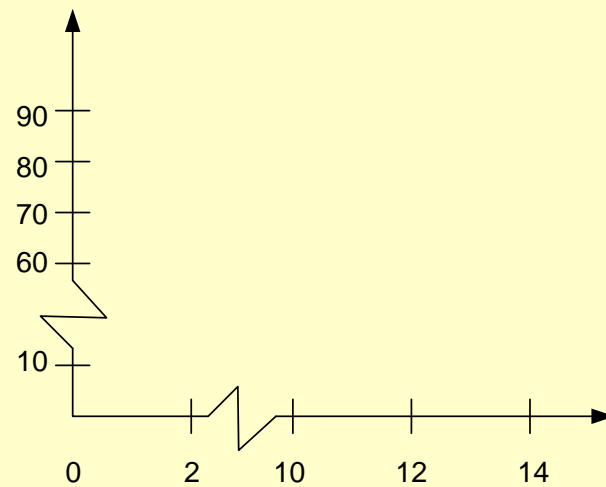
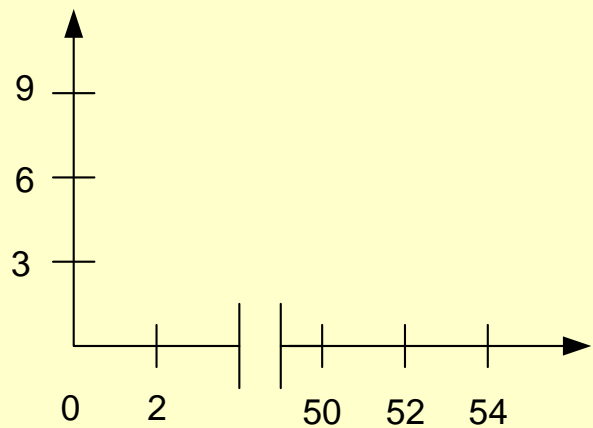
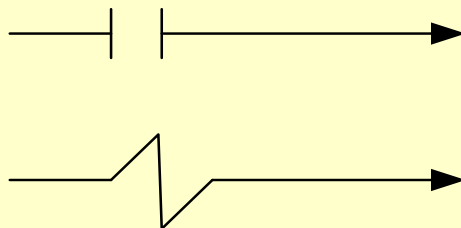
Primer za bar dijagram sa dva obeležja
(urađeno u Excel-u)

Broj radnika prema odeljenjima i polu

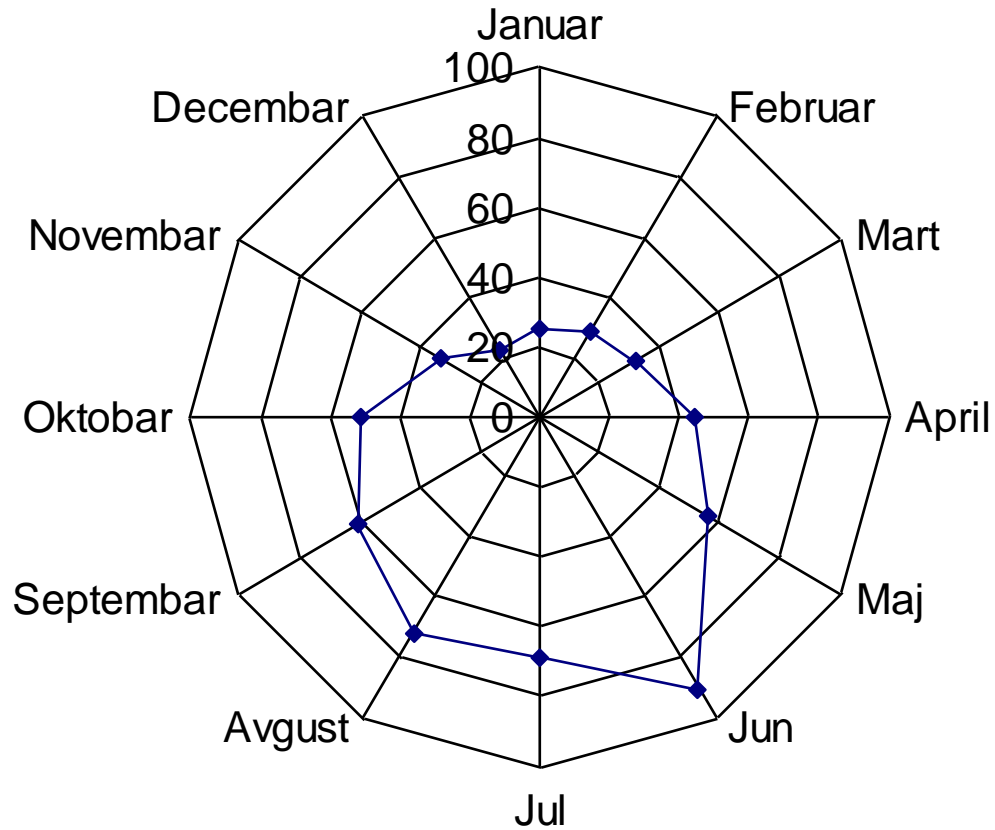


Primer za stereogram (urađeno u Excel-u)

Pravljenje preseka na osama



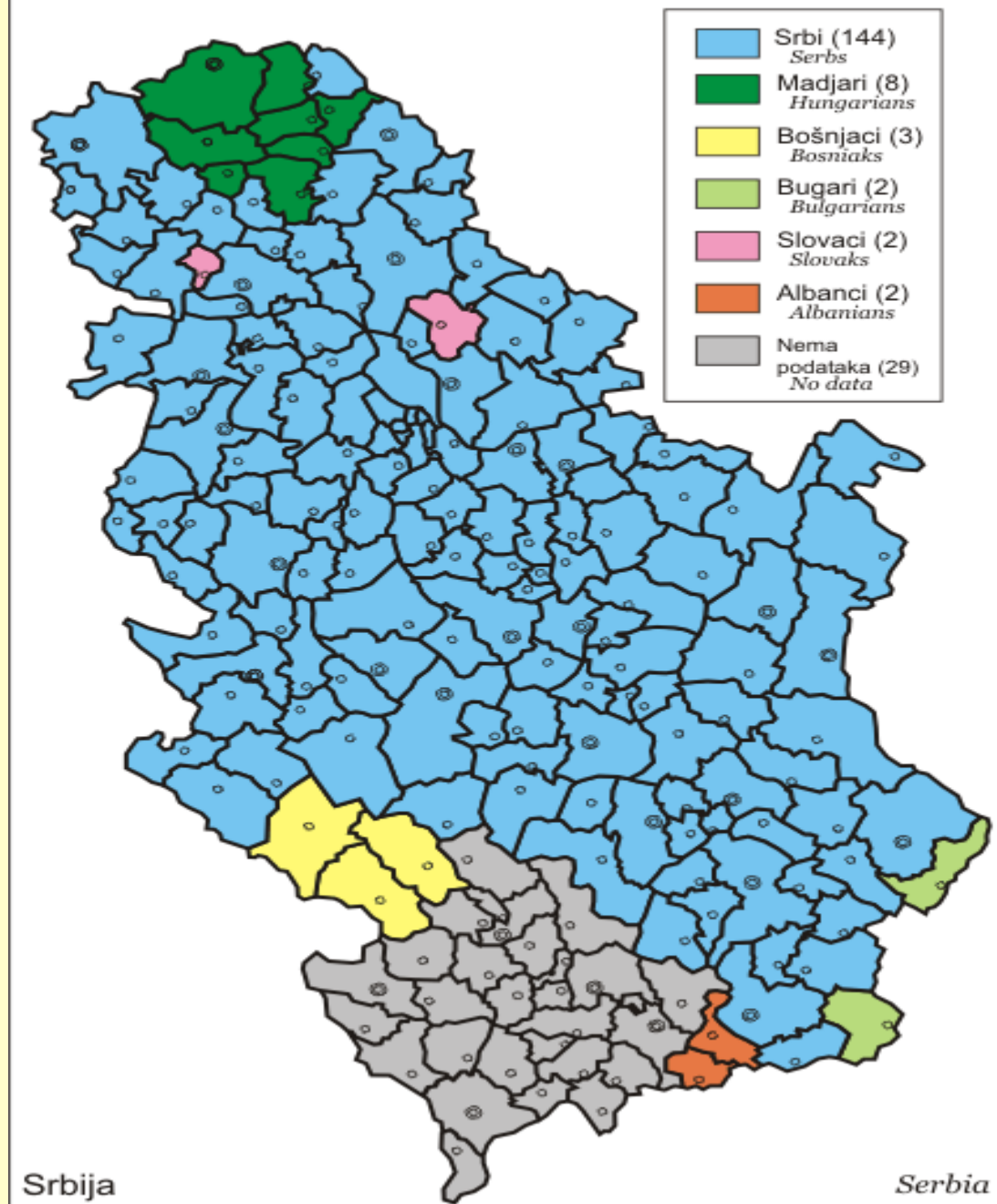
Poseta turista u hiljadama



Primer za polarni dijagram (linijski dijagram)

Konačni rezultati Popisa 2002.
Census 2002 - Final Results

Primer za
kartogram



DES-001 K(05)2-1 Grupisanje i grafičko prikazivanje statističkih podataka (prekidna numerička obeležja)

DES-002 K(05)2-2 Grupisanje i grafičko prikazivanje statističkih podataka (neprekidna num. obeležja)

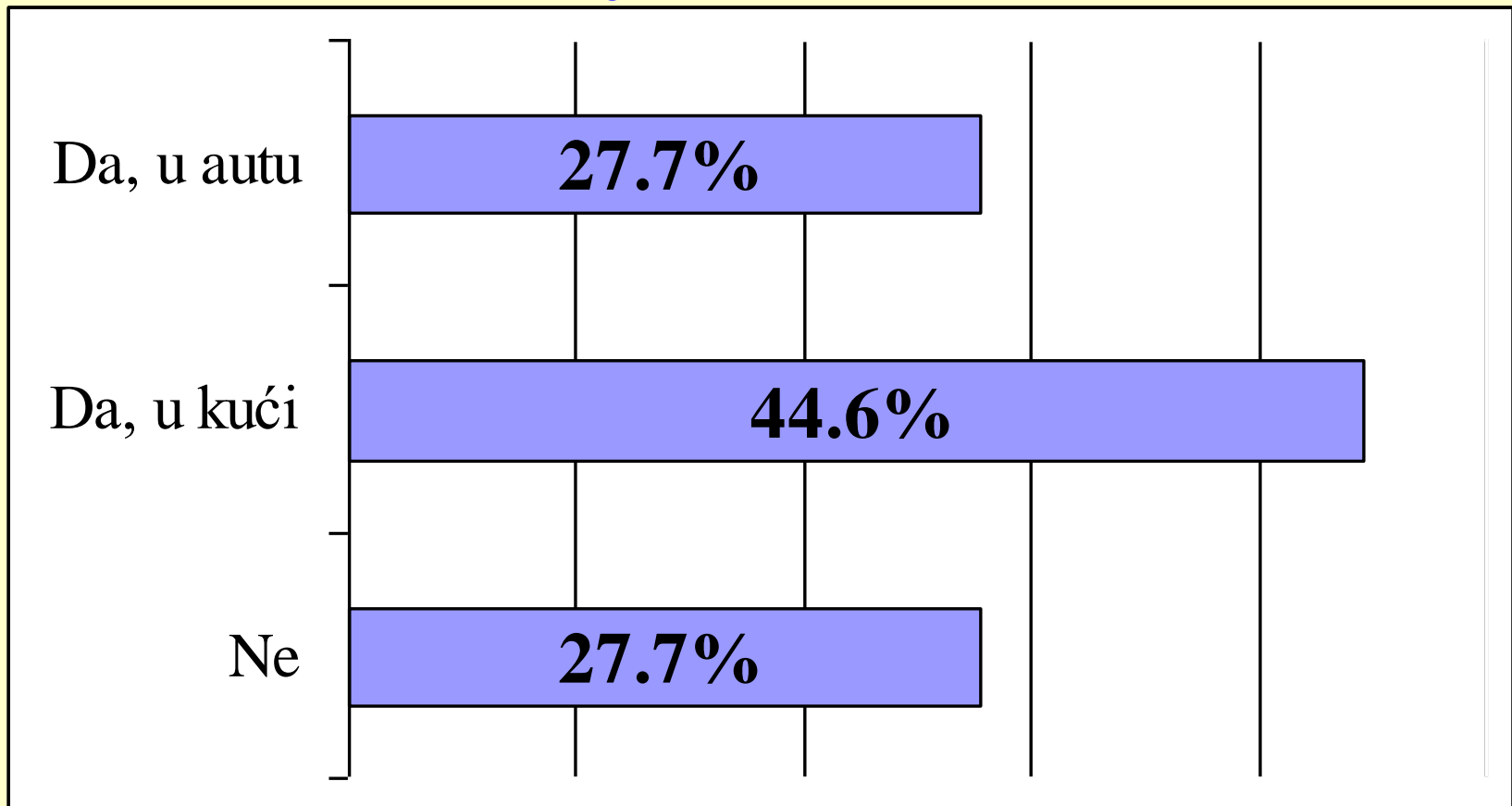
DES-097 Z(06)1-1 Grupisanje, prekidna obeležja

DES-098 Z(06)1-2 Grupisanje, neprekidna obeležja

DES-128 Poligon frekvencija i poligon kum. frekvencija

DES-131 Grafičko prikazivanje serija sa otvorenim intervalima

Primer za loše grupisanje i grafičko prikazivanje:
Imate li klima uređaj?



Izvor: Blic, 7. avgust 2006.

Primer za loše grafičko prikazivanje (Fresh&Co)

Grafičko prikazivanje u softveru



- Pravljenje linijskog dijagrama
- Dijagram rasturanja
- Pie Chart (površinski dijagrami)
- Bar dijagram
- Histogram
- Polarni dijagram
- Kreiranje grafa na osnovu pivot tabela

I Q_2 V

Osnovne mere statističkih serija σ

U osnovne mere statističkih serija spadaju (nema u udžbeniku):

- Srednje vrednosti (mere centralne tendencije).
- Mere varijacije (mere disperzije, raspršenosti).
- Mere oblika rasporeda.

 σ^2 μ M_0 α_4 Q_1 α_3 M_e Q_3

Srednje vrednosti

Srednje vrednosti su vrednosti obeležja koje na specifičan način reprezentuju čitavu statističku masu, odnosno zamenjuju sve vrednosti u statističkoj seriji i karakterišu statističku masu u celini.

Srednje vrednosti poseduju sledeće osobine:

- Ne mogu biti veće od najveće vrednosti obeležja niti manje od najmanje vrednosti obeležja u seriji.
- Mogu imati vrednost koja uopšte ne postoji u numeričkoj seriji.
- Mogu biti izražene i decimalnim brojem bez obzira da li je u pitanju serija sa prekidnim ili neprekidnim obeležjem.

Srednje vrednosti se dele u dve grupe:

- Izračunate srednje vrednosti.
- Srednje vrednosti po položaju.

Izračunate srednje vrednosti se mogu utvrditi samo računskim putem. Tu spadaju:

- aritmetička sredina* $(\bar{x}; \mu)$
- geometrijska sredina* (G) ,
- harmonijska sredina (H) ,
- kvadratna sredina,
- kubna sredina,
- logaritamska sredina.

Srednje vrednosti po položaju

To su srednje vrednosti koje se mogu odrediti na osnovu pozicije na kojoj se nalaze kada su vrednosti obeležja poredane u rastući niz.

Tu spadaju:

- modus* (M_o),
- medijana* (M_e),
- medijala (M_l),
- kvartili* (Q),
- kvintali (Kv),
- decili (D),
- percentili* (P).

Aritmetička sredina (prosek)

Simboli koji se koriste:

Aritmetička sredina za uzorak: \bar{x} ("iks-bar")

Aritmetička sredina za osnovni skup: μ ("mi")

Prema tome da li su podaci grupisani ili ne, razlikuju se:

- prosta aritmetička sredina,
- ponderisana (složena, vagana) aritmetička sredina.

Formule za aritmetičku sredinu:

Prosta, za osnovni skup:

$$\mu = \frac{\sum_{i=1}^N x_i}{N};$$

Prosta, za uzorak:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n};$$

Ponderisana, za osnovni skup:

$$\mu = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i};$$

Ponderisana, za uzorak:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i};$$

DES-132 Prosta aritmetička sredina za uzorak

DES-133 Složena aritmetička sredina za osnovni skup

Geometrijska sredina

Simbol: G

Geometrijska sredina spada u izračunate srednje vrednosti koja se koristi kada u numeričkoj seriji obeležja pokazuju neke relativne pokazatelje (indekse) ili karakteristike geometrijske progresije.

Formule za geometrijsku sredinu:

Prosta geometrijska sredina:

$$\log G = \frac{\sum_{i=1}^n \log x_i}{n};$$

Ponderisana geometrijska sredina:

$$\log G = \frac{\sum_{i=1}^k f_i \log x_i}{\sum_{i=1}^k f_i};$$

Geometrijska sredina:

$$G = \sqrt{\log G}$$

Geometrijsku sredinu nije moguće izračunati ako je neka vrednost obeležja jednaka nuli!

Primer za antilogaritam

$$\log G = 0,9542425$$

Antilogaritam:

$$G = 10^{\log G} = 10^{0,9542425} = 9$$

DES-134 Prosta geometrijska sredina

Modus

Simbol: M_o

Modus je ona vrednost obeležja koja se najčešće javlja u statističkoj seriji, odnosno ona vrednost obeležja koja ima najveću frekvenciju.

*Zašto je modus nekad bolji od aritmetičke sredine?
Veličina obuće ili odeće.*

DES-081 Modus, negrupisani podaci

DES-135 Modus za neintervalnu numeričku seriju

Medijana

Simbol: M_e

Medijana je srednja vrednost po položaju koja deli numeričku seriju na dva jednaka dela. Jedna polovina vrednosti obeležja je manja od nje, a druga polovina veća.

Formule za medijanu:

Neparan broj podataka: $M_e = x_{\frac{n+1}{2}};$

Paran broj podataka: $M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2};$

Medijana može da se odredi i grafički uz pomoć kumulacija ispod i iznad.

DES-137 – Medijana za negrupisane podatke – neparan broj podataka

Zašto je nekad bolje koristiti medijanu nego aritmetičku sredinu?

Primer:

“U našem preduzeću prosečna plata je 400 evra!”

$$\mu=400$$

**Preduzeće ima 6 radnika sa platama:
100, 100, 150, 150, 400, 1500**

$$M_e=150$$



Kvartili

Kvartili su srednje vrednosti po položaju koje dele statističku seriju na četiri jednaka dela kada su vrednosti obeležja poređane u rastući niz.

Postoji ukupno tri kvartila.

Simboli: Q_1 , Q_2 , Q_3



Prvi kvartil (Q_1) deli numeričku seriju tako da je jedna četvrtina podataka manja od njega a tri četvrtine su veće.

Drugi kvartil (Q_2) je jednak sa medijanom (Me) i deli numeričku seriju tako da je jedna polovina podataka manja od njega a druga polovina veća.

Treći kvartil (Q_3)...

Na primer, zarada radnika: $Q_1=30000$

Percentili

Percentili su srednje vrednosti po položaju koje dele statističku seriju na sto jednakih delova.

Simbol: P

Na primer, zarada radnika: $P_{80}=85000$

Primeri za srednje vrednosti:

DES-021 K:2-7

Aritmetička sredina, M_o i M_e , negrupisani podaci

DES-022 K:2-8

Aritmetička sredina, M_o i M_e , neintervalna serija

DES-063; K(05)z 2-11 Geometrijska sredina

DES-064; K(05)z 2-13 Geometrijska sredina

Mere varijacija (mere disperzije)

Mere varijacije su pokazatelji relativnih i apsolutnih odstupanja vrednosti obeležja od neke srednje vrednosti, obično od aritmetičke sredine.

U statističkoj praksi postoji velik broj mera varijacije:

- *interval varijacije I
- *varijansa σ^2
- *standardna devijacija σ
- *koeficijent varijacije V
- *normalizovano (standardizovano) odstupanje (z-skor) U
- *interkvartilna varijacija I_Q
- srednje apsolutno odstupanje D

Podaci o antropomerama građana SFRJ (16-55 godina starosti) (udžbenik, strana 52)

Muškarci

<i>Naziv obeležja</i>	<i>Aritmetička sredina</i>	<i>Standardna devijacija</i>
<i>Težina tela</i>	<i>72,8</i>	<i>10,51</i>
<i>Visina tela</i>	<i>174,64</i>	<i>6,89</i>
<i>Dužina nosa</i>	<i>5,01</i>	<i>0,55</i>
<i>Širina ramena</i>	<i>48,80</i>	<i>2,38</i>
<i>Širina kukova</i>	<i>39,78</i>	<i>2,55</i>
<i>Broj cipela</i>	<i>42,91</i>	<i>1,41</i>

Žene

<i>Naziv obeležja</i>	<i>Aritmetička sredina</i>	<i>Standardna devijacija</i>
<i>Težina tela</i>	<i>70,07</i>	<i>12,91</i>
<i>Visina tela</i>	<i>166,59</i>	<i>9,25</i>
<i>Dužina nosa</i>	<i>5,05</i>	<i>2,49</i>
<i>Širina ramena</i>	<i>40,78</i>	<i>2,20</i>
<i>Širina kukova</i>	<i>38,70</i>	<i>1,40</i>
<i>Broj cipela</i>	<i>37,68</i>	<i>1,06</i>

Primeri za mere varijacije:

DES-023 K:2-9

Mere varijacije, negrupisani podaci, uzorak

DES-069 Z(06)3-1

Srednje vrednosti, mere varijacije, neintervalna, uzorak

DES-028; K(05)z 2-9 Mere varijacije – negrupisani, prekidna

DES-032; K(05)z 2-10 Srednje vred., Mere v. - grup. i prekidna

DES-068; K(05)z 2-12 Srednje v. i mere v.-grupis., prekidna

Mere oblika rasporeda

Za izračunavanje asimetrije i spljoštenosti rasporeda koriste se sledeće mere:

- mera asimetrije (α_3),
- mera spljoštenosti (ekscesa) (α_4).

Mere oblika rasporeda se izračunavaju preko pomoćnih i centralnih momenata rasporeda.

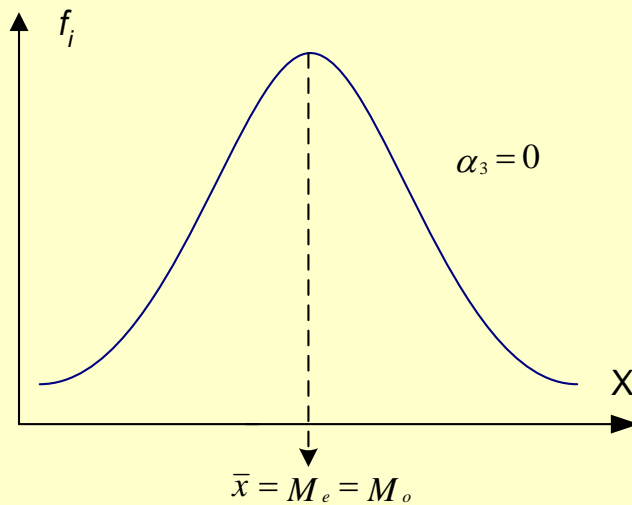
Koeficijent asimetrije

Simbol: α_3

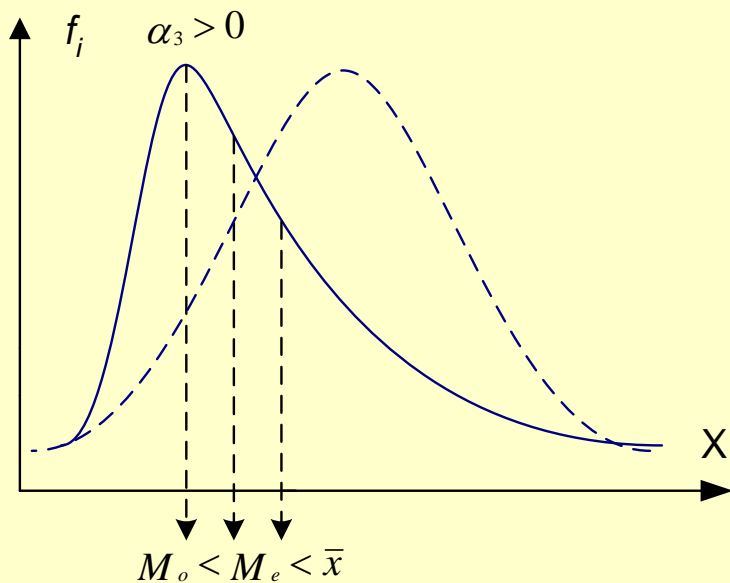
Numerički pokazatelj koji izražava u kojoj meri je neki raspored asimetričan u odnosu na normalni raspored.

Ako je:

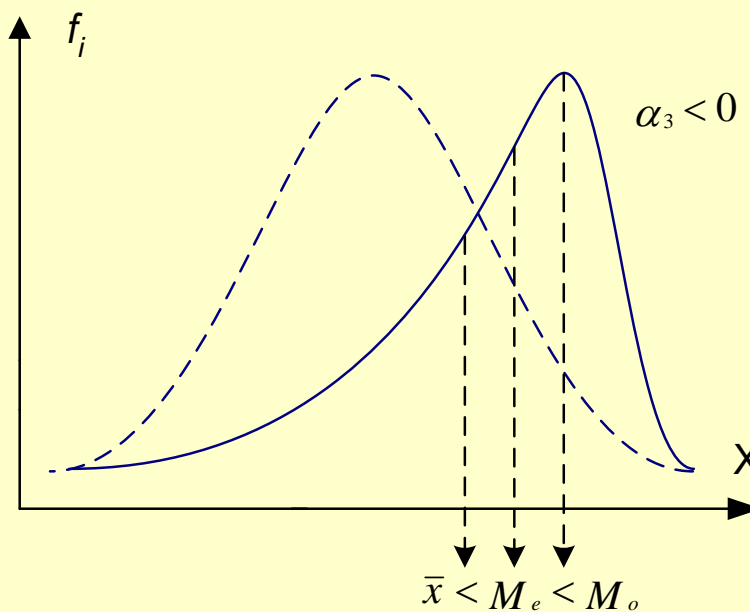
- $\alpha_3 = 0$, raspored je simetričan,
- $\alpha_3 > 0$, raspored je asimetričan u desno (pozitivna asimetrija),
- $\alpha_3 < 0$, raspored je asimetričan u levo (negativna asimetrija).



Simetrični raspored



Raspored asimetričan u desnu stranu
(pozitivna asimetrija)



Raspored asimetričan u levu stranu
(negativna asimetrija)

U zavisnosti od veličine koeficijenta, određuje se i jačina asimetrije. Gradacija je sledeća:

$|\alpha_3| \leq 0,25$ – mala asimetrija,

$0,25 < |\alpha_3| \leq 0,50$ – srednja asimetrija,

$|\alpha_3| > 0,50$ – jaka asimetrija.

Formula za koeficijent asimetrije:

$$\alpha_3 = \frac{M_3}{\sigma_u^3}$$

Koeficijent spljoštenosti

Simbol: α_4

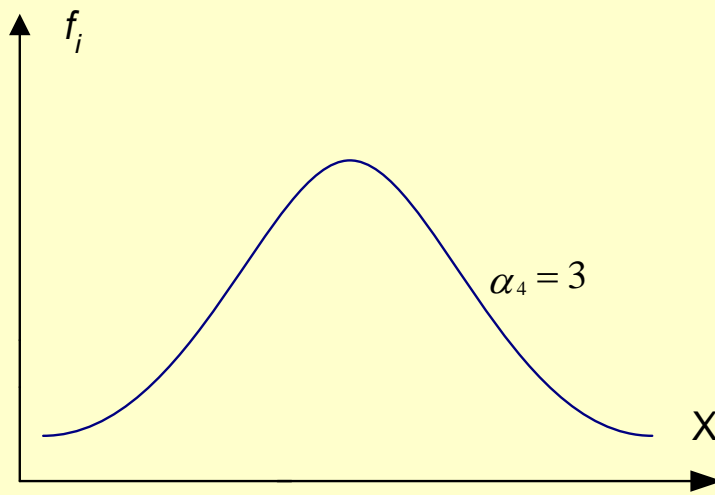
Numerički pokazatelj koji izražava u kojoj meri je neki raspored spljošten u odnosu na normalni raspored.

Formula za koeficijent spljoštenosti:

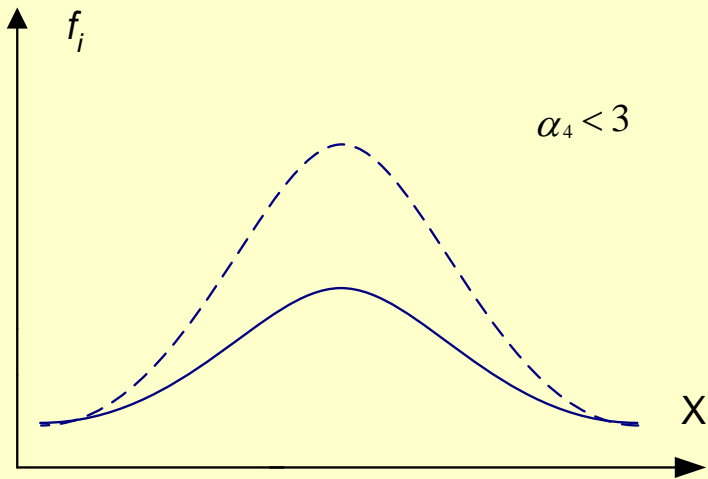
$$\alpha_4 = \frac{M_4}{\sigma_u^4}$$

Na osnovu ove formule, koeficijent pruža sledeću informaciju:

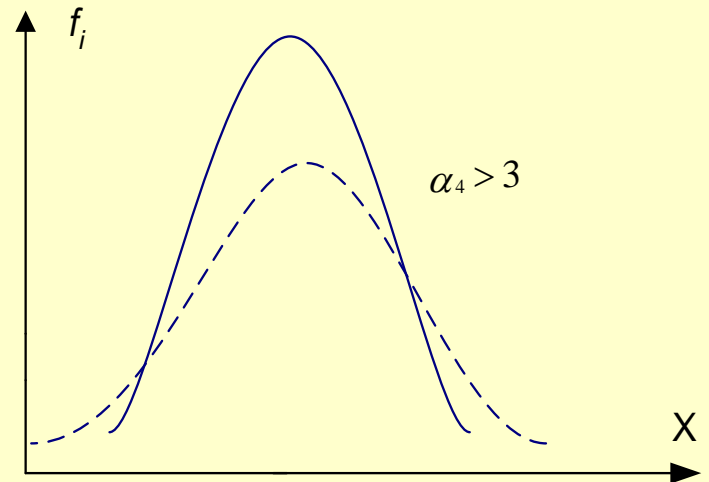
- $\alpha_4 = 0$, raspored je normalno spljošten,
- $\alpha_4 > 0$, raspored je više izdužen u odnosu na normalni raspored,
- $\alpha_4 < 0$, raspored je više spljošten u odnosu na normalni raspored.



Normalna spljoštenost rasporeda



Više spljošten raspored



Više izdužen raspored

DES-074 K:2-13 Skiciranje mera oblika rasporeda

Provera teorijskih pretpostavki kod statističkih podataka

- Normalna distribucija podataka
- Homogenost varijanse
- Postojanje intervalnih obeležja
- Nezavisnost opservacija