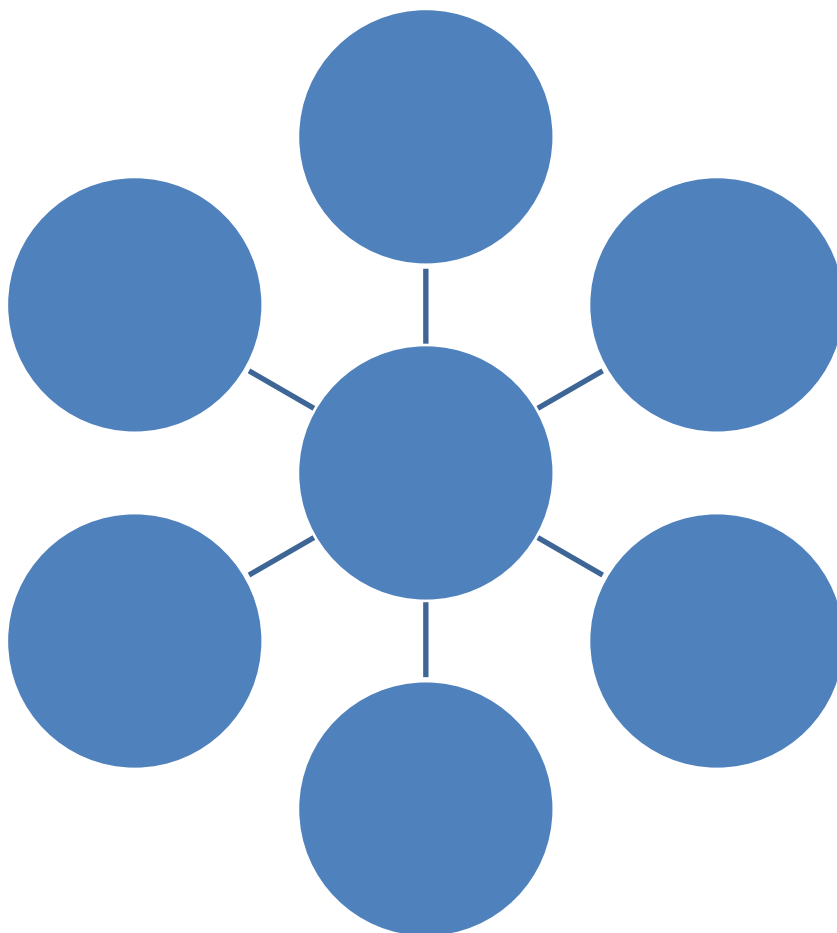




Srednja poklicna in strokovna šola Bežigrad-Ljubljana
Ptujska ulica 6, 1000 Ljubljana,
Tel.: 01/280 53 00
Fax: 01/280 53 33

OBDELAVA PODATKOV IN OSNOVE STATISTIKE



- INTERNO GRADIVO -

- SREDNJE POKLICNO IZOBRAŽEVANJE

- OBDELAVA PODATKOV IN OSNOVE STATISTIKE*Teorija podkrepljena s primeri*

Gradivo je namenjeno interni uporabi pri pouku na Srednji poklicni in strokovni šoli Bežigrad – Ljubljana in je fotokopiranje prepovedano.

AVTORJA: Mojca Rožič in Nikolaj Lipič

IZDALA: Srednja poklicna in strokovna šola Bežigrad – Ljubljana

ŠOLSKO LETO: 2010/11

NAMENJENO dijakom SREDNJEGA POKLICNEGA IZOBRAŽEVANJA

Uporabo je odobrila ravnateljica ga. Fani Al-Mansour.

VSEBINA s kazalom

<i>OSNOVE STATISTIKE</i>	2
<i>1 OSNOVNI POJMI STATISTIKE</i>	2
<i>2 UREJANJE IN GRUPIRANJE PODATKOV</i>	3
<i>3 GRAFIČNO PRIKAZOVANJE PODATKOV</i>	4
<i>4 SREDNJE VREDNOSTI</i>	7
<i>4.1. POVPREČJE ALI ARITMETIČNA SREDINA</i>	7
<i>4.2. MEDIANA (SREDIŠČNICA)</i>	8
<i>4.3. MODUS (GOSTIŠČNICA)</i>	8

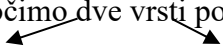
OSNOVE STATISTIKE

Beseda »statistika« izvira iz latinske besede status in ima naslednje pomene:

- prvotno: državoslovje, nauk o upravi, gospodarstvo, naseljenosti itd. države;
- veda, ki proučuje gospodarske, socialne ipd. procese na osnovi številčnih podatkov, vzetih iz opazovanja množičnih primerov danega pojava;
- številčni podatki, npr. v obliki tabele.

(Vir: Verbinc F., Slovar tujk, Cankarjeva založba, Ljubljana)

1 OSNOVNI POJMI STATISTIKE

POJEM	DEFINICIJA	PRIMER			
<i>Populacija</i>	je množica, ki jo statistično proučujemo.	Dijaki vseh šol po Sloveniji			
<i>Statistične enote</i>	so elementi populacije.	En dijak			
<i>Reprezentativen vzorec</i>	je podmnožica, s katero predstavimo celotno množico in katere elementi kar najboljše predstavljajo značilnosti celotne množice.	Ustrezen delež gimnazijcev, ustrezen delež dijakov poklicnih šol, osnovnih šol,....			
<i>Slučajen vzorec</i>	vzorec, kjer imajo vsi elementi osnovne množice enako možnost, da bodo izbrani.	Naključno izbrani dijaki naključnih šol			
<i>Numerus</i>	imenujemo moč vzorca in ga označimo z N .	število dijakov v vzorcu			
<i>Statistični znak, spremenljivka ali podatek</i>	imenujemo vrednost ali lastnost statistične enote, ki jo proučujemo. Ločimo dve vrsti podatkov: 				
	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; border-right: 1px dashed black; padding: 5px; vertical-align: top;"> <p>Številski (numerični) podatki <u>Primeri:</u> številka čevlja, starost, teža, višina ...</p> <p>Ločimo naslednji vrsti numeričnih podatkov: <i>celoštevilčni (diskretni)</i> in <i>zvezni</i> - so podatki, ki lahko dosežejo vsako vrednost iz nekega intervala</p> </td> <td style="width: 50%; padding: 5px; vertical-align: top;"> <p>Stvarni (opisni, atributivni) podatki <u>Primeri:</u> spol, rojstni kraj, državljanstvo, izobrazba, poklic, stalno bivališče, mesto zaposlitve ...</p> </td> </tr> </table>	<p>Številski (numerični) podatki <u>Primeri:</u> številka čevlja, starost, teža, višina ...</p> <p>Ločimo naslednji vrsti numeričnih podatkov: <i>celoštevilčni (diskretni)</i> in <i>zvezni</i> - so podatki, ki lahko dosežejo vsako vrednost iz nekega intervala</p>	<p>Stvarni (opisni, atributivni) podatki <u>Primeri:</u> spol, rojstni kraj, državljanstvo, izobrazba, poklic, stalno bivališče, mesto zaposlitve ...</p>	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; border-right: 1px dashed black; padding: 5px; vertical-align: top;"> <p>-numerični celoštevilski podatek: letnica rojstva dijaka (celo število)</p> <p>- numerični zvezni podatek: povprečna ocena dijaka (decimalno število)</p> <p>- stvarni podatek: interesna dejavnost dijaka (opisno – besedno)</p> </td> <td style="width: 50%; padding: 5px; vertical-align: top;"></td> </tr> </table>	<p>-numerični celoštevilski podatek: letnica rojstva dijaka (celo število)</p> <p>- numerični zvezni podatek: povprečna ocena dijaka (decimalno število)</p> <p>- stvarni podatek: interesna dejavnost dijaka (opisno – besedno)</p>
<p>Številski (numerični) podatki <u>Primeri:</u> številka čevlja, starost, teža, višina ...</p> <p>Ločimo naslednji vrsti numeričnih podatkov: <i>celoštevilčni (diskretni)</i> in <i>zvezni</i> - so podatki, ki lahko dosežejo vsako vrednost iz nekega intervala</p>	<p>Stvarni (opisni, atributivni) podatki <u>Primeri:</u> spol, rojstni kraj, državljanstvo, izobrazba, poklic, stalno bivališče, mesto zaposlitve ...</p>				
<p>-numerični celoštevilski podatek: letnica rojstva dijaka (celo število)</p> <p>- numerični zvezni podatek: povprečna ocena dijaka (decimalno število)</p> <p>- stvarni podatek: interesna dejavnost dijaka (opisno – besedno)</p>					

2 UREJANJE IN GRUPIRANJE PODATKOV

V raziskavah so najpogosteje pridobljeni podatki neurejeni in nepregledni, zato jih moramo urediti, torej urediti po velikosti v **ranžirno vrsto** ali jih združiti v skupine v **frekvenčne razrede**. Urejanje podatkov je odvisno od samih podatkov: ali so podatki diskretni ali zvezni ter od količine podatkov.

Starostni razred v letih	Število ponesrečencev - frekvenca f_j	Relativna frekvenca f_0 $f_0 = \frac{f_j}{N}$	Relativna frekvenca v %	Kumulativna frekvenca $F_j = F_{j-1} + F_j$	Relativna kumulativna frekvenca $F_j^0 = \frac{F_j}{N}$ v%
0-9	12	=12:84=0,143	14,3	12	12:84=0,143 = 14,3 %
10-19	22	0,262	26,2	12+22=34	34:84 = 0,405 =40,5 %
20-29	14	0,167	16,7	34+14=48	48:84 = 0,571 = 57,1%
30-39	9	0,107	10,7	57	0,679 = 67,5%
40-49	6	0,071	7,1	63	0,75 = 75 %
50-59	13	0,155	15,5	76	0,905 = 90,5%
60-69	8	0,095	9,5	84	1,00 = 100 %
Skupaj	84	1,00			

POJEM	DEFINICIJA
Frekvenca (f)	imenujemo posamezno število diskretnih statističnih enot iste vrednosti.
Frekvenčni razredi	<ul style="list-style-type: none"> Če je <u>diskretnih</u> podatkov zelo <i>veliko</i> ali če so podatki zvezni, jih združujemo v skupine ali frekvenčne razrede. Vse podatke, s katerimi razpolagamo, razdelimo v določeno število frekvenčnih razredov, ki so ponavadi enako široki, kar pa ni nujno. Če pa so podatki <u>zvezni</u>, so frekvenčni razredi kar intervali na abscisni osi.
Širina frekvenčnega razreda (d_k)	je razlika med zgornjo in spodnjo mejo razreda. $d_k = z_k - s_k$ z_k – zgornja meja razreda s_k – spodnja meja razreda
Sredina frekvenčnega razreda (x_k)	$x_k = \frac{z_k + s_k}{2}$ x_k – sredina frekvenčnega razreda s_k – spodnja meja razreda z_k – zgornja meja razreda
Relativna frekvenca (f^0)	pove, kolikšen delež celote pomeni posamezna vrednost statističnega znaka. Največkrat jo podajamo v %.
Kumulativna frekvenca (F)	pove, koliko podatkov je doseglo manjšo vrednost od zgornje meje frekvenčnega razreda. To je način grupiranja, ko združujemo frekvence ali frekvenčne razrede od spodaj navzgor oziroma ko »kopičimo« podatke.

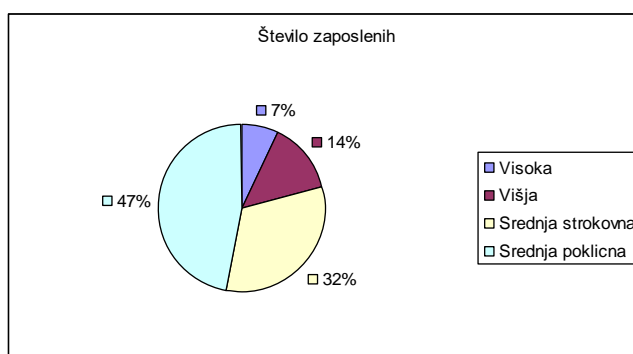
3 GRAFIČNO PRIKAZOVANJE PODATKOV

Podatke lahko prikazujemo:

- s tabelami in
- z grafikoni (vrsta grafikona je odvisna od vrste podatkov in števila frekvenčnih razredov).

Poznamo naslednje vrste grafikonov:

a) krožni diagram ali strukturalni krog



je najbolje uporabiti pri majhnem številu frekvenčnih razredov.

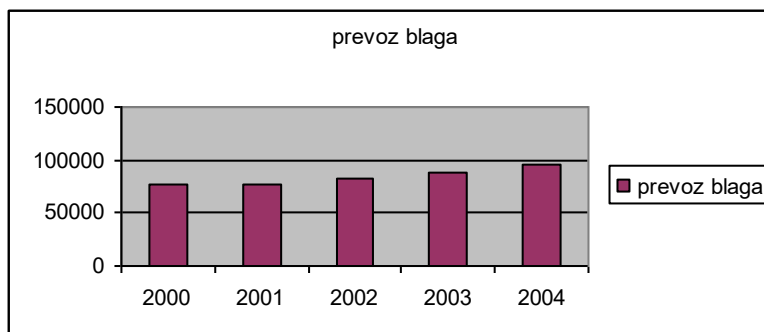
Velja: Celota v tem primeru pomeni 360° , središčne kote, ki pripadajo posameznih vrednostim, pa dobimo s procentnim računom.

b) stolpčni diagram

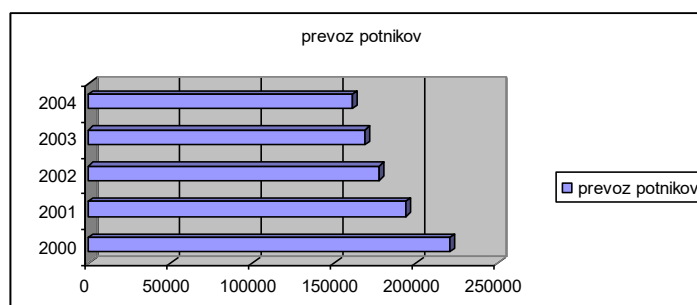
uporabljamo, če so podatki razvrščeni v veliko frekvenčnih razredov ali lahko dosežejo veliko različnih diskretnih vrednosti.

Ločimo naslednje vrste stolpčnih diagramov:

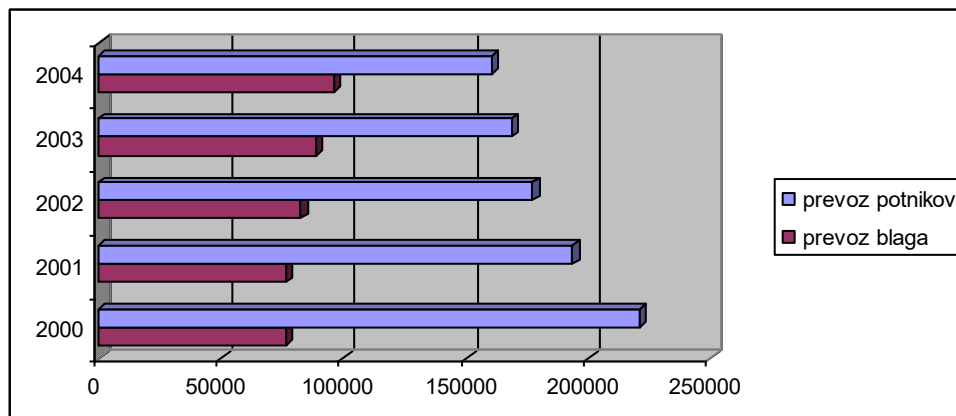
i. pokončne,



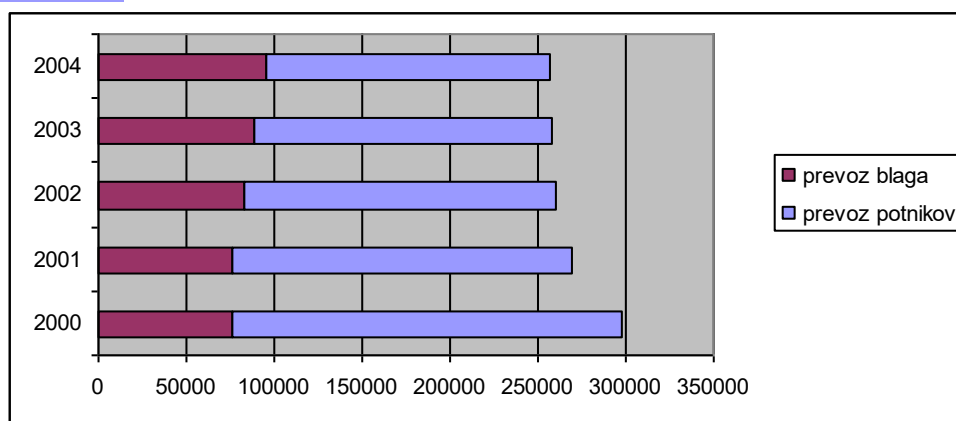
ii. ležeče,



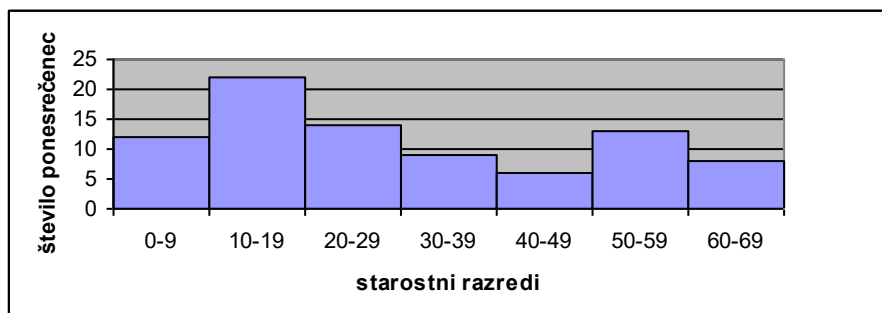
iii. sestavljene



iv. strukturne



HISTOGRAM: z njim najlažje grafično predstavimo grupirane podatke. Pri tem ni pomembno, ali so širine frekvenčnih razredov enake.



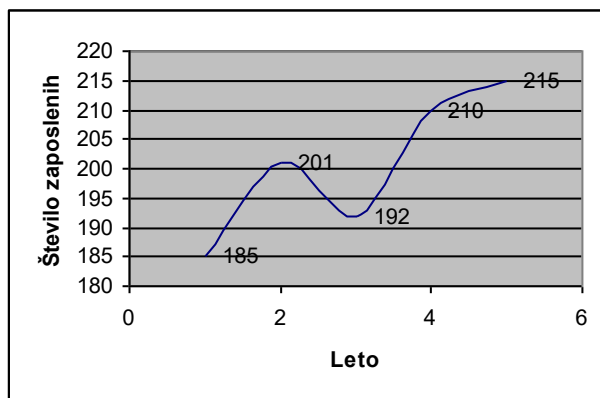
Velja:

- Na vodoravno os narišemo meje razredov;
- Na navpično os pa frekvence posameznih razredov.
- Nastanejo pravokotniki, ki so drug ob drugem, ploščina posameznega pravokotnika pa je sorazmerna frekvenci tistega razreda.

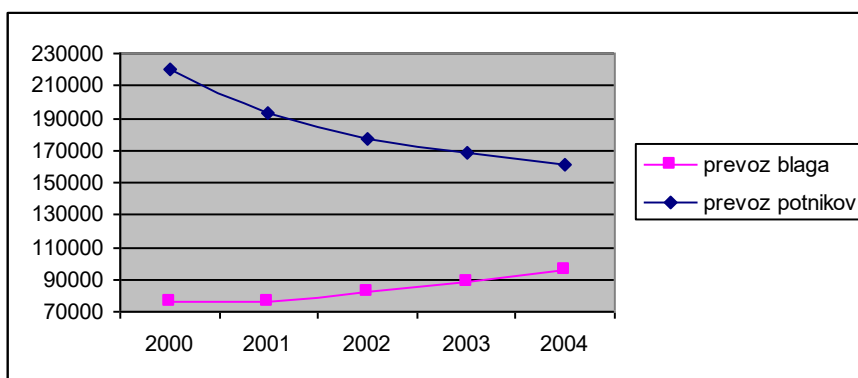
c) linijski diagram

uporabljamo za opisovanje postopnega spreminjanja vrednosti nekega podatka, največkrat skozi daljše časovno obdobje. Podatki so lahko:

- grupirani: pri teh podatkih dobimo krivuljo tako, da povežemo vrednosti sredin frekvenčnih razredov;
- zvezni: pri zveznih podatkih pa krivulja predstavlja funkcijsko odvisnost podatka od časa.



FREKVENČNI POLIGON – je poseben primer linijskega diagrama, ki je lahko sestavljen iz več krivulj, kar nam omogoča lažjo primerjavo vrednosti, lahko pa je tudi kombiniran s stolpčnim diagramom.

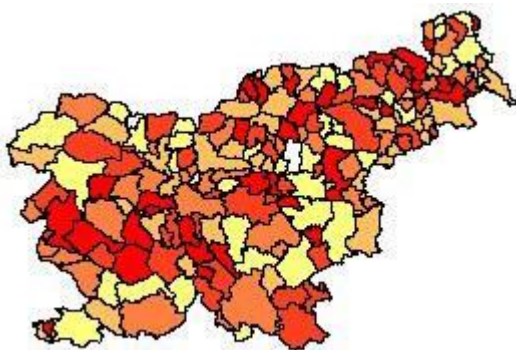


d) piktogram ali slikovni diagram

- za grafično prikazovanje uporabljamo različne figurice, oblake, deževne kapljice ...



e) obarvani zemljevid



4 SREDNJE VREDNOSTI

PRIMERI: 1, 2, 3:

Tabela predstavlja zaslužek dijaka po posameznih dnevih:

ZASLUŽEK
50 €
150 €
75 €
25€
100 €
N = 5

Tabela predstavlja ocene pisnega testa 26 dijakov:

OCENA	f
5	3
4	5
3	8
2	6
1	4
N	26

Tabela s frekvenčnimi razredi predstavlja točke dosežene pri pisnem testu:

TOČKE	f	x _s
18-20	3	19
15-17	5	16
12-14	8	13
9-11	6	10
0-8	4	4
N	26	

4.1. Povprečje ali aritmetična sredina

- je kvocient vsote vseh vrednosti statistične spremenljivke s številom vseh vrednosti.

<p>Če so vse vrednosti različne, aritmetično sredino izračunamo po formuli:</p>	$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$	<p>PRIMER 1:</p> $\bar{x} = \frac{50 + 150 + 75 + 25 + 100}{5} = \frac{400}{5} = 80$
<p>Če se vrednosti statistične spremenljivke ponavljajo (k_1 vrednosti $x_1 \dots$), je formula naslednja:</p>	$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + \dots + k_m x_m}{k_1 + k_2 + \dots + k_m}$	<p>PRIMER 2:</p> $\begin{aligned} \bar{x} &= \frac{3 \cdot 5 + 5 \cdot 4 + 8 \cdot 3 + 6 \cdot 2 + 4 \cdot 1}{3 + 5 + 8 + 6 + 4} = \\ &= \frac{15 + 20 + 24 + 12 + 4}{26} = \\ &= \frac{75}{26} = 2,88 = 2,9 \end{aligned}$

Pri grupiranih podatkih za vrednosti statistične spremenljivke vzamemo sredine frekvenčnih razredov.	$\bar{x} = \frac{k_1 x_{s1} + k_2 x_{s2} + \dots + k_m x_{sm}}{k_1 + k_2 + \dots + k_m}$	PRIMER 3: $\bar{x} = \frac{3 \cdot 19 + 5 \cdot 16 + 8 \cdot 13 + 6 \cdot 10 + 4 \cdot 4}{3 + 5 + 8 + 6 + 4} = \frac{57 + 80 + 104 + 60 + 16}{26} = \frac{317}{26} = 12,2$
--	--	--

4.2. Mediana (središčnica)

- je tista vrednost statistične spremenljivke, pri kateri je polovica vrednosti večjih, druga polovica pa manjših od nje.

Mediano izračunamo:

- pri lihem številu vrednosti kot vrednost, ki je na $\frac{n+1}{2}$ mestu po velikosti urejenih podatkov;

PRIMER 1: 25€, 50 €, 75 €, 100 €, 150 € → Me = 75 €

- pri sodem številu podatkov pa ni pravega srednjega podatka, zato za mediano vzamemo aritmetično sredino srednjih dveh podatkov.

PRIMER 2:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5

$$Me = \frac{3+3}{2} = 3$$

4.3. Modus (gostiščnica)

- je vrednost podatka, ki se v množici vseh vrednosti najpogosteje ponavlja, če sta taki vrednosti dve, govorimo o bimodalni porazdelitvi podatkov.

PRIMER 2:

$$Mo = 3$$

Za grupirane podatke namesto modusa poiščemo modalni razred: to je tisti razred, ki ima največjo frekvenčno gostoto oz. mu ustreza ploščinsko največji pravokotnik v histogramu.

PRIMER 3:

$$\text{Modalni razred} = 12-14$$