Anuška Ferligoj, Katja Lozar Manfreda, Aleš Žiberna:

OSNOVE STATISTIKE NA PROSOJNICAH

Študijsko gradivo pri predmetu Statistika. Fakulteta za družbene vede, Univerza v Ljubljani

Ljubljana, 2010

2 OPISNA STATISTIKA

2.1 KORAKI STATISTIČNE ANALIZE

1. Določitev **vsebine in namena** statističnega proučevanja; opredelitev predmeta opazovanja (enote in populacije) in vsebine opazovanja (spremenljivk).
2. **Statistično opazovanje**; vrste opazovanj:
   - opazovanje cele populacije (npr. popisi, tekoče registracije),
   - opazovanje vzorca (npr. ankete).
3. **Enostavna obdelava**: urejanje, razvrščanje podatkov ter izračun osnovnih statističnih karakteristik.
4. **Analitična obdelava**.

2.2 UREJANJE IN PRIKAZOVANJE PODATKOV

- Z namenom preglednosti podatke uredimo v frekvenčno porazdelitev.
- **Frekvenčna porazdelitev** spremenljivke je tabela, ki jo določajo vrednosti ali skupine vrednosti in njihove frekvence.
- Skupine vrednosti (razrede) oblikujemo, če gre za intervalno ali razmernostno spremenljivko, ki ima veliko (nepregledno) število različnih vrednosti.
- Če je spremenljivka vsaj ordinalnega značaja, vrednosti (ali skupine vrednosti) uredimo od najmanjše do največje.
- Frekvenčno porazdelitev lahko tudi grafično predstavimo.

2.2.1 Frekvenčne tabele in grafi za nominalne in ordinalne spremenljivke

PRIMER

Enote: študenti

Spremenljivka: *X* – "*Brez česa bi najlaže živeli?*" z odgovori "*brez TV*", "*brez mobilnika*", "*brez interneta*"

Število enot: $N = 300$

Frekvenčna porazdelitev

Vrednosti spremenljivke (anketni odgovori) — (absolutne) frekvence (št. študentov) — relativne frekvence strukturni odstotki (% študentov)

x_i	fi	$fi\%$
brez TV	30	10
brez mobilnika	180	60
brez interneta	90	30
Skupaj N	300	100

število vseh enot

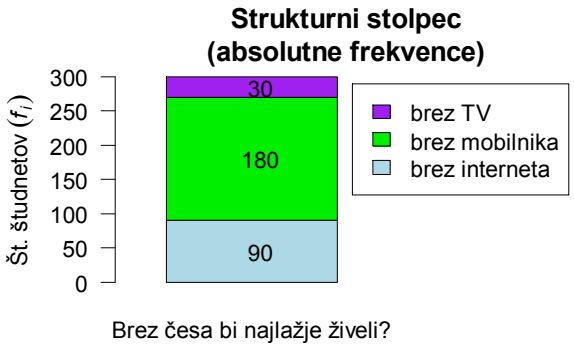
Osnovni pojmi

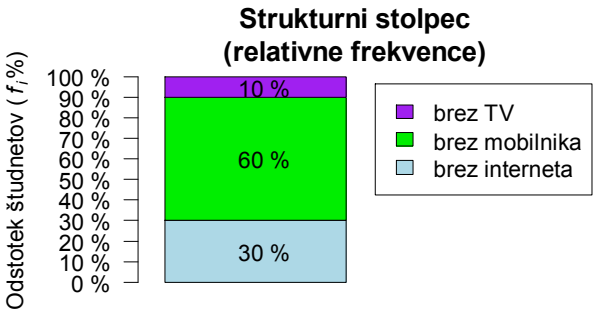
- Podatki so urejeni v **frekvenčno porazdelitev**, t.j. tabelo, kjer je v vsaki vrstici zapisana ena izmed možnih vrednosti spremenljivke ($x_1, x_2, \ldots x_m$), poleg pa njena frekvenca.
- **Frekvenca** je število enot, ki imajo določeno vrednost spremenljivke. Označujemo jih z f_i – frekvenca za i-to vrednost spremenljivke.
- Smiselno je izračunati **relativne frekvence**, im. **strukturni odstotki**. Relativna frekvenca je % enot med vsemi enotami, ki imajo določeno vrednost spremenljivke. Označujemo jih z $f_i\%$ - relativna frekvenca za i-to vrednost spremenljivke. Izračunamo jih po naslednjem obrazcu:

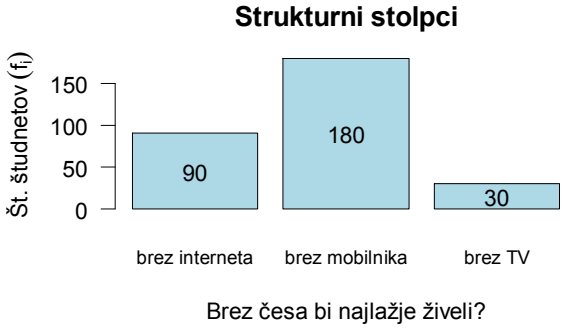
$$f_i\% = \frac{f_i}{N} \cdot 100$$

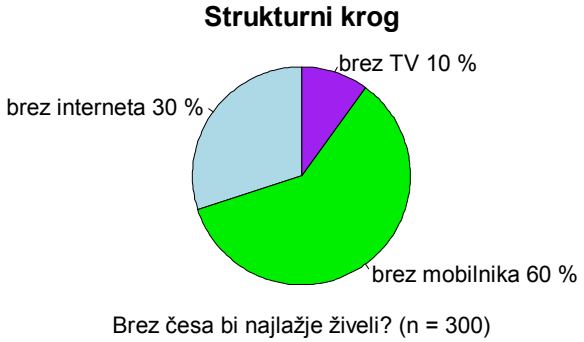
Grafična predstavitev

- Frekvenčno porazdelitev za nominalne in ordinalne spremenljivke grafično prikazujemo z liki, npr. **strukturnimi stolpci** (angl. *barchart*) ali **strukturnimi krogi** (angl. *pie*).
- Ploščine likov naj bodo sorazmerne frekvencam.
- Legenda naj bo primerno urejena.
- Vsaka grafična predstavitev naj ima naslov.
- Grafična predstavitev naj bo samozadostna.









2.2.2 Frekvenčne tabele in grafi za intervalne in razmernostne spremenljivke

PRIMER

Enota: gospodinjstvo

Spremenljivka: X – število članov gospodinjstva

Število enot: N = 50

Podatki so: 1 3 4 3 5 6 2 3 10 4 3 6 4 7 3 7 6 8 2 6
2 4 5 4 4 4 3 2 3 4 4 4 1 4 8 3 2 1 4 5
9 6 4 6 8 7 5 5 5 9

Frekvenčna porazdelitev

x_i	f_i	f_i%	F_i	F_i%
1	3	6	3	6
2	5	10	8	16
3	8	16	16	32
4	13	26	29	58
5	6	12	35	70
6	6	12	41	82
7	3	6	44	88
8	3	6	47	94
9	2	4	49	98
10	1	2	50	100
N	50	100		

Osnovni pojmi

- Podatke uredimo v **frekvenčno porazdelitev**, t.j. tabelo, kjer je v vsaki vrstici zapisana ena izmed možnih vrednosti spremenljivke ($x_1, x_2, \ldots x_m$), poleg pa njena frekvenca.
- **Frekvenca** je število enot, ki imajo določeno vrednost spremenljivke. Označujemo jih z f_i – frekvenca za i-to vrednost spremenljivke.
- Smiselno je izračunati **relativne frekvence.** Relativna frekvenca je % enot med vsemi enotami, ki imajo določeno vrednost spremenljivke. Označujemo jih z f_i% - relativna frekvenca za i-to vrednost spremenljivke. Izračunamo jih po naslednjem obrazcu:

$$f_i\% = \frac{f_i}{N} \cdot 100$$

- V primeru ordinalnih, intervalnih in razmernostnih spremenljivk je smiselno računati tudi **kumulativne frekvence (kumulative)**. Kumulativna frekvenca za neko vrednost spremenljivke, vključujoč to vrednost, je seštevek frekvenc za vrednosti pred njo ter to frekvenco. Označimo jo z F_i - kumulativna frekvenca za i-to vrednost. Pove nam, koliko enot ima vrednosti manjše ali enake od določene vrednosti. Izračunamo jo po naslednjem obrazcu:

$$F_i = F_{i-1} + f_i = f_1 + f_2 + \ldots + f_i$$

- Izračunamo lahko tudi **relativno kumulativno frekvenco**, ki nam pove, kolikšen % enot ima vrednosti manjše ali enake od določene vrednosti. Izračunamo jo po naslednjem obrazcu:

$$F_i\% = \frac{F_i}{N} \cdot 100$$

Grupiranje – razvrščanje v razrede

- Če je število možnih vrednosti spremenljivke veliko, je preglednost podatkov majhna. Zato sorodne vrednosti razvrstimo v **skupine – razrede**.
- Skupine vrednosti morajo biti določene **enolično**: vsaka enota s svojo vrednostjo je lahko uvrščena v eno in samo eno skupino vrednosti.
- Običajno so skupine vrednosti (razredi) enako široki.
- Podatke uredimo v **frekvenčno porazdelitev**, t.j. tabelo, kjer je v vsaki vrstici zapisana ena izmed skupin vrednosti (en razred), poleg pa njena frekvenca.

Nezvezna ureditev - če so vrednosti spremenljivke diskretne

Razredi	f_i	f_i%	F_i	F_i%	x_{imin}	x_{imax}	x_i	d_i
1-2	8	16	8	16	0.5	2.5	1.5	2
3-4	21	42	29	58	2.5	4.5	3.5	2
5-6	12	24	41	82	4.5	6.5	5.5	2
7-8	6	12	47	94	6.5	8.5	7.5	2
9-10	3	6	50	100	8.5	10.5	9.5	2
Skupaj N	50	100						

Zvezna ureditev - če so vrednosti spremenljivke zvezne (ali diskretne)

Razredi	f_i	f_i%	F_i	F_i%	x_{imin}	x_{imax}	x_i	d_i
1 – pod 3	8	16	8	16	1	3	2	2
3 – pod 5	21	42	29	58	3	5	4	2
5 – pod 7	12	24	41	82	5	7	6	2
7 – pod 9	6	12	47	94	7	9	8	2
9 – pod 11	3	6	50	100	9	11	10	2
Skupaj N	50	100						

Razredi	f_i	f_i%	F_i	F_i%	x_{imin}	x_{imax}	x_i	d_i
nad 0- 2	8	16	8	16	0	2	1	2
nad 2 - 4	21	42	29	58	2	4	3	2
nad 4 - 6	12	24	41	82	4	6	5	2
nad 6 - 8	6	12	47	94	6	8	7	2
nad 8 - 10	3	6	50	100	8	10	9	2
Skupaj N	50	100						

- **Frekvenca** je v tem primeru število enot, ki imajo vrednost v določeni skupini vrednosti (razredu). Označujemo jih z f_i – frekvenca za *i*-ti razred.
- Tudi v tem primeru je smiselno izračunati **relativne frekvence**. Relativna frekvenca je % enot med vsemi enotami, ki imajo vrednost v določenem razredu. Označujemo jih z f_i% - relativna frekvenca za *i*-ti razred. Izračunamo jih po naslednjem obrazcu:

$$f_i\% = \frac{f_i}{N} \cdot 100$$

- **Kumulativna frekvenca** za neko skupino vrednosti in vključujoč to skupino (razred) je seštevek frekvenc za skupine vrednosti (razrede) pred njo ter frekvenco te skupine. Označimo jo z F_i - kumulativna frekvenca za *i*-ti razred. Pove nam, koliko enot ima vrednosti manjše ali enake od te skupine vrednosti, natančneje manjše od zgornje meje tega razreda. Izračunamo jo po naslednjem obrazcu:

$$F_i = F_{i-1} + f_i = f_1 + f_2 + \ldots + f_i$$

- **Relativna kumulativna frekvenca** za nek razred pa pove, kolikšen % enot ima vrednosti manjše od zgornje meje razreda. Izračunamo jo po naslednjem obrazcu:

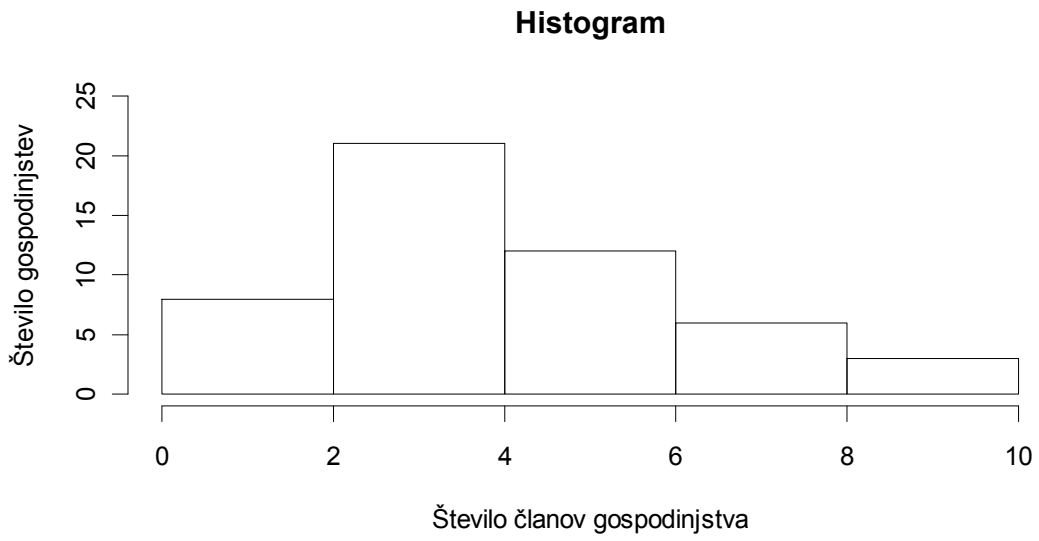
$$F_i\% = \frac{F_i}{N} \cdot 100$$

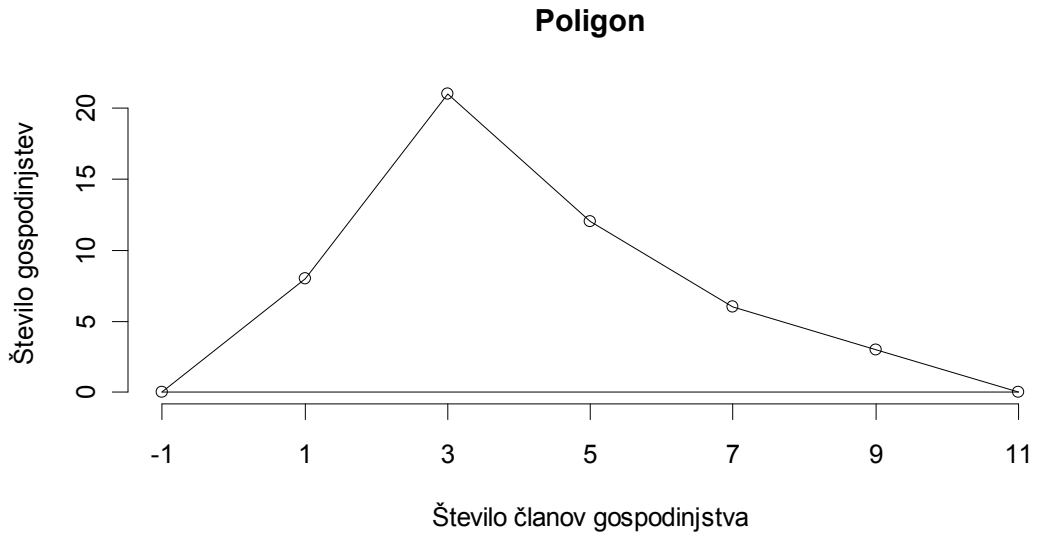
Grafična predstavitev

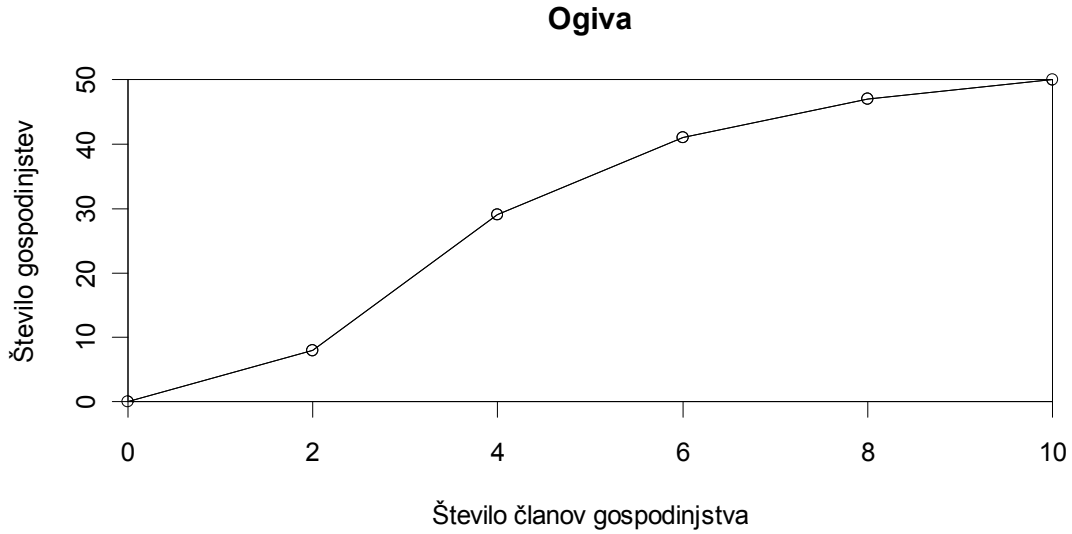
Predpostavimo, da so razredi enako široki.

- **HISTOGRAM**: drug poleg drugega rišemo stolpce – pravokotnike, katerih višina je sorazmerna frekvenci v razredu. Širina pravokotnikov je enaka, ker so razredi enako široki.
- **POLIGON**: v koordinatnem sistemu zaznamujemo točke (x_i, f_i), kjer je x_i sredina i-tega razreda in f_i njegova frekvenca. K tem točkam dodamo še točki $(x_o, 0)$ in $(x_{k+1}, 0)$, če je v frekvenčni porazdelitvi k razredov. Točke zvežemo z daljicami.
- **OGIVA**: grafična predstavitev kumulativne frekvenčne porazdelitve s poligonom, kjer v koordinatni sistem nanašamo točke $(x_{i,max}, F_i)$. Dodamo tudi točko $(x_{1,min}, 0)$.

Razredi	f_i	f_i%	F_i	F_i%	x_{imin}	x_{imax}	x_i
nad 0- 2	8	16	8	16	0	2	1
nad 2 - 4	21	42	29	58	2	4	3
nad 4 - 6	12	24	41	82	4	6	5
nad 6 - 8	6	12	47	94	6	8	7
nad 8 - 10	3	6	50	100	8	10	9
Skupaj N	50	100					





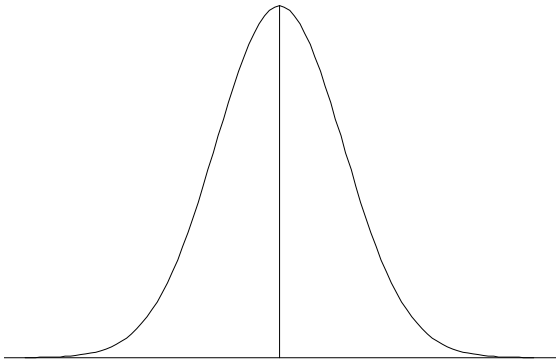


Oblike frekvenčnih porazdelitev

Frekvenčna porazdelitev prikazuje variiranje ali razpršenost vrednosti spremenljivke. Razpršenost je rezultat individualnih, posamičnih faktorjev, ki vplivajo na posamezne enote. Ti vplivi so najrazličnejši in njihova posledica so različne oblike frekvenčne porazdelitve.

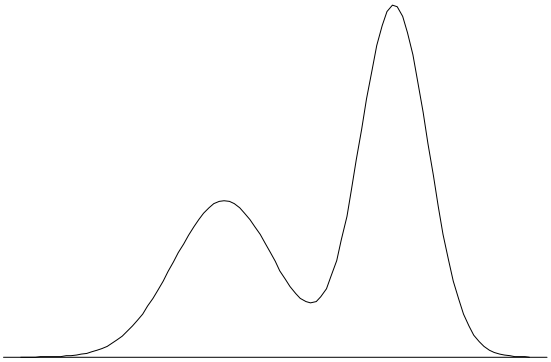
Nekatere oblike frekvenčnih porazdelitev se večkrat pojavljajo in te porazdelitve so dobile svoja imena: normalna, unimodalna, bimodalna, polimodalna, asimetrična v levo, asimetrična v desno, sploščena, koničasta, J oblike, U oblike.

Frekvenčna porazdelitev, s katero običajno primerjamo določeno frekvenčno porazdelitev, je **normalna porazdelitev**, ki je unimodalna (ima en vrh), simetrična in zvonaste oblike.

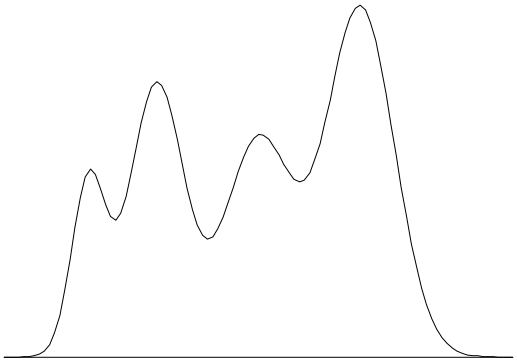


Oblika porazdelitev se lahko od normalne bolj ali manj razlikuje zaradi nehomogenosti populacije, okrnjenega delovanja določenih faktorjev itd. Zato je oblika porazdelitve lahko:

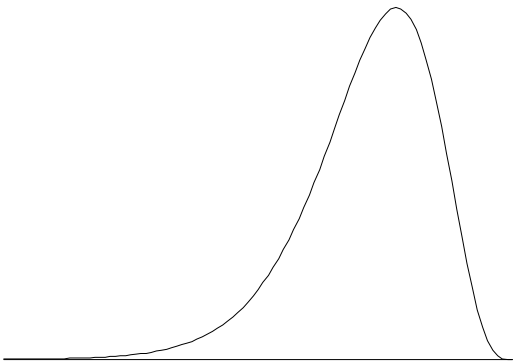
- **bimodalna**, če ima dva vrha**;**



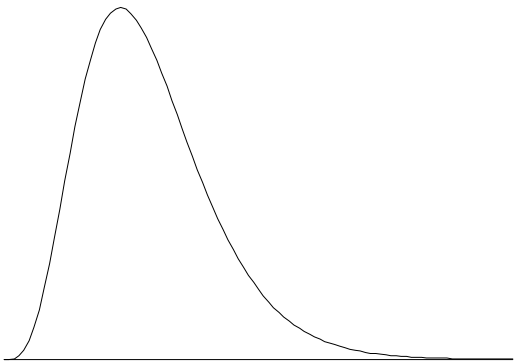
- **polimodalna** z več vrhovi;



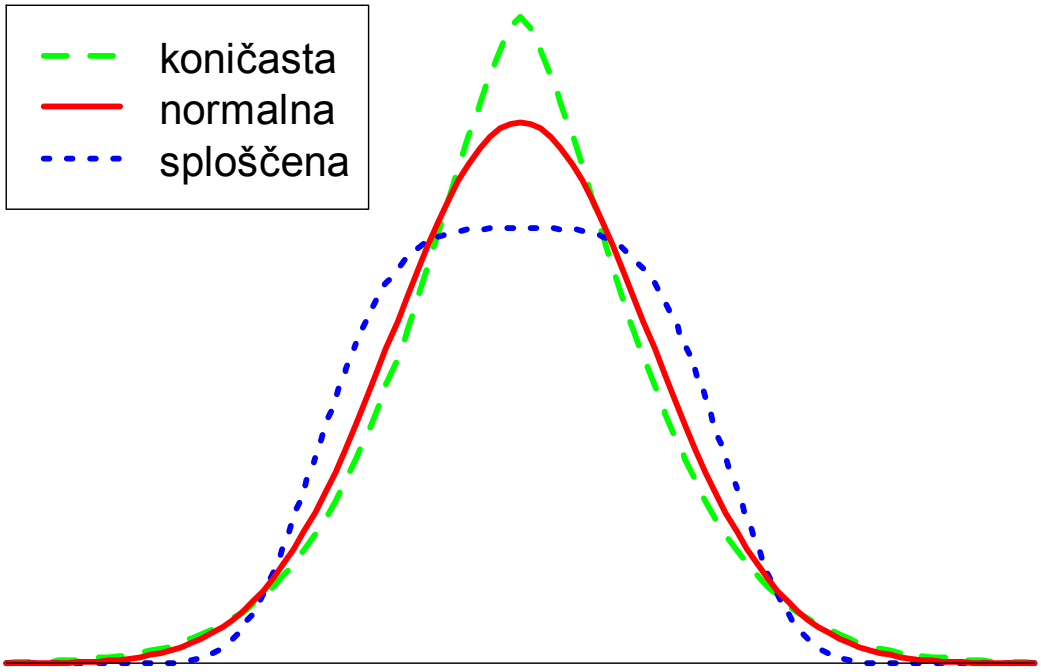
- **asimetrična v levo**, če se rep vleče na levo;



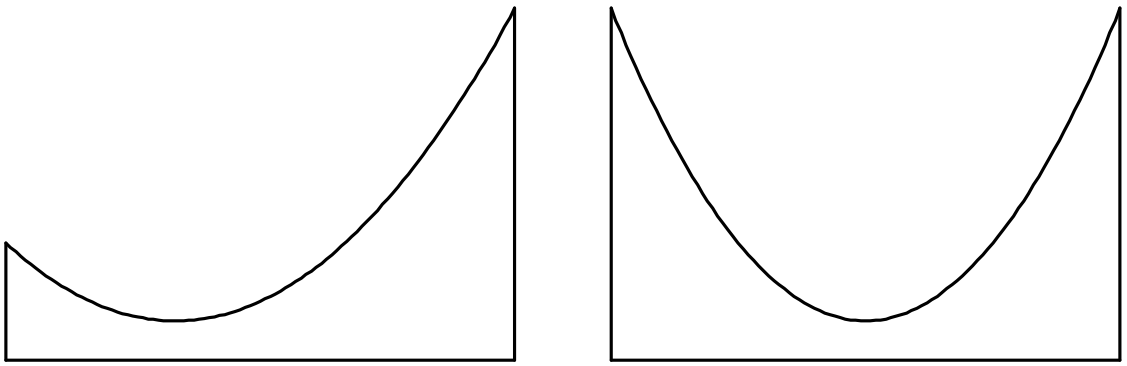
- **asimetrična v desno**, če se rep vleče v desno;



- bolj **koničasta** ali **sploščena** od normalne porazdelitve;



- **J** ali **U** oblike.



Zavajajoči grafi

- Grafi morajo biti narisani tako, da realno prikažejo osnovne značilnosti podatkov.
- Včasih so grafi nepravilno narisani, tako da celo zavajajo bralca. Npr. potrebna je posebna pozornost pri izbiri velikosti merske enote v koordinatnem sistemu.

2.3 OSNOVNI STATISTIČNI IZRAČUNI

2.3.1 Kvantili

Kvantili so statistični izračuni, ki nam pomagajo določiti, kje se nahaja posamezna enota v primerjavi z drugimi enotami in njihovimi vrednostmi. Računamo jih lahko le za intervalne in razmernostne spremenljivke.

Primer

Enota: študent 1. letnika FDV

Spremenljivka: število doseženih točk na izpitu

S kvantili lahko odgovorimo npr. na naslednji dve vprašanji:

1. Kolikšno je število doseženih točk na izpitu za 10 % najboljših študentov?
2. Kolikšen del študentov je doseglo 55 točk ali več?

Osnovni pojmi:

- **Ranžirna vrsta:** enote z ustreznimi vrednostmi spremenljivke uredimo od tiste z najmanjšo vrednostjo do tiste z največjo vrednostjo. Im. tudi ranžirni razmik.
- **Rang *R***: vsaki enot v ranžirni vrsti priredimo zaporedno mesto.

  Primer:

  Izpitne ocene (0-100) 10 študentov so: *82, 78, 58, 68, 86, 59, 46, 45, 17, 92*

  Ranžirna vrsta je:

x_i	17	45	46	58	59	68	78	82	86	92	←vrednosti
R_i	1	2	3	4	5	6	7	8	9	10	←rangi

- **Kvantilni rang *P***: pove, na katerem delu celotne ranžirne vrste / ranžirnega razmika leži določena enota oz. koliki del celotne ranžirne vrste / ranžirnega razmika ima manjše vrednosti od dane vrednost. Izračunamo ga po naslednjem obrazcu:

$$P = \frac{R - 0{,}5}{N}$$

- **Kvantil:** vrednost spremenljivke, ki pripada določenemu kvantilnemu rangu.

Običajni kvantili:

- **Mediana** *Me (P=0.5)*

  = kvantil, ki razdeli ranžirno vrsto na dva dela.

  = kvantil, ki pripada kvantilnemu rangu *P=0.5*.

  = vrednost, od katere ima ½ enot manjšo, ½ pa večjo vrednost.

- **Kvartili** Q_1 *(P=0.25),* Q_2 *(P=0.5),* Q_3 *(P=0.75)*

  = kvantili, ki razdelijo ranžirno vrsto na četrtine.

  = npr. prvi kvartil je vrednost, ki pripada kvantilnemu rangu *P=0.25*.

  = npr. prvi kvartil je vrednost, od katere ima 1/4 enot manjšo, 3/4 pa večjo vrednost.

- **Decili** D_1 *(P=0.1),* D_2 *(P=0.2), ...,* D_9 *(P=0.9),*

  = kvantili, ki razdelijo ranžirno vrsto na desetine.

  = npr. prvi decil je vrednost, ki pripada kvantilnemu rangu *P=0.1*.

  = npr. prvi decil je vrednost, od katere ima 1/10 enot manjšo, 9/10 enot pa večjo vrednost.

- **Centili** C_1 *(P=0.01),* C_2 *(P=0.02), ...,* C_{99} *(P=0.99),*

  = kvantili, ki razdelijo ranžirno vrsto na stotine.

  = npr. prvi centil je vrednost, ki pripada kvantilnemu rangu *P=0.01*.

  = npr. prvi centil je vrednost, od katere ima 1/100 enot manjšo, 99/100 enot pa večjo vrednost.

$$Me=Q2=D5=C50$$

$$D1=C10$$

$$Q1=C25$$

Računanje kvantilov

Xi	17	45	46	58	59	68	78	82	86	92
Ri	1	2	3	4	5	6	7	8	9	10

Koliko točk je dosegla polovica študentov z najmanjšim oz. največjim številom točk? Iščemo vrednost, ki se nahaja točno na polovici med 5. in 6. enoto v ranžirni vrsti. To je vrednost, ki je točno na polovici med 59 in 68, torej 63.5.

Takšno vrednost natančno določimo z **linearno interpolacijo**, ki upošteva razmerja med rangi ter vrednostmi, ki rangom pripadajo. Če je R med rangoma R_0 in R_1, je ustrezni x med vrednostima x_0 in x_1.

$$\frac{R-R_0}{R_1-R_0}=\frac{x-x_0}{x_1-x_0}$$

Linearna interpolacija je seveda mogoča le pri intervanih in razmernostnih spremenljivkah.

Primer

1. **Kakšno oceno ima polovica študentov, ki ima najnižjo oceno? (Izračunajmo za navedene ocene študentov mediano, t.j. kvantil, ki pripada kvantilnemu rangu *P*=0.5.)**

$$R = N \cdot P + 0.5 = 10 \cdot 0.5 + 0.5 = 5.5$$

Rang leži med rangoma $R_0 = 5$ in $R_1 = 6$ in ustrezna vrednost (mediana) leži med vrednostima $x_0 = 59$ in $x_1 = 68$. Uporabimo linearno interpolacijo:

$$\frac{R-R_0}{R_1-R_o}=\frac{x-x_0}{x_1-x_o}$$
$$\frac{5.5-5}{6-5}=\frac{x-59}{68-59}$$
$$x = Me = x_{0.5} = 63.5$$

Ocena, ki razdeli ranžirno vrsto na polovico oz. leži točno na polovici ranžirne vrste, torej mediana, je 63.5.

Polovica študentov je dosegla manj kot 63.5 točk, polovica študentov pa več kot 63.5 točk.

2. **Kolikšen del študentov ima manj kot 50 točk? (Izračunajmo kvantilni rang za vrednost 50 (*x*=50).)**

Vrednost x=50 leži med sosednjima vrednostima x_0=46 in x_1=58, ki mu pripadata ranga R_0=3 in R_1=4. Uporabimo linearno interpolacijo:

$$\frac{R-R_0}{R_1-R_o}=\frac{x-x_0}{x_1-x_o}$$
$$\frac{R-3}{4-3}=\frac{50-46}{58-46}$$
$$R = 3.33$$

Vrednost x=50 bi bila 3.33 enota v ranžirni vrsti. Izračunajmo kvantilni rang (kaj to pomeni v relativnem smislu):

$$P=\frac{R-0.5}{N}=\frac{3.3-0.5}{10}=0.283$$

Vrednost 50 je vrednost, za katero velja, da ima 28.3% enot manjšo vrednost.

28% študentov ima manj kot 50 točk.

Računamo torej lahko:

1) Vrednost, ki pripada določenemu kvantilnemu rangu - katera vrednost leži na določenem mestu v ranžirni vrsti.
   x_p = ?
   P je podan; iz *P* in *N* izračunamo *R*; z linearno interpolacijo določimo *x*, ki pripada *R*.
2) Kvantilni rang za določeno vrednost - na katerem mestu v ranžirni vrsti leži določena vrednost.
   P = ?
   x je podan; z linearno interpolacijo določimo *R* za x, iz R in N izračunamo *P*.

2.3.2 Srednje vrednosti

Pregled vrednosti spremenljivke dobimo z ranžirno vrsto ali - v primeru večjega števila enot - s frekvenčno porazdelitvijo. Pri tem nas zanima, ali iz podatkov lahko razberemo neko reprezentativno vrednost spremenljivke in jo im. **srednja vrednost**. Ta vrednost naj bila najbolj tipična, običajna, reprezentativna, normalna, pričakovana, pogosta ...

Ostale vrednosti se od srednje vrednosti nekoliko odklanjajo, variirajo. Bolj kot se posamezne vrednosti odklanjajo od srednje vrednosti, tem slabše ta srednja vrednost predstavlja spremenljivko.

Obstaja več vrst srednjih vrednosti. Mi bomo spoznali naslednje:
- mediana *Me*
- modus *Mo*
- aritmetična sredina μ

2.3.2.1 Mediana

Mediana je tista vrednost spremenljivke, od katere je ravno toliko manjših vrednosti od nje, kolikor jih je večjih od nje. Torej tista vrednosti, ki razdeli ranžirno vrsto na polovico, t.j. kvantil, ki pripada kvantilnemu rangu *P*=0.5.

Mediana je primerna srednja vrednost za ordinalne spremenljivke.
Računamo jo iz ranžirne vrste.

- V ranžirni vrsti z **lihim številom enot** $N = 2m + 1$ je mediana x_{m+1} vrednost v ranžirni vrsti.

  Primer: 3, 6, 7, 8, 10, 13, 14

  Ker je N=7, je mediana na 4. mestu ($Me=x_{3+1}$), torej Me = 8.

- V ranžirni vrsti s **sodim številom enot** $N = 2m$ je mediana vrednost, ki je na sredini med srednjima dvema vrednostima, torej $Me= (x_m+x_{m+1})/2$.

  Primer: 3, 3, 8, 10, 11, 14

  Ker je N=6, je mediana med 3. in 4. vrednostjo, torej Me=(8+10)/2=9.

2.3.2.2 Modus

Modus je edina srednja vrednost, ki je primerna za nominalne spremenljivke.

Spremenljivka z diskretnimi vrednostmi

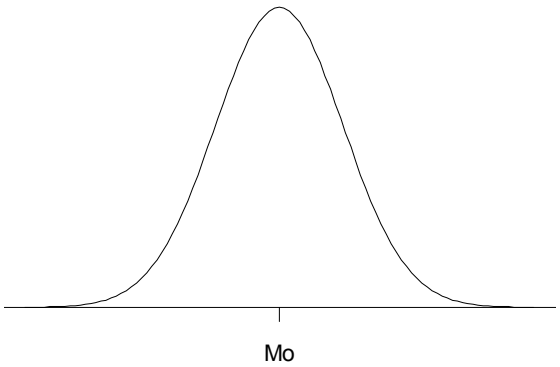
Modus = Vrednost spremenljivke, ki se najpogosteje pojavlja.

Primeri:

2, 3, 4, 4, 4, 5, 7, 9	*Mo*=4	
2, 2, 5, 5, 5, 6, 7, 7, 7, 8, 12	*Mo1=5 Mo2=7*	Modusov je lahko več
2, 3, 7, 9, 12	Modusa ni.	

Spremenljivka z zveznimi vrednostmi

Modus = Vrednost spremenljivke, okoli katere se ostale vrednosti najbolj gostijo.



Globalni in lokalni modusi

Globalni modus

Pravi modus je "globalni", to je tista vrednost spremenljivke, ki se **najpogosteje** pojavlja oz. okoli katere se ostale vrednosti **najbolj** gostijo.

Lokalni modus

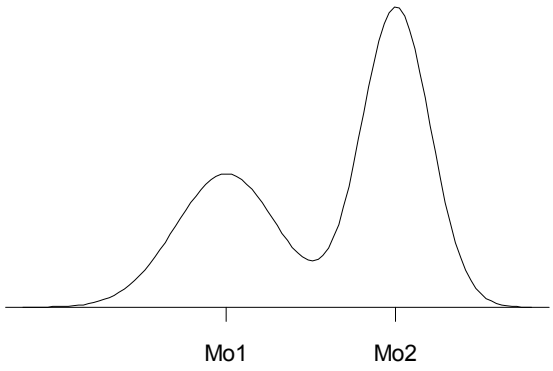
"Lokalni" modus je tista vrednost spremenljivke, ki se pojavlja pogosteje kot njej bližnje vrednosti oz. okoli katere se ostale vrednosti gostijo. To ni pravi modus.

Primeri:

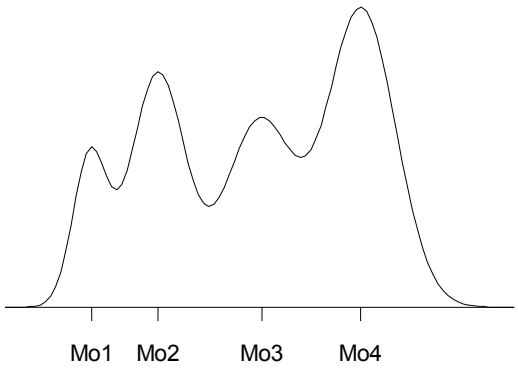
2, 2, 3, 4, 4, 4, 5, 7, 7, 9	Globalni modus: 4 Lokalna modusa: 2 in 7
2, 2, 5, 5, 5, 6, 7, 7, 7, 8, 12	Dva globalna modusa: 5 in 7
2, 3, 4, 4, 4, 5, 5, 7, 9	Globalni modus: 4 Lokalnega modusa ni

Primer: več **lokalnih** modusov in en globalni modus pri spremenljivki z zveznimi vrednostmi

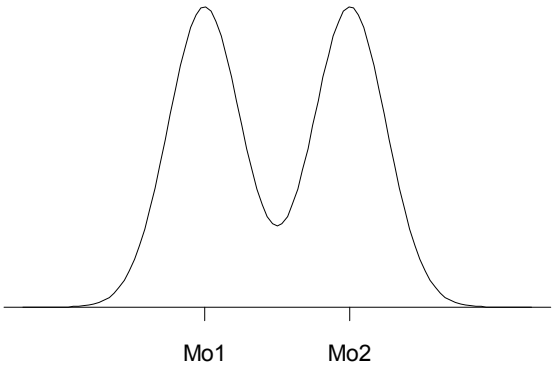
- **Bimodalna porazdelitev**



- **Polimodalna porazdelitev**



Primer: več **globalnih** modusov pri spremenljivki z zveznimi vrednostmi



Izračun modusa iz frekvenčne porazdelitve

Če modus razumemo kot vrednost spremenljivke, okoli katere se vrednosti najbolj gostijo, ga lahko najlažje določamo iz frekvenčne porazdelitve. Razred z največjo frekvenco vsebuje modus, zato ga imenujemo **modalni razred**.

- Prvi približek modusu je lahko sredina modalnega razreda.
- Natančneje modus določimo s pomočjo naslednjega obrazca:

$$Mo = x_{0,min} + d\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_{+1}}$$

$x_{o,min}$ ... spodnja meja modalnega razreda
d ... širina razreda
f_o ... frekvenca modalnega razreda
f_{+1} ... frekvenca razreda za modalnim razredom
f_{-1} ... frekvenca razreda pred modalnim razredom

Primer: Frekvenčna porazdelitev prikazuje število članov 50 gospodinjstev:

št. članov	f_i
1-2	8
3-4	21
5-6	12
7-8	6
9-10	3
	50

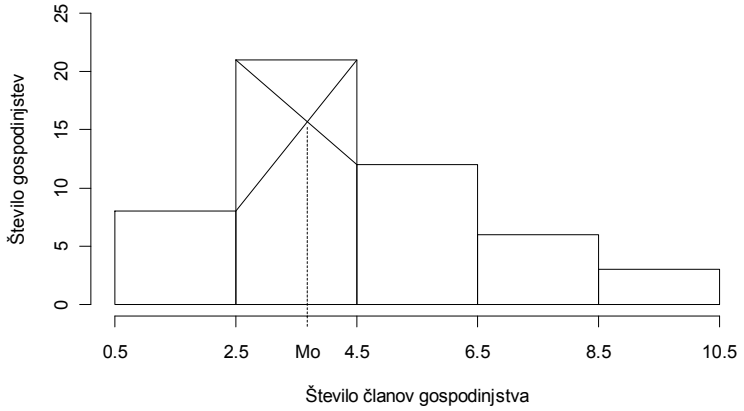
Iz frekvenčne porazdelitve razberemo potrebne podatke za izračun modusa:

$x_{0,min} = 2.5$, $d = 2$, $f_0 = 21$, $f_{-1} = 8$, $f_{+1} = 12$

$$Mo = x_{0,\min} + d\frac{f_o - f_{-1}}{2f_o - f_{-1} - f_{+1}} \qquad Mo = 2.5 + 2\frac{21 - 8}{2 \cdot 21 - 8 - 12} = 3.68$$

Največ gospodinjstev ima 3.68 članov.

Grafično pa si modus ponazorimo takole:



2.3.2.3 Aritmetična sredina

Aritmetična sredina ali povprečje je vsota vseh vrednosti, deljena s številom enot v populaciji. Primerna je za intervalne in razmernostne, približno normalno porazdeljene spremenljivke.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}(x_1 + x_2 + ... + x_N)$$

μ ... aritmetična sredina

x_i ... vrednost spremenljivke X na i-ti enoti

N ... število vseh enot

Σ ... sumacijski znak – "Seštej vrednosti spremenljivke X, ko i teče od *1* do N, torej za vse enote."

Primer: 1, 2, 3, 4, 5

$$\mu = \frac{1+2+3+4+5}{5} = 3$$

Aritmetično sredino lahko definiramo kot vrednost, za katero velja:

$$\sum_{i=1}^{N}(x_i - \mu) = 0$$

Zgornji obrazec za izračun aritmetične sredine velja, kadar imamo podatke, ki **niso urejeni v frekvenčno porazdelitev**.

Če pa imamo podatke, ki **so urejeni v frekvenčno porazdelitev**, uporabljamo naslednji obrazec:

$$\mu = \frac{1}{N}\sum_{i=1}^{k} f_i x_i = \frac{1}{N}(f_1 x_1 + f_2 x_2 + ... + f_k x_k)$$

μ ... aritmetična sredina

x_i ... reprezentativna vrednost (običajno sredina) i-tega razreda spremenljivke X

f_i ... frekvenca za i-ti razred spremenljivke X (št. enot, ki so v tem razredu)

k ... število vseh razredov v frekvenčni porazdelitvi

N ... število vseh enot

Σ ... sumacijski znak – "Seštej produkt reprezentativne vrednosti razreda spremenljivke X in frekvence, ko i teče od *1* do k, torej za vse vrednosti."

Pozor: V primeru, da uporabimo kot reprezentativno vrednost sredino razreda in ne aritmetične sredine vrednosti v razredu (ki običajno ni na voljo, razen če je v vsakem razredu le ena vrednost), z zgornjim izračunom dobimo le približek aritmetične sredine vseh vrednosti.

Primer: frekvenčna porazdelitev, ki prikazuje število članov gospodinjstva.

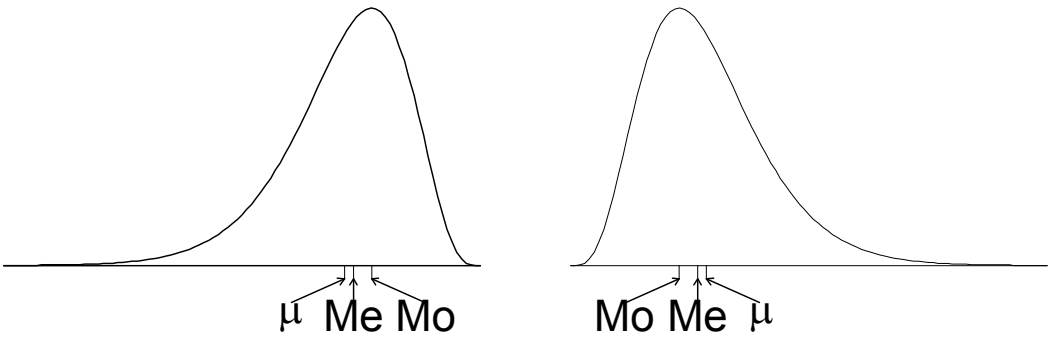
x_i	f_i	$x_i f_i$
1	3	3
2	5	10
3	8	24
4	13	52
5	6	30
6	6	36
7	3	21
8	3	24
9	2	18
10	1	10
	50	228

$$\mu = \frac{1}{N}\sum_{i=1}^{k} f_i x_i = \frac{1}{50} \cdot 228 = 4.56$$

Gospodinjstva imajo v povprečju 4.56 članov gospodinjstva.

2.3.2.4 Odnos med Me in μ

- Za unimodalne, simetrične porazdelitve velja: $\mu = Me = Mo$.
- Za unimodalne porazdelitve, asimetrične v levo, velja: $\mu < Me$.
- Za unimodalne porazdelitve, asimetrične v desno, velja: $\mu > Me$.



2.3.2.5 Katero srednjo vrednost izbrati?

	Modus	**Mediana**	**Aritmetična sredina**
Nominalna	+	-	-
Ordinalna	+	+	-
Intervalna	+	+	+
Razmernostna	+	+	+

Drugi kriteriji:

Modus:

- Če imamo spremenljivko, ki je samo nominalna.
- Če je porazdelitev bimodalna ali polimodalna.

Mediana:

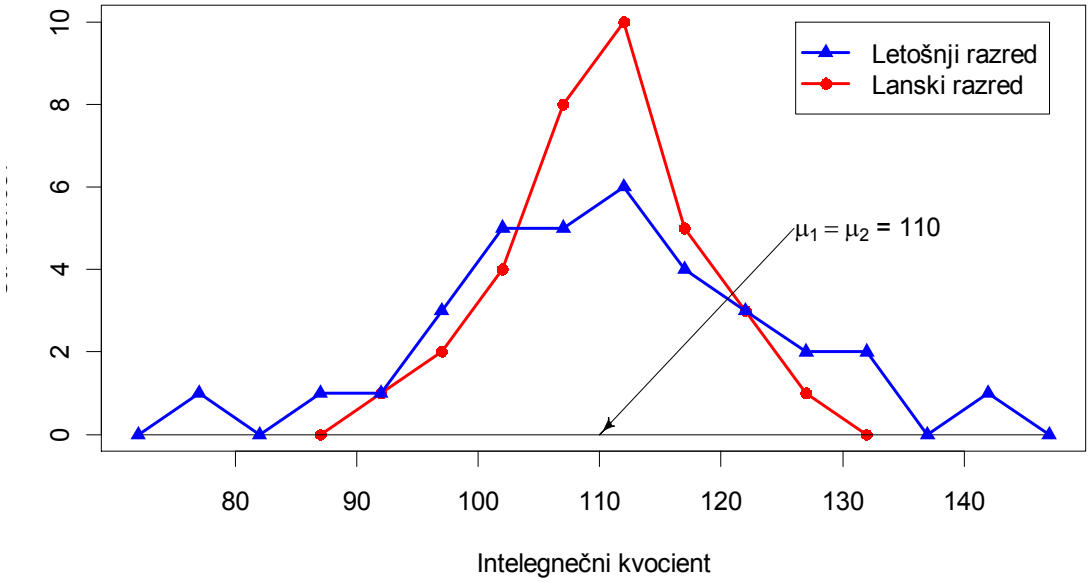
- Če imamo spremenljivko, ki je samo ordinalna.
- Če kakšna enota močno izstopa od ostalih enot oz. je porazdelitev izrazito asimetrična (Pri izračunu mediane ekstremnih vrednosti namreč ne upoštevamo.)

Aritmetična sredina:

- Če je spremenljivka številska in približno normalno porazdeljena.
- Če bomo srednjo vrednost potrebovali kot osnovo za nadaljnje izračune.

2.3.3 Mere variabilnosti

Primer: IK za 2 razreda učencev



Im. tudi mere razpršenosti.

Variabilnost (razpršenost, variacija) govori o tem:

- koliko so podatki variabilni (koliko so vrednosti različne med seboj),
- koliko se vrednosti odklanjajo od srednje vrednosti,
- koliko se vrednosti razlikujejo od srednje vrednosti.

2.3.3.1 Absolutne mere variabilnosti

Variacijski razmik $R = x_{\max} - x_{\min}$ x_{max} ... največja vrednost

x_{min} ... najmanjša vrednost

- Razlika med največjo in najmanjšo vrednostjo.
- Večja kot je variabilnost, večji je variacijski razmik.

Kvartilni odklon $Q = \frac{Q_3 - Q_1}{2}$, kjer sta Q_3 in Q_1 kvartila.

- Polovica razdalje med prvim in tretjim kvartilom.
- Meri variabilnost okoli mediane.
- Večja kot je variabilnost, večji je kvartilni odklon.

Povprečni absolutni odklon od *Me* in μ

$$AD_{\mu} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \mu|$$

$$AD_{Me} = \frac{1}{N}\sum_{i=1}^{N}|x_i - Me|$$

x_i ... vrednost i-te enote
N ... število enot

$$AD_{\mu} = \frac{1}{N}\sum_{i=i}^{k}|x_i - \mu| \cdot f_i$$

$$AD_{Me} = \frac{1}{N}\sum_{i=i}^{k}|x_i - Me| \cdot f_i$$

Če so podatki urejeni v frekvenčno porazdelitev ...
x_i ... reprezentativna vrednost (običajno sredina) *i*-tega razreda spremenljivke X
k ... število vseh razredov v frekvenčni porazdelitvi
f_i ... frekvenca za *i*-ti razred

Pozor: V primeru, ko vse vrednosti v razredu niso enake reprezentativni vrednosti razreda (kar se zgodi le, če je v vsakem razredu le ena vrednost, ki je enaka reprezentativni), z zgornjim izračunom dobimo le približek absolutnega odklona od mediane oz. aritmetične sredine.

- Povprečni odklon vrednosti od srednje vrednosti (*Me* oz. μ).
- Meri variabilnost okoli aritmetične sredine (μ) oz. mediane (*Me*).
- Večja kot je variabilnost, večji je povprečni absolutni odklon – bolj se posamezne vrednosti odklanjajo od srednje vrednosti.

Varianca in standardni odklon

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2$$

$$\sigma = \sqrt{\sigma^2}$$

σ^2 ... varianca
σ ... standardni odklon
μ ... aritmetična sredina
N ... št. enot
x_i ... vrednost sprem. X za *i*-to enoto
$(x_i-\mu)$ ... odklon od aritmetične sredine za *i*-to enoto

- Standarden (tipičen, običajen) odklon vrednosti od aritmetične sredine.
- Meri variabilnost okoli aritmetične sredine.
- Večja kot je variabilnost, večji je standardni odklon – bolj se posamezne vrednosti odklanjajo (razlikujejo) od aritmetične sredine.

Varianca in standardni odklon – če so podatki urejeni v frekvenčno porazdelitev

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \mu)^2 f_i$$

$$\sigma = \sqrt{\sigma^2}$$

σ^2... varianca
σ ... standardni odklon
μ ... aritmetična sredina
N ... št. enot
x_i ... reprezentativna vrednost (običajno sredina) *i*-tega razreda spremenljivke X
k ... število razredov
f_i ... frekvenca *i*-tega razreda
$(x_i-\mu)$ ... odklon od aritmetične sredine

Pozor: V primeru, ko vse vrednosti v razredu niso enake reprezentativni vrednosti razreda (kar se zgodi le, če je v vsakem razredu le ena vrednost, ki je enaka reprezentativni), z zgornjim izračunom dobimo le približek variance oz. standardnega odklona.

Primer 1
Spremenljivka *X:* število ur gledanja TV na teden.
Enota: oseba.

x_i	$x_i-\mu$	$(x_i-\mu)^2$
10	-5	25
12	-3	9
15	0	0
18	3	9
20	5	25
Σ75	Σ 0	Σ 68

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{75}{5} = 15$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2 = \frac{68}{5} = 13{,}6$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{13{,}6} = 3{,}69$$

Opazovane osebe v povprečju gledajo TV 15 ur na teden. Standardni odklon pa je 3,69 ure, kar pomeni, da se število ur gledanja TV pri posameznikih od povprečja standardno odklanja za slabe 4 ure.

Primer 2
Spremenljivka *X*: število članov gospodinjstva.
Enota: 1 gospodinjstvo.

x_i	f_i	$x_i \cdot f_i$	$x_i-\mu$	$(x_i-\mu)^2$	$(x_i-\mu)^2 f_i$
1	3	3	-3,56	12,67	38,02
2	5	10	-2,56	6,55	32,77
3	8	24	-1,56	2,43	19,47
4	13	52	-0,56	0,31	4,08
5	6	30	0,44	0,19	1,16
6	6	36	1,44	2,07	12,44
7	3	21	2,44	5,95	17,86
8	3	24	3,44	11,83	35,50
9	2	18	4,44	19,71	39,43
10	1	10	5,44	29,59	29,59
	Σ50	Σ228			Σ230,32

Gospodinjstva imajo v povprečju 4.56 članov gospodinjstva. Število članov posameznih gospodinjstev pa se od povprečja standardno odklanja za 2,15 člana.

$$\mu = \frac{1}{N}\sum_{i=1}^{k} f_i x_i = \frac{1}{50} \cdot 228 = 4.56$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2 f_i = \frac{230{,}32}{50} = 4{,}61$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{4{,}61} = 2{,}15$$

2.3.3.2 Relativne mere variabilnosti

Absolutne mere variabilnosti lahko le redko primerjamo med seboj. Zato računamo relativne mere variabilnosti, ki so absolutne mere, deljene z ustrezno srednjo vrednostjo.
POZOR: Relativne mere variabilnosti lahko uporabljamo le pri razmernostnih spremenljivkah

$$\text{relativna mera} = \frac{\text{absolutna mera}}{\text{srednja vrednost}}$$

Kdaj jih uporabljamo?

- Če želimo primerjati dve porazdelitvi z zelo različnima srednjima vrednostima.
- Če želimo primerjati dve spremenljivki, kjer so uporabljene različne merske enote.

Relativni variacijski razmik: $\frac{(x_{\max} - x_{\min}) \cdot 2}{x_{\max} + x_{\min}}$

Relativni kvartilni odklon: $\frac{Q_3 - Q_1}{2 \cdot Me}$

Relativni povprečni absolutni odklon od mediane oz. aritm. sredine: $\frac{AD_{Me}}{Me}$ $\frac{AD_{\mu}}{\mu}$

Relativni standardni odklon – **koeficient variacije**: $KV = \frac{\sigma}{\mu}$

Primer 1
V neki raziskavi so ocenili, da je povprečno število ur gledanje televizije na teden za ženske $\mu_Z=10$ in za moške $\mu_M=13$. Standardna odklona pa sta bila enaka, in sicer $\sigma_Z = \sigma_M = 6$. V katerem primeru so – relativno gledano – razlike med osebami večje? Kdo se bolj razlikuje med seboj, moški ali ženske?

$$KV_Z = \frac{\sigma_Z}{\mu_Z} = \frac{6}{10} = 0{,}6 \qquad KV_M = \frac{\sigma_M}{\mu_M} = \frac{6}{13} = 0{,}46$$

Podatki kažejo, da v povprečju moški gledajo za 3 ure več televizijo na teden kot ženske. Razlike v gledanju pa so med ženskami večje kot med moškimi, ker je relativna razpršenost pri ženskah večja.

Primer 2
Na izpitu iz Osnov programiranja je bilo povprečno število točk (od 100 možnih) 50, standardni odklon pa 10. Pri izpitu iz Statistike pa je bilo povprečno število točk 16 (od 30 možnih), standardni odklon pa 4.
V katerem primeru so razlike med točkami večje?

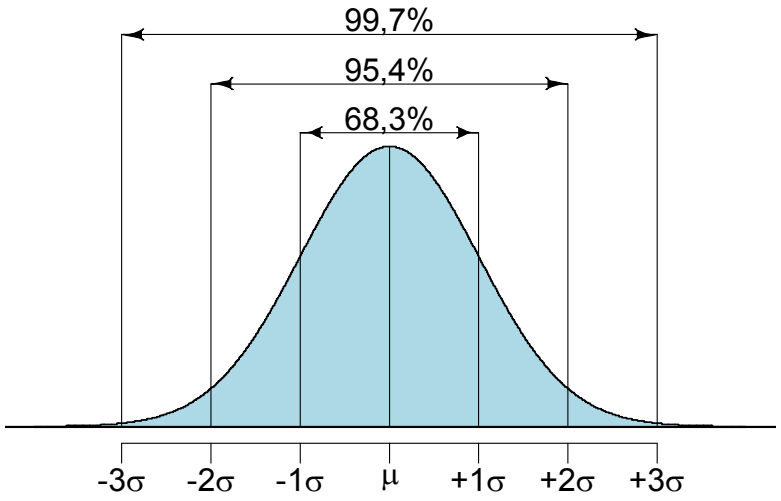
$$KV_{OP} = \frac{\sigma_{OP}}{\mu_{OP}} = \frac{10}{50} = 0{,}2 \qquad KV_S = \frac{\sigma_S}{\mu_S} = \frac{4}{16} = 0{,}25$$

Razlike med doseženimi točkami so večje pri izpitu iz Statistike.

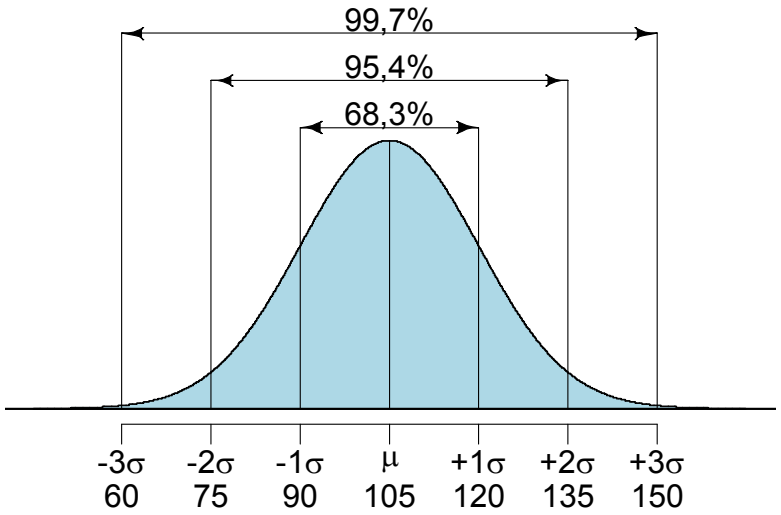
2.3.3.3 Variabilnost pri normalni porazdelitvi

Denimo, da se spremenljivka X porazdeljuje normalnoz aritmetično sredino μ in standardnim odklonom σ. Tedaj velja, da v razmiku:

- $[\mu - \sigma; \mu + \sigma]$ leži 68.3% enot,
- $[\mu - 2\sigma; \mu + 2\sigma]$ leži 95,4% enot,
- $[\mu - 3\sigma; \mu + 3\sigma]$ leži 99,7% enot.



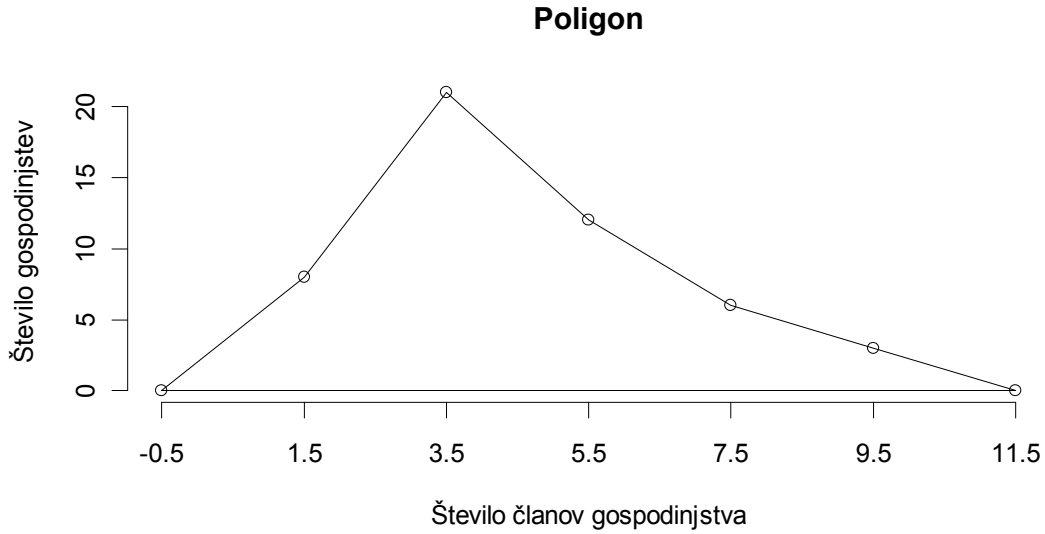
Primer
Denimo, da se inteligenčni kvocient na populaciji študentov porazdeljuje približno normalno z aritmetično sredino $\mu=105$ in standardnim odklonom $\sigma=15$.
Potem vemo, da ima 95,4% študentov inteligenčni kvocient v intervalu [75; 135].
$[\mu - 2\sigma; \mu + 2\sigma]=[105 - 2 \cdot 15; 105 + 2 \cdot 15]=[75; 135]$

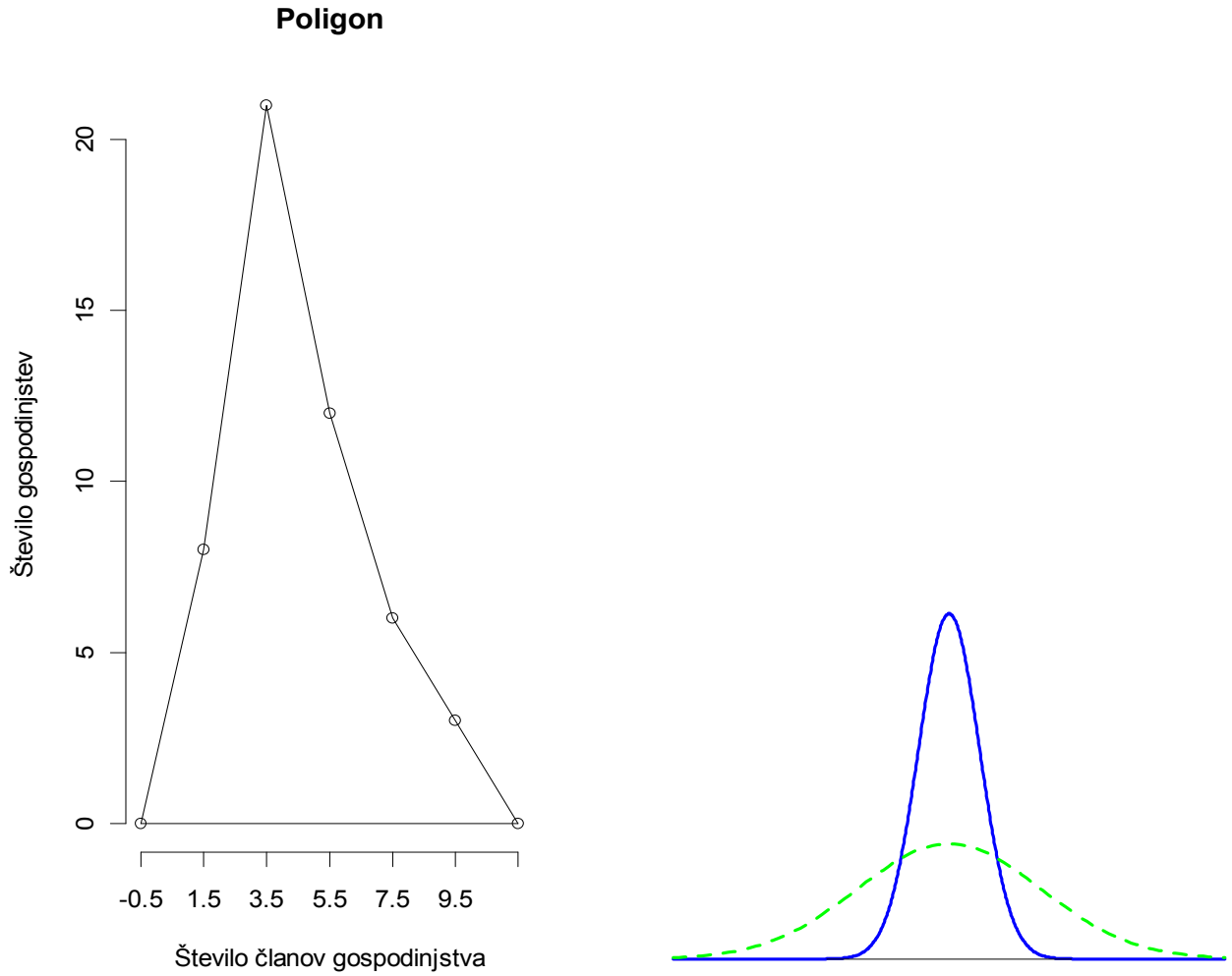


2.3.4 Mere asimetrije in sploščenosti

Če je spremenljivka približno normalno porazdeljena, potem jo statistični karakteristiki aritmetična sredina in standardni odklon zelo dobro opisujeta. V primeru unimodalne

porazdelitve spremenljivke, ki pa je bolj ali manj asimetrična in bolj ali manj sploščena (koničasta), pa je potrebno izračunati še stopnjo asimetrije in sploščenosti (koničavosti). To lahko na več načinov merimo s koeficienti asimetrije in sploščenosti.





Na katerem poligonu in na kateri krivulji (modri-polni ali zeleni-črtkani) je sploščenost večja?

Koeficienta asimetrije in sploščenosti s centralnimi momenti

Več različnih mer, npr. koeficient asimetrije g_1 in koeficient sploščenosti g_2 (razvil ju je Karl Pearson), izračunana s pomočjo t. im. centralnih momentov.

l-ti centralni moment je: $m_l = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^l$

Gre za razlike med posameznimi vrednostmi in aritmetično sredino na *l*-to potenco. Meri asimetrije in sploščenosti torej upoštevata odklone vrednosti od srednje vrednosti.

$m_1=0$ ... ker je seštevek vseh odklonov (na 1. potenco) enak 0.
$m_2=\sigma^2$ ... ker odklone kvadriramo, je to ravno varianca.

2.3.4.1 Koeficient asimetrije (angl. skewness)

Porazdelitev spremenljivke je lahko simetrična ali asimetrična.

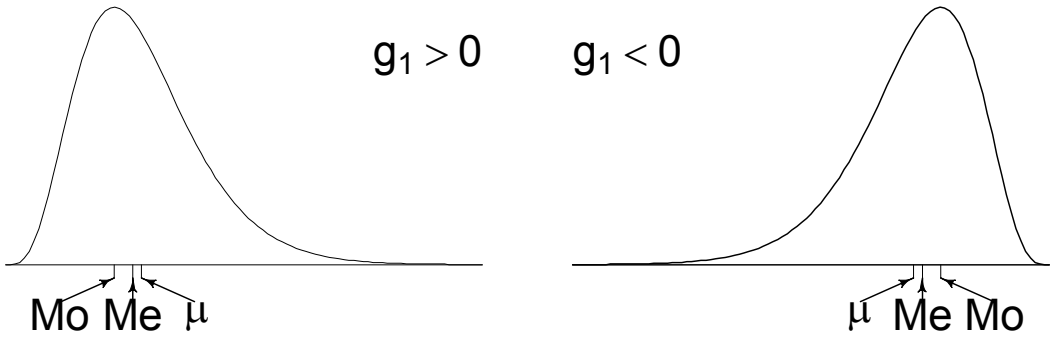
Simetrična p. - vrednosti se enako odklanjajo od srednje vrednosti navzdol in navzgor.

Asimetrična p. v levo – če se rep porazdelitve vleče v levo stran, v negativno smer. Večina enot ima visoke vrednosti in malo enot ima (ekstremno) nizke vrednosti. Aritm. sredina je manjša od mediane, ker nizke vrednosti zmanjšujejo aritmetično sredino.

Asimetrična p. v desno - če se rep porazdelitve vleče v desno stran, v pozitivno smer. Večina enot ima majhne vrednosti in malo enot ima (ekstremno) visoke vrednosti. Aritm. sredina je večja od mediane, ker visoke vrednosti povečujejo aritmetično sredino.

$$g_1 = \frac{m_3}{\sqrt{m_2}^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}\right)^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\left(\sqrt{\sigma^2}\right)^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3}$$

g1 > 0 ... asimetrija v desno
g1 = 0 ... simetrična
g1 < 0 ... asimetrična v levo



2.3.4.2 Koeficient sploščenosti (angl. kurtosis)

Porazdelitev spremenljivke je koničasta, sploščena ali normalno sploščena/koničasta.

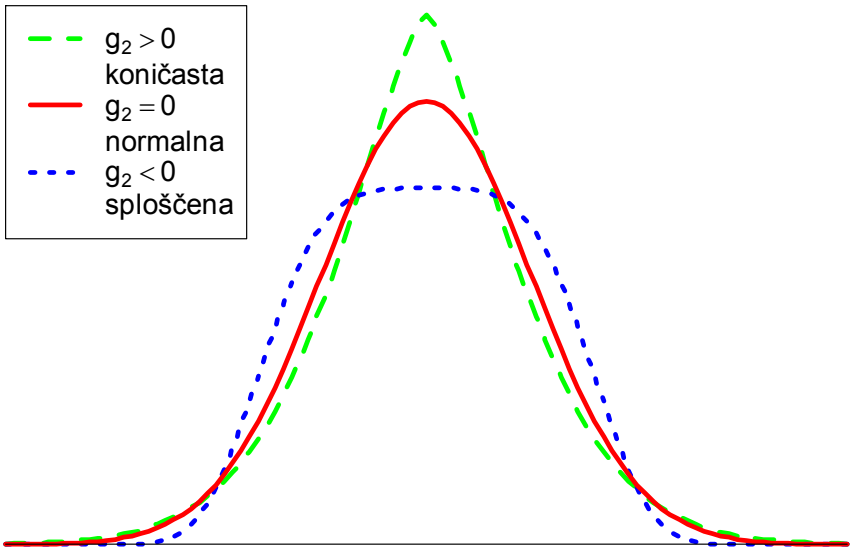
Koničasta p. – bolj koničasta od normalne porazdelitve. Za porazdelitev sta značilna daljša repa in ožji osrednji del. Z naraščanjem vrednosti frekvenca zelo počasi narašča do določene vrednosti, ko začne naenkrat hitro naraščati in hitro doseže vrh. Z nadaljnjimi vrednostmi pa frekvenca najprej hitro upade in nato počasi upada do ekstremno visokih vrednosti.

Sploščena p. – bolj sploščena od normalne porazdelitve. Za porazdelitev sta značilna krajša repa ter debelejši osrednji del. Frekvenca začne naraščati že pri nižjih vrednostih in z višjimi vrednostmi enakomerno narašča, dokler ne doseže vrha. Nato pa z višjimi vrednostmi počasi enakomerno upada.

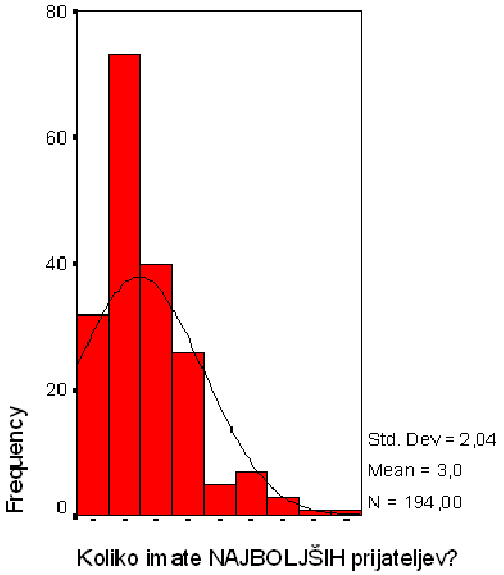
Zelo ekstremne vrednosti in/ali vrednosti čisto na sredini se pojavljajo pri koničasti porazdelitvi pogosteje kot pri sploščeni. Vrednosti med sredino in ekstremnimi vrednostmi pa se pogosteje pojavljajo pri sploščeni porazdelitvi.

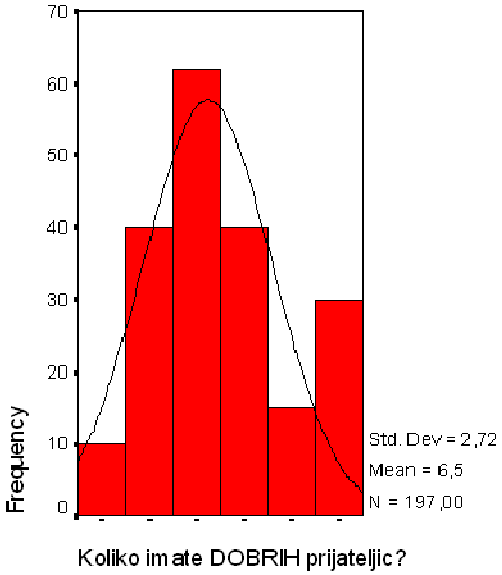
Lahko tudi rečemo, da je variabilnost pri sploščeni porazdelitvi predvsem posledica velikega števila srednje-velikih razlik (med vrednostmi), pri koničasti pa manjšega števila zelo velikih razlik.

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2)^2} - 3 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{(\sigma^2)^2} - 3 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\sigma^4} - 3$$



Primer: anketa med študenti FDV (2002)





Koef. asimetrije 2.95 ... asimetrija v desno
Koef. sploščenosti 5.97 ... koničasto

Koef. asimetrije 0.21 ... simetrična
Koef. sploščenosti -0.74 ... normalno sploščena

2.3.5 Standardizacija

Postopek standardizacije

Vsaki vrednosti x_i spremenljivke X odštejemo njeno aritmetično sredino μ_x in delimo z njenim standardnim odklonom σ_x:

$$z_i = \frac{x_i - \mu_X}{\sigma_X}$$

Vrednosti z_i imenujemo standardizirane vrednosti. Spremenljivko Z, ki ima vrednosti z_i, pa imenujemo standardizirana spremenljivka Z.

Standardizirana vrednost z_i za vrednost x_i predstavlja relativni odklon od aritmetične sredine.

Vrednosti različnih spremenljivk v splošnem niso primerljive. Če pa spremenljivki standardiziramo, lahko primerjamo njihove standardizirane vrednosti.

Značilnosti standardizirane spremenljivke in njenih vrednosti

1. Standardizirana spremenljivka Z ima aritmetično sredino enako 0 (μ_Z=0) in standardni odklon enak 1 (σ_Z=1).

Dokaz (izpeljava):

$$\mu_Z = \frac{1}{N}\sum_{i=1}^{N} z_i = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i - \mu_X}{\sigma_X} = \frac{1}{N}\cdot\frac{1}{\sigma_X}\cdot\sum_{i=1}^{N}(x_i - \mu_X) = \frac{1}{N}\cdot\frac{1}{\sigma_X}\cdot 0 = 0$$

$$\sigma_Z^2 = \frac{1}{N}\sum_{i=1}^{N}(z_i - \mu_Z)^2 = \frac{1}{N}\sum_{i=1}^{N}(z_i - 0)^2 = \frac{1}{N}\sum_{i=1}^{N} z_i^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^2 = \frac{1}{\sigma_X^2}\cdot\frac{1}{N}\cdot\sum_{i=1}^{N}(x_i - \mu_X)^2 = \frac{1}{\sigma_X^2}\cdot\sigma_X^2 = 1$$

2. Če je spremenljivka *X* porazdeljena približno normalno, standardizirana vrednost določa mesto enote v populaciji:

   - Predznak pove, ali je enota nad (+) ali pod (-) povprečjem.
   - Absolutna vrednost pove, za koliko se vrednost odklanja od aritmetične sredine v relativnem smislu – za koliko standardnih odklonov se odklanja

   Primer: Teža in višina dojenčkov, starih 9 mesecev

   Lucija: 68 cm, 8 kg Matevž: 74 cm, 10.2 kg

   Kateri dojenček je večji glede na ostale otroke istega spola in starosti?

	Deklice	Dečki
Povprečna višina	$\mu = 70$ cm	$\mu = 72$ cm
Standardni odklon	$\sigma = 2.5$ cm	$\sigma = 2.5$ cm
Povprečna teža	$\mu = 8.6$ kg	$\mu = 9.4$ kg
Standardni odklon	$\sigma = 0.8$ kg	$\sigma = 0.8$ kg

   Standardizirana višina $z_L = \frac{x_L - \mu}{\sigma} = \frac{68-70}{2,5} = -0,8$ $z_M = \frac{x_M - \mu}{\sigma} = \frac{74-72}{2,5} = 0,8$

   Standardizirana teža $z_L = \frac{x_L - \mu}{\sigma} = \frac{8-8,6}{0,8} = -0,75$ $z_M = \frac{x_M - \mu}{\sigma} = \frac{10,2-9,4}{0,8} = 1$

   Lucija je manjša in lažja v primerjavi z ostalimi deklicami svoje starosti, medtem ko je Matevž večji in težji v primerjavi z ostalimi dečki svoje starosti. Vendar ne odstopata ekstremno. Npr. Lucija je manjša za 0.8 standardnega odklona, torej spada med 68.3% deklic svoje starosti, ki so najbližje povprečju (ob predpostavki, da se spremenljivka *teža deklic* porazdeljuje normalno).

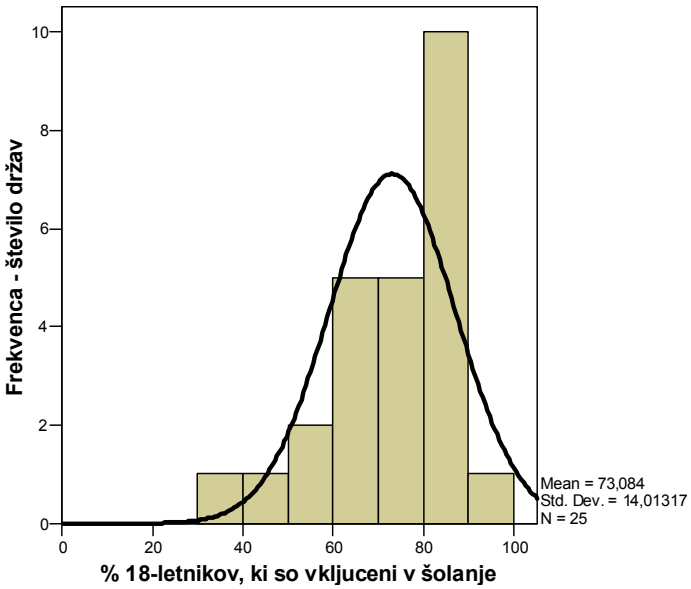
2.4 VAJE

2.4.1 Urejanje in prikazovanje podatkov

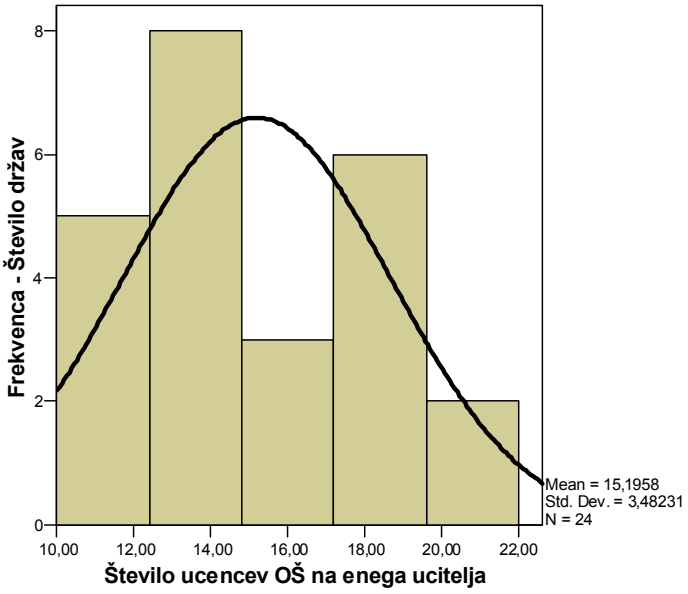
1. Ferligoj (1994): Naloge iz statistike, naloge: 3.1 – 3.6.

2. Opišite naslednja dva grafa:
- razberite, kaj je enota in kaj je spremenljivka,
- določite mersko lestvico spremenljivke,
- opišite obliko porazdelitev (asimetrija, sploščenost/koničavost, modalnost).

Vkljucenost mladih v šolanje v EU državah



Izobraževalni kadri v osnovnem šolstvu



3. V nadaljevanju so grafično predstavljene porazdelitve dveh razmernostnih spremenljivk. Za vsako spremenljivko je narisan histogram za več načinov združevanja vrednosti v razrede. Za vsako spremenljivko odgovorite na naslednja vprašanja:
   - Kaj je spremenljivka?
   - Kaj je enota in kolikšno je število enot?
   - Kako se spreminja oblika porazdelitve, če spreminjamo število (in posledično širino) razredov?

Primer 1: Kadri v osnovnošolskem izobraževanju – Povprečno število učencev na enega učitelja v OŠ za 25 držav članic EU

Število učencev OŠ na 1 učitelja

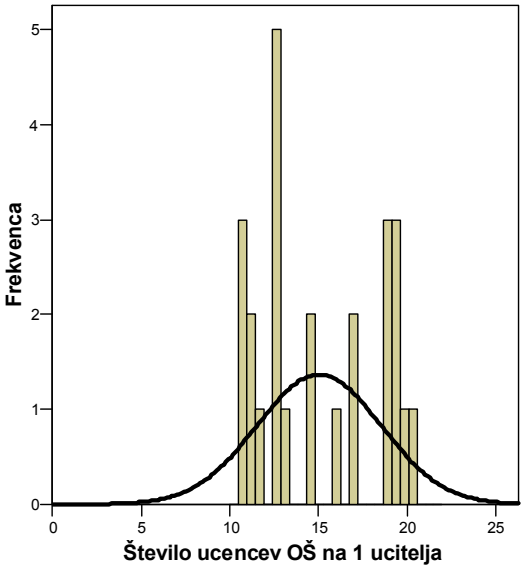
	Frekvenca (Število držav)	Relativna frekvenca
10,60	1	4,0
10,80	1	4,0
10,90	1	4,0
11,00	1	4,0
11,20	1	4,0
11,60	1	4,0
12,40	1	4,0
12,50	2	8,0
12,60	1	4,0
12,80	1	4,0
13,10	1	4,0
14,40	1	4,0
14,60	1	4,0
15,80	1	4,0
16,90	1	4,0
17,00	1	4,0
18,90	2	8,0
19,10	1	4,0
19,40	2	8,0
19,50	1	4,0
19,90	1	4,0
20,10	1	4,0
	25	100,0

Primer 2: Število avtomobilov na 1000 prebivalcev za 25 držav članic EU

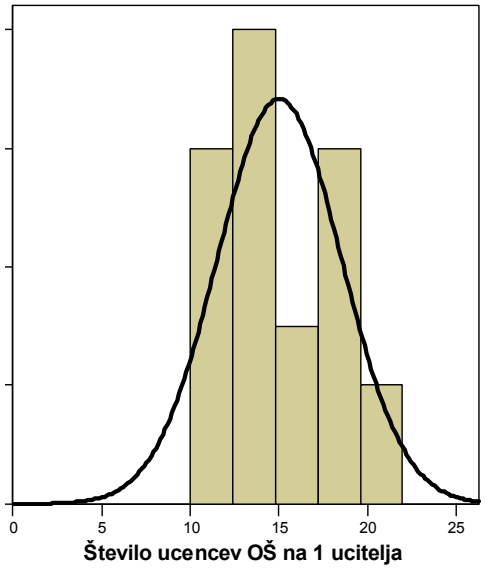
Število osebnih avtomobilov na 1000 prebivalcev

	Frekvenca (Število držav)	Relativna frekvenca
247	1	4,0
259	1	4,0
265	1	4,0
287	1	4,0
295	1	4,0
331	1	4,0
340	1	4,0
351	1	4,0
357	1	4,0
371	1	4,0
405	1	4,0
422	1	4,0
424	1	4,0
447	1	4,0
453	1	4,0
459	1	4,0
460	1	4,0
463	1	4,0
490	1	4,0
495	1	4,0
508	1	4,0
541	1	4,0
558	1	4,0
590	1	4,0
643	1	4,0
	25	100,0

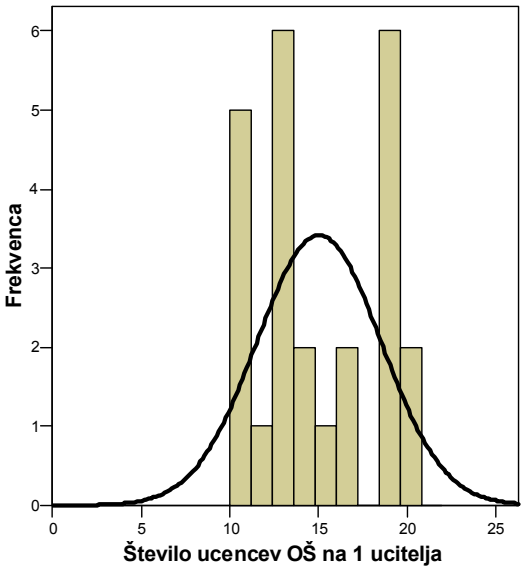
Histogram: Kadri v osnovnošolskem izobraževanju



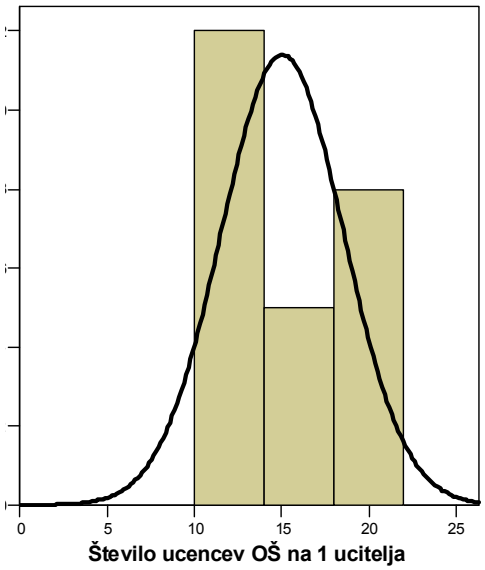
Histogram: Kadri v osnovnošolskem izobraževanju



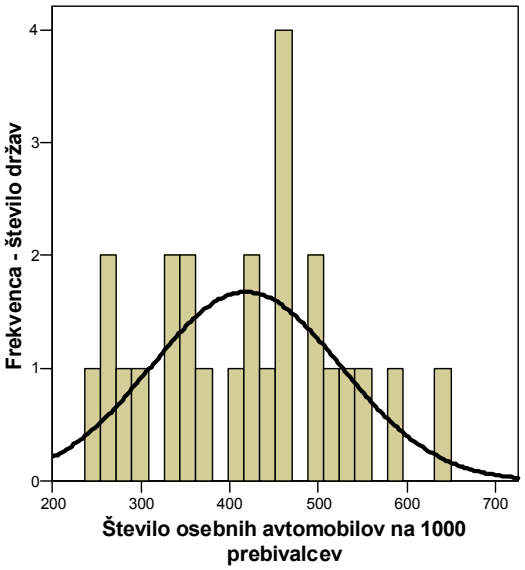
Histogram: Kadri v osnovnošolskem izobraževanju



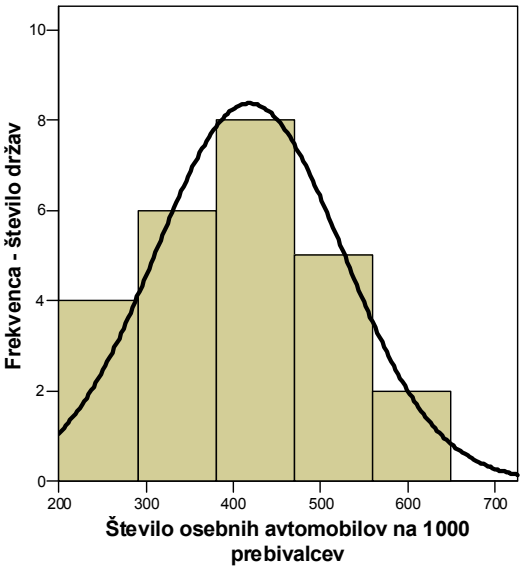
Histogram: Kadri v osnovnošolskem izobraževanju



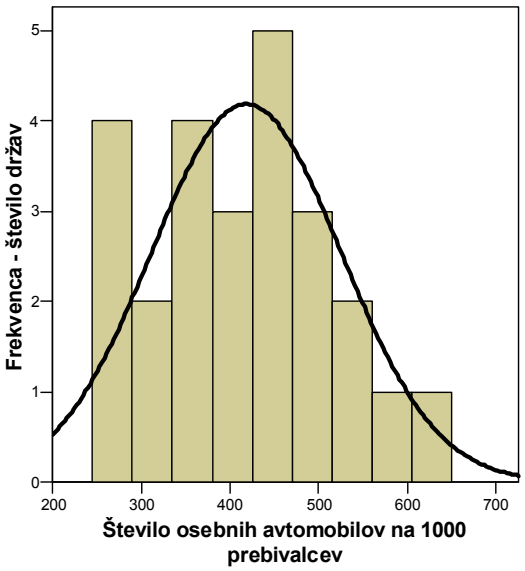
Histogram: število osebnih avtomobilov v državah EU



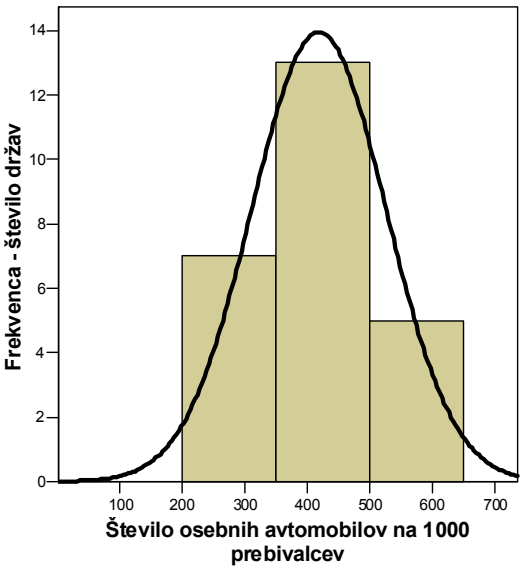
Histogram: število osebnih avtomobilov v državah EU



Histogram: število osebnih avtomobilov v državah EU



Histogram: število osebnih avtomobilov v državah EU



2.4.2 Kvantili

1. Ferligoj (1994): Naloge iz statistike, naloge: 4.1, 4.4.

2. Primer izpitnega vprašanja.
   Podane so starosti 7 študentov: 22, 19, 21, 24, 19, 18, 23. Pr vi kvartil ima vrednost:

a) 18.5
b) 19
c) 19.5
d) 25%

2.4.3 Srednje vrednosti

1. Ferligoj (1994): Naloge iz statistike, naloge: 4.6, 4.7, 4.8, 4.9 (samo modus), 4.10, 4.11, 4.12 (samo modus), 4.13 (a, b samo frekvenčna porazdelitev), 4.17, 4.20.

2. Neka multinacionalna tovarna avtomobilov na leto proizvede naslednje število avtomobilov (v 10.000) v svojih obratih v sedmih državah:
   6, 8, 6, 9, 11, 5, 60
   a) Kaj je spremenljivka in kaj so enote?
   b) Kakšna je merska lestvica spremenljivke?
   c) Narišite histogram, poligon in ogivo.
   d) Kolikšno je celotno število proizvedenih avtomobilov v tej tovarni?
   e) Kolikšni so arimetična sredina, mediana, modus? Označite te tri sredine na
   f) histogramu.
   g) Kaj lahko na osnovi grafov rečete o asimetriji te porazdelitve?
   h) Za neko drugo tovarno, ki izdeluje avtomobile v 10-ih državah, so podatki za letno število proizvedenih avtomobilov (v 10.000 naslednji): 7.8 v povprečju na državo, mediana je 6.5 in modus 5.0. Kakšno je skupno število proizvedenih avtomobilov v 10-ih državah?

3. Da bi videli, ali so upravičene izjave o vzdržljivosti baterij, so na Zvezi potrošnikov testirali vzorec 20-ih baterij. Zabeleženi so podatki o življenjski dobi (v minutah) teh baterij:

58 58,4 59,4 64,2 64,9 65,1 65,4 67,8 68 73,3 74,7 74,9 75,1 75,4 76 76 76,6 76,7 77,6 81,3

   a) Kaj je spremenljivka in kaj so enote?
   b) Kakšna je merska lestvica spremenljivke?
   c) Vrednosti razporedite v frekvenčno porazdelitev z razredi, širokimi 5 minut
   d) Relativno frekvenčno porazdelitev grafično predstavite.
   e) Na osnovi frekvenčne porazdelitve izračunajte modus porazdelitve.
   f) Vrednosti razporedite v frekvenčno porazdelitev s tremi razredi, pri čemer naj bodo srednje vrednosti razredov 60, 70 in 80 minut. Ponovno grafično predstavite porazdelitev in izračunajte modus. Primerjajte rezultat dveh različnih grupiranj.

4. V manjšem podjetju je prejšnji mesec 5 zaposlenih na najnižjem delovnem mestu dobilo 2.500 EUR bruto, dve zaposleni osebi na višjem delovnem mestu 6.000 EUR bruto ter direktor podjetja 25.500 EUR bruto.
   a) Kaj je spremenljivka in kaj so enote?
   b) Kakšna je merska lestvica spremenljivke?
   c) Kolikšna je bila povprečna bruto plača v tem podjetju?
   d) Koliko zaposlenih zasluži manj od povprečja?
   e) Kolikšna je mediana za bruto plačo?
   f) Kolikšen je modus za bruto plačo?
   g) Katera srednja vrednost je najprimernejša za ponazoritev višine bruto plače v tem podjetju?

5. Predstavljene so štiri frekvenčne porazdelitve neke spremenljivke (negrupirana porazdelitev, porazdelitev z razredi širine 2 enoti, z razredi širine 3 enote ter razredi širine 4 enote). Za vsako porazdelitev izračunajte modus in jih primerjajte. Kaj se dogaja, ko vrednosti združujemo v razrede?

Negrupirani podatki		Grupirani podatki d=2		Grupirani podatki d=3		Grupirani podatki d=3	
x_i	f_i	x_i	f_i	x_i	f_i	x_i	f_i
22	1	22-23	2	22-24	2	22-25	3
23	1						
24	0	24-25	1				
25	1			25-27	17		
26	3	26-27	16			26-29	36
27	13						
28	12	28-29	20	28-30	28		
29	8						
30	8	30-31	14			30-33	20
31	6			31-33	12		
32	4	32-33	6				
33	2						
34	1	34-35	2	34-36	2	34-37	2
35	1						
36	0	36-37	0				
37	0						

6. Izračunajte aritmetično sredino za naslednjih pet vrednosti: 3, 7, 8, 12, 15.
   a) Za vsako vrednost izračunajte (pozitiven ali negativen) odklon od aritmetične sredine, t.j. x_i-μ. Izračunajte povprečje teh odklonov.
   b) Zamislite si nek poljuben niz števil. Izračunajte njihovo aritmetično sredino in nato povprečni odklon od aritmetične sredine.
   c) Dokažite, da je za vsak možen niz *n*-tih vrednosti povprečni odklon od aritmetične sredine enak nič.

7. Škotski vic: "Če se Škot preseli iz Škotske v Anglijo, se v obeh regijah zviša povprečni inteligenčni kvocient". V katerem primeru je to mogoče (ali je nemogoče)? (Kakšno mora biti povprečje v vsaki regiji in kakšna mora biti vrednost Škota v primerjavi s povprečjem?)

2.4.4 Mere variabilnosti, asimetrije, sploščenosti, standardizacija

1. Ferligoj (1994): Naloge iz statistike: 5.1, 5.2, 5.7, 5.9, 5.14.

2. Podani so podatki o višini inteligenčnega kvocienta za nek šolski razred. Odgovorite na spodnja vprašanja:

77	87	92	97	97	97	102	102	102	102	102	107	107	107	107	107	112
112	112	112	112	112	117	117	117	117	122	122	122	127	127	132	132	142

   a) Kaj je enota analize in kaj spremenljivka? Kakšna je merska lestvica spremenljivke?
   b) Za dane podatke izračunajte vse absolutne in relativne mere variabilnosti?
   c) Interpretirajte izračunani standardni odklon.
   d) Podatke uredite v frekvenčno porazdelitev s 5 razredi in jo grafično predstavite s histogramom in poligonom. Ocenite oblike porazdelitve iz grafične predstavitve.
   e) Izračunajte koeficienta asimetrije in sploščenosti s pomočjo centralnih momentov in ju interpretirajte. Ali se rezultat ujema z ugotovitvijo v točki d)?
   f) Kakšen inteligenčni kvocient bi imelo 95.4% vseh učencev, če bi bila porazdelitev popolnoma normalna?

3. Za skupino starejših oseb, obolelih s Parkinsovo boleznijo, je podan podatek o starosti ob nastopu bolezni:

67	68	60	64	68	63
68	70	63	70	68	69
70	69	69	69	69	68
62	70	70	64	66	66
66	69	67	67	70	67

   a) Izračunajte koeficient asimetrije za to porazdelitev in ga interpretirajte.
   b) Izračunajte koeficient sploščenosti za to porazdelitev in ga interpretirajte.
   c) Narišite histogram. Ali grafična predstavitev vrednosti potrjuje zgornje ugotovitve?
   d) Gospod X je imel ob nastopu bolezni 64.2 leti. S pomočjo standardizacije ugotovite, ali je – v primerjavi s povprečjem – za to boleznijo zbolel mlad ali star.

4. Primeri izpitnih vprašanj

Dane so vrednosti: 10, 8, 14, 14, 16, 10. Aritmetična sredina danih vrednosti je 12. Varianca je:

a) 2,83
b) 3,27
c) 0
d) 8

68.3% vseh vrednosti standardizirane normalno porazdeljene slučajne spremenljivke

a) leži na intervalu od -1 do +1
b) leži na intervalu od -3 do +3
c) je samo pozitivnih
d) leži na intervalu od minus do plus neskončno

Če je koeficient asimetrije, izračunan s centralnimi momenti, enak 2.05, je porazdelitev spremenljivke

a) sploščena
b) asimetrična v levo
c) koničasta
d) asimetrična v desno

Če želimo primerjati variabilnost ocen pri dveh predmetih, moramo primerjati

a) normalizirane vrednost
b) koeficienta variacije
c) najmanjši vrednosti
d) asimetrični sredini

Če ima enota standardizirano vrednost 1.3, pomeni, da ima

a) vrednost, ki je večja od aritmetične sredine,
b) vrednost, ki je manjša od aritmetične sredine,
c) vrednost, ki je 1.3-krat večja od aritmetične sredine,
d) koničasto porazdelitev