



Europska unija
Ulaganje u budućnost



SREDNJA ŠKOLA
KRAPINA



MATEMATIKA I JA, NA TI

Skripta iz statistike
1. - 4. razred srednje škole

Ova publikacija izrađena je u okviru projekta **Gimnazija 100+** kojeg je sufinancirala Europska unija iz Europskog socijalnog fonda.



PROJEKT GIMNAZIJA 100+

Korisnik: Srednja škola Krapina

Partner: Srednja škola Pregrada

Vrijednost projekta: 1.117.336,41 kn

Bespovratna sredstva: 1.117.336,41 kuna

Trajanje projekta: 12 mjeseci

U sklopu projekta Gimnazija 100+ želimo unaprijediti ishode učenja u području matematike i prirodoslovlja, omogućiti stjecanje dodatnih kompetencija, osigurati veću uspješnost na ispitima državne mature i uspješan nastavak obrazovanja. Razvijamo nove fakultativne programe iz matematike, kemije, biologije i fizike.

Projekt Gimnazija 100+ sufinancirala je Europska unija iz
Europskog socijalnog fonda.

IMPRESSUM

Pripremili: Renata Spiegl, prof. matematike
Ivana Polanović, mag. educ. matematike
Jasminka Mikša, prof. matematike
dr. sc. Željka Milin Šipuš – vanjska konzultantica

Nakladnik: Srednja škola Krapina,
Šetalište hrvatskog narodnog preporoda 6, 49 000 Krapina

Za nakladnika: Ivica Rozijan, prof., ravnatelj Srednje škole Krapina

Grafičko oblikovanje: Aldini d.o.o., Sesvete

Tisak: Aldini d.o.o., Sesvete

Prvo izdanje, 2016.

Sadržaj ove publikacije isključiva je odgovornost Srednje škole Krapina.

I	UVOD U OSNOVE STATISTIČKOG ISTRAŽIVANJA	1
1.	Pojam, predmet i zadaće statistike	1
2.	Podjela statistike	1
3.	Statistički skup. Obilježje	3
4.	Vrste statističkih obilježja (varijabla) i skala	6
4.1.	Kvalitativne varijable (obilježja)	7
4.2.	Kvantitativne (numeričke) varijable (obilježja)	8
5.	Statistički podaci i njihova svojstva	11
5.1.	Izvori podataka	12
5.2.	Sekundarni podaci	12
5.3.	Primarni podaci	13
5.4.	Načini prikupljanja primarnih podataka	14
5.5.	Priprema podataka za statističku obradu	16
II	UREĐIVANJE I PRIKAZIVANJE PODATAKA	18
1.	Formiranje statističkih nizova, frekvencije	18
2.	Vrste statističkih nizova	28
2.1.	Nizovi kvalitativnih podataka	28
2.2.	Numerički statistički niz	29
2.3.	Vremenski niz	31
3.	Tabeliranje	31
4.	Grafičko prikazivanje kvalitativnih nizova	33
4.1.	Stupčasti dijagram (eng. BarChart)	33
4.2.	Kružni dijagram	36
4.3.	Dijagram točaka (eng. DotPlot)	37
4.4.	Dijagram stablo-list (eng. StemPlot)	37
4.5.	Histogram	38
4.6.	Poligon frekvencija (eng. FrequencyPolygon)	39
4.7.	Kumulanta	40
III	SREDNJE VRIJEDNOSTI STATISTIČKIH NIZOVA	48
1.	Aritmetička sredina (prosjek)	48
2.	Geometrijska i harmonijska sredina	52
3.	Mod. Medijan	56
3.1.	Mod	56
3.2.	Medijan	58
IV	MJERE RASPRŠENOSTI (DISPERZIJE)	67
1.1.	Apsolutne mjere disperzije	67
1.2.	Relativne mjere disperzije	75
2.	Standardizirana vrijednost	77
3.	Pravilo normalne distribucije (Empirijsko pravilo)	79
V	UTJECAJ DODAVANJA ILI UKLANJANJA PODATAKA NA	84
	SREDNJE VRIJEDNOSTI NIZA	84
1.	Netipične vrijednosti	84
2.	Ublažavanje problema utjecaja dodavanja ili uklanjanja podataka na srednje vrijednosti	86
3.	Uspoređivanje srodnih skupova podataka	87
3.1.	Grafičko uspoređivanje srodnih skupova podataka	87
3.2.	Uspoređivanje srednjih vrijednosti i mjera raspršenosti srodnih skupova podataka	90
VI	PRAVAC REGRESIJE	95
VII	STATISTIKA UZ POMOĆ PROGRAMA GEOGEBRA	102
2.	Konstrukcije i naredbe u GeoGebri	103
VIII	PROJEKTA NASTAVA	122
1.1.	Što je projektna nastava?	122
1.2.	Koji su ciljevi projektne nastave?	122
1.3.	Realizacija jednog matematičkog projekta	123
1.4.	Uloga učenika i nastavnika u projektu	124
1.5.	Primjer projektne teme	125
	LITERATURA	126
	PRILOG 1	127
	PRILOG 2	128

I UVOD U OSNOVE STATISTIČKOG ISTRAŽIVANJA

1. Pojam, predmet i zadaće statistike

Statistika je znanstvena metoda prikupljanja, uređivanja, analiziranja i tumačenja raznovrsnih brojčanih podataka i donošenja zaključaka o pojavama i procesima koje ti podaci predaju. Matematička statistika je dio teorije vjerojatnosti koja se bavi matematičkom obradom empirijskih podataka (rezultata mjerenja). Realizacija ciljeva statistike zahtjeva metode koje se nazivaju *statističkim metodama i tehnikama*.

Statističke metode i tehnike primjenjuju se u ekonomiji, zdravstvu, kulturi, prometu itd. U tim područjima statistika nudi izravnu pomoć i potporu procesu odlučivanja.

Predmet statistike kao analitičke metode jesu svojstva pojava i procesa u prirodi i društvu te njihovi međusobni odnosi predloženi podacima dobivenima opažanjem ili mjerenjem u statističkim pokusima.

Zadaca statistike jest da svojim metodama ustanovi i brojčano izrazi svojstva pojava i procesa i njihovih međusobnih odnosa na temelju svojstava njihovih elemenata te pruži valjanu podlogu za statističko zaključivanje.

Osnova za bilo kakvu statističku analizu jesu *podaci*.

2. Podjela statistike

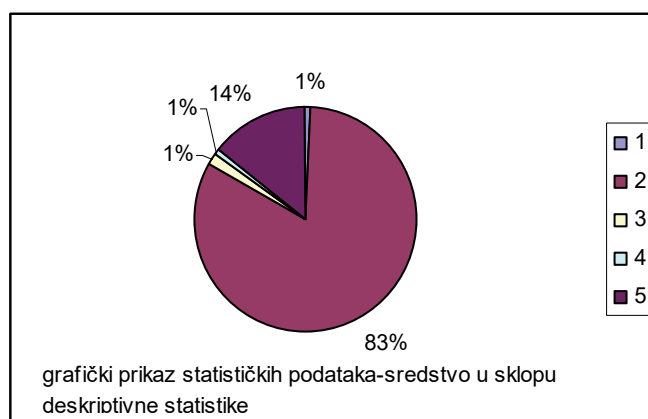
S obzirom na temeljna obilježja, statistika kao metoda analize pojava i procesa dijeli se na deskriptivnu (opisnu) i inferencijalnu statistiku (inferencija je zaključivanje metodom indukcije ili dedukcije).

Deskriptivna statistika jest skupina metoda s pomoću kojih se utvrđuju činjenice o pojavama. Pri tom se rabe postupci uređivanja, tabličnog i grafičkog prikazivanja te utvrđuju različiti brojčani pokazatelji. Zaključci doneseni u sklopu deskriptivne statistike obavijesti su koje se odnose isključivo na pojavu za koju su dani podaci i ne služe poopćivanju.

Rezultati deskriptivne statistike najčešće su dani u obliku tablica, grafičkih prikaza i jednostavnih brojčanih pokazatelja.

Primjer 2.1.

Oblici vlasništva	Broj aktivnih pravnih osoba	Struktura (u %)
državno	1031	0,78
privatno	109121	82,37
zadružno	1932	1,46
mješovito	1408	1,06
nema vlasnika	18981	14,33
ukupno	132473	100,00



Inferencijalna statistika obuhvaća skupinu metoda s pomoću kojih se donose sudovi o obilježjima cjeline-statističkog skupa, na temelju podataka iz slučajnih uzoraka-podskupova statističkog skupa.

Primjer 2.2.

Ispituje se gledanost neke TV emisije kako bi se dobila obavijest o postotku građana koji gledaju tu emisiju. Kako je financijski zahtjevno i relativno dugotrajno ispitati sve građane, umjesto anketiranja svih građana navedeni se postotak procjenjuje anketiranjem određenog manjeg broja građana, odnosno uzorka. Znači, na temelju dijela (uzorka) zaključuje se o cjelini (statističkom skupu).

3. Statistički skup. Obilježje

Statistički skup (ili populacija) je skup elemenata (osobe, predmeti, vozila, države i slično) koji se podvrgavaju promatranju, odnosno mjerenju nekih veličina. To je skup istovrsnih elemenata. Npr. skup svih učenika jednog razreda, skup svih zaposlenika u srednjoj školi, skup država u Europskoj uniji itd.

Ako se pritom mjeri samo jedna veličina, za promatrani statistički skup kažemo da je *jednodimenzionalan*. Ako se promatraju i mjere dvije veličine, tada je skup *dvodimenzionalan* a rezultati mjerenja su parovi brojeva. Analogno se definira *n-dimenzionalan* statistički skup.

Statistički skup mora biti definiran *pojmovno, prostorno i vremenski*.

Pojmovnim definiranjem statističkog skupa utvrđuje se pripadnost tom skupu s obzirom na pojmovno određenje svake pojedine jedinice (elementa). Npr. kod statističkog skupa učenika element tog skupa je osoba upisana u odgovarajuću srednju školu. Pojmovno određenje je i starost stanovnika, vrste bolesti, turist itd.

Prostornim definiranjem statističkog skupa utvrđuje se prostor kojem mora pripadati element da bi bio uključen u statistički skup. Na primjer, učenici Srednje škole Krapina, stanovnici Pregrade, bolesti u dječjem vrtiću u Zaboku, turisti u RH.

Vremenskim definiranjem statističkog skupa utvrđuje se vrijeme za koje su vezani elementi statističkog skupa. Vrijeme se može odnositi na vremenski trenutak ili razdoblje-vremenski interval. Na primjer, ako se promatraju učenici upisani 2015./2016. godine na dan 1.9.2015., statistički skup definiran je u trenutku, dok je kod učenika koji su pristupili polaganju državne mature školske godine 2014./2015. statistički skup definiran u vremenskom intervalu (ljetni i jesenski rok).

Primjer 3.1.

Istražuju se obilježja turista koji su boravili u Republici Hrvatskoj u 2012. godini. Statistički skup čine turisti. Jedinica skupa je turist. Opseg skupa je broj turista u 2012. Skup je konačan jer ima konačan broj jedinica (turista).

Pojmovnom definicijom utvrđuje se pojam turista (turist je svaka osoba koja u mjestu izvan svog prebivališta provede najmanje jednu noć u nekom smještaju za goste radi odmora, poslova itd.). *Prostornom* definicijom određuje se područje na kojem su boravili turisti

(Republika Hrvatska), a *vremenska* definicija odnosi se na vremenski interval, tj. 2012. godinu.

Predmet statističke analize nisu elementi statističkog skupa već njihova različita svojstva. Skup podataka o promatranom svojstvu za svaki element predstavlja **osnovni skup ili populaciju**.

Statističko obilježje je ono svojstvo elemenata populacije po kojem se ti elementi razlikuju ili nalikuju. Elementima statističkog skupa može se pridružiti jedno ili više obilježja.

Npr. spol, stručna sprema, ocjena na ispitu, površina stana. **Vrijednosti obilježja** su stanja koja to obilježje može imati.

Primjer 3.2.

Prošli tjedan u Zavodu za transfuziju 50 je osoba dobrovoljno dalo krv. Svakom je davatelju određena krvna grupa i dobiven je ovaj niz podataka:

A A B 0 B AB 0 0 0 0 B B 0 A 0 A 0 0 0 0 A 0 0 0 B
0 0 0 A B 0 0 0 B 0 0 0 B AB AB 0 0 B A A 0 0 A A A

Ovo su tzv. primarni podaci, oni još nisu sređeni. Prva faza sređivanja je prikazivanje u tablici:

tip krvne grupe	broj osoba koje imaju taj tip krvne grupe(frekvencija)
0	27
A	11
B	9
AB	3

Svojstvo koje nas je zanimalo bio je tip krvne grupe. To svojstvo je **obilježje** ili **varijabla**.

Vrijednost obilježja je tip krvne grupe: A,B, AB ili 0 (četiri vrijednosti). **Populacija** je skup svih dobrovoljnih darivatelja krvi koji su prošli tjedan dali krv. Brojevi u 2.stupcu nazivaju se **frekvencije**, a kazuju nam koliko puta se u skupu podataka pojavljuje neka vrijednost. Dobivena tablica zove se **tablica frekvencija**.

Zadaci 1., 2., 3.

1. Napiši primjer statističkog skupa koji je:
 - a. pojmovno definiran
 - b. prostorno definiran
 - c. vremenski definiran
2. Anketiraj učenike u svom razredu o tome koji im je dan u tjednu najdraži. Podatke prikaži u tablici frekvencija. Što je obilježje, a što su vrijednosti tog obilježja? Koji dan im je najdraži? Koliko je učenika anketirano?
3. Učenici 3a razreda su na ispitu iz fizike ostvarili ove ocijene:

5 4 4 4 3 4 2 2 1 1 2 3 2 5 1 3 3 1 4 5 2 2

 Napravi tablicu frekvencija. Što je obilježje, što su vrijednosti tog obilježja?
4. Učenici svih prvih razreda škole odgovarali su na pitanje: "Koliko imaš kućnih ljubimaca?". Dobiveni su ovi odgovori:

0	0	0	2	1	3	1	1	1	0	3	2	1	1	1
0	0	1	1	1	2	4	1	1	2	2	1	1	0	0
1	0	1	1	0	2	2	3	2	1	1	2	1	0	1
4	1	0	0	1	1	1	0	1	2	2	1	1	1	1

- a. Što je obilježje, a što su vrijednosti tog obilježja?
 - b. Napravi tablicu frekvencija.
 - c. Koliko djece nema ni jednog kućnog ljubimca?
 - d. Koliko djece ima bar jednog kućnog ljubimca?
 - e. Koliki dio djece ima više od 2 ljubimca?
5. a. Kako se definira statistika?
 - b. Što je predmet a što zadaća statistike?
 - c. Kako se dijeli statistika?
 - d. Navedi temeljna obilježja deskriptivne statistike.

Potražite primjere na internetskim stranicama tvrtki koje sami odaberete.
 - e. Odaberite neku konkretnu pojavu i objasnite u čemu bi se sastojala njezina analiza u sklopu inferencijalne statistike.

Pronađite primjere tvrtki kojima su predmet poslovanja anketna istraživanja.

Rješenja

1. a. Skup gimnazijalaca SŠ Krapina:3 modaliteta(jezična,opća,mat.)
 b. Tramvajski promet u RH: Osijek, Zagreb
 c. Broj učenika upisano u 1.razred jezične gimnazije u šk.g.2015./16.
2. Obilježje je dan u tjednu a vrijednosti su pon, ut, sri, čet, pet, sub, ned
3. Obilježje: ocjena na ispitu; vrijednosti su 1,2,3,4,5.
4. Obilježje je "broj kućnih ljubimaca", a vrijednosti su 0,1,2,3,4.

4. Vrste statističkih obilježja (varijabla) i skala

Svojstva elemenata statističkog skupa se mjere. Mjerenjem se smatra pridruživanje brojeva ili oznaka jedinicama statističkog skupa prema **mjernim skalama**. Svaka mjerna skala sastoji se od oblika statističke varijable. Osnova za bilo kakvu statističku analizu jesu podatci.

Postoji više vrsta statističkih varijabli, odnosno statističkih obilježja. Razlikovanje vrsta temelji se na mogućnosti provođenja brojčanih radnji nad njihovim oblicima.

VRSTA OBILJEŽJA	PODVRSTA		PRIMJERI
KVALITATIVNO	nominalno- atributivno		zanimanje, spol, nacionalnost, bračno stanje, vrsta gospodarske djelatnosti, vrsta robe, marka automobila
	nominalno-geografsko (prostorno)		mjesto boravka, rođenja, zemlja podrijetla robe, mjesto registracije automobila
	redosljedno (ordinalno)		školska sprema, ocjena na ispitu, kategorija kakvoće robe, stupanj zadovoljstva uslugom
KVANTITATIVNO (numeričko)	intervalno		temperaturna skala u °C
	omjerno	diskretno (prekidno)	broj sagrađenih stanova, broj zaposlenih, dnevna prodaja televizora, broj turista
		kontinuirano (neprekidno)	trajanje telefonskih razgovora, visina osobe, plaća zaposlenika u kunama, starost osobe

4.1. Kvalitativne varijable (obilježja)

Skupinu kvalitativnih varijabli čine *nominalne i redoslijedne varijable* pa tako imamo i nominalne i ordinalne (redoslijedne) skale.

Nominalna skala

- dana je u obliku nenumeričkog skupa naziva po kojima se elementi statističkog skupa međusobno razlikuju tj. dana je opisno (npr. strani jezik koji su studenti učili u srednjoj školi, mjesto rođenja neke osobe, spol, nacionalnost, odgovor u upitniku da, ne, ne znam)
- modalitetima nominalnog obilježja mogu se pridružiti brojevi ali nad njima nisu dopuštene nikakve brojčane operacije (npr. označi li se ženski spol brojem 0, a muški 1, očito nad tim brojevima nema smisla provoditi računske radnje. Oni služe kao identifikatori jednako, različito)

Primjer 4.1.1. (nominalno obilježje)

U srednjoj školi zapisivali su se učenici jednog razreda koji bi strani jezik željeli učiti. Dobiveni su ovakvi rezultati:

strani jezik	broj učenika koji žele učiti taj strani jezik
engleski jezik	22
njemački jezik	9
talijanski jezik	5
španjolski jezik	3

Ordinalna (redoslijedna) skala

- pridružuje brojeve, slovne oznake ili simbole elementima statističkog skupa prema intenzitetu nekog svojstva (npr. ocjena na ispitu - brojčano od 1 do 5, ekonomska razvijenost - nerazvijene zemlje, zemlje u razvoju, razvijene zemlje)
- također nad njima nisu dopuštene nikakve brojčane operacije (+, -, :,), ali modaliteti se mogu urediti prema nekom redoslijedu (jednako, veće, manje)

Ako učenik na ispitu dobije vrlo dobar razlikuje se od učenika koji je dobio dovoljan i imat će različita mjesta na ordinalnoj skali, ali ne možemo reći da je ocjena vrlo dobar dva puta veća od ocjene dovoljan.

Primjer 4.1.2. (redosljedno obilježje)

Na ispitu znanja iz matematike učenici 1.b razreda postigli su sljedeće rezultate: ocjenu odličan dobilo je 3 učenika, ocjenu vrlo dobar 9 učenika, ocjenu dobar 9 učenika, ocjenu dovoljan 5 učenika, a ocjenu nedovoljan 4 učenika.

Prikažimo te podatke u tablici i odredimo frekvenciju pojedine ocjene. Što je obilježje, a što vrijednost obilježja?

Ocjena	odličan	vrlo dobar	dobar	dovoljan	nedovoljan
Broj učenika	3	9	9	5	4

4.2. Kvantitativne (numeričke) varijable (obilježja)

Statističke varijable kojima se svojstva elemenata statističkog skupa izražavaju brojačno nazivaju se kvantitativne varijable. Nad oblicima numeričkih varijabli dopuštene su sve računske operacije.

Kvantitativne varijable dijele se na *diskretne* i *kontinuirane* i one se mjere omjernom skalom. Intervalnom skalom mjeri se numerička varijabla koja se naziva *intervalom*.

Intervalna skala

- čine je brojevi kojima se mjeri neko svojstvo elemenata statističkog skupa (npr. temperaturna skala: 0 na skali ne znači nepostojanje pojave već $0^{\circ}C$)
- dopuštene su matematičke operacije zbrajanja i oduzimanja (npr. prosječna temperatura u siječnju je $6^{\circ}C$, a u srpnju $30^{\circ}C$, što znači da je temperatura u srpnju u prosjeku viša za $24^{\circ}C$, a ne da je u srpnju 5 puta toplije nego u siječnju)

Omjerna skala

- sastoji se od brojeva za koje također vrijedi da njihove jednake razlike predstavljaju jednake razlike mjerenog svojstva, ali za razliku od intervalne skale nula na mjernoj skali znači nepostojanje svojstva i nije utvrđena dogovorno
- starost, težina, cijena, štednja primjeri su obilježja čiji su modaliteti izraženi brojevima do kojih se dolazi primjenom omjerne skale i nazivaju se numeričkim obilježjima (npr. ako je neki učenik uštedio za maturalac 500 kn, a drugi 1500 kn tada drugi ima tri puta veću uštedevinu od prvog)

Numeričko obilježje koje poprima konačan broj cjelobrojnih vrijednosti (broj djece u obitelji, broj prodanih televizora tijekom mjeseca itd.) naziva se **diskretno**. Vrijednost diskretnog obilježja nije nužno cjelobrojna (npr. broj cipela 37, 37.5, itd.)

Primjer 4.2.1. (numeričko diskretno obilježje)

Bolnički je administrator pregledavajući kartone pacijenata primljenih na bolničko liječenje u rujnu dobio sljedeći niz podataka koji sadrži broj dana provedenih na bolničkom liječenju:

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

broj dana bolničkog liječenja	1	2	3	4	5	6	7	8
broj pacijenata koji su proveli u bolnici taj broj dana	3	5	5	4	2	8	1	1

Primjer 4.2.2.

Kontrola proizvoda vrši se uzimanjem uzorka od 20 proizvoda i zapisivanjem broja neispravnih među njima. Pregledano je 18 uzoraka i dobiveni su podaci:

3 0 0 3 1 0 2 1 0
0 4 0 0 2 2 0 0 0

Broj neispravnih proizvoda u jednom uzorku može biti bilo koji broj iz skupa $\{0,1,2,\dots,20\}$.

Statistički skup (čine ga svi promatrani proizvodi) u ovom slučaju je jednodimenzionalan i diskontinuiran (diskretan).

Neprekidno (kontinuirano) obilježje je obilježje ako poprima bilo koju vrijednost iz nekog intervala npr. visina, težina.

(Za statistički skup kažemo da je *kontinuiran* ako veličina koja se mjeri teoretski može primiti *bilo koju vrijednost* iz nekog konačnog ili prebrojivog skupa).

Primjer 4.2.3. (numeričko kontinuirano obilježje)

U razredu je 20 učenika. Vršeno je mjerenje njihove visine i dobiveni su sljedeći rezultati (u centimetrima):

163	172	165	185	181	172	172	185	164	169
170	169	172	168	162	165	181	181	168	170

20 učenika predstavlja u ovom slučaju jednodimenzionalan statistički skup. Najmanja dobivena vrijednost je 162, a najveća 185. Teoretski, bilo koji broj iz intervala $[162,185]$ mogao je biti rezultat mjerenja (npr. 166,712 ili 170,0007 itd.). Možemo uzeti i širi interval od ovog i još uvijek tvrditi da su svi brojevi unutar njega mogući ishodi mjerenja visine učenika. Dakle, visina učenika je kontinuirana veličina pa je pripadni statistički skup kontinuiran.

Vrijednosti diskretnog obilježja dobivaju se obično brojanjem a neprekidnog mjerenjem.

Zadaci 4.

- Objasnite pojam statističke varijable, odnosno obilježja. Za koje se varijable kaže da su:
 - kvalitativne
 - kvantitativne?Navedite primjere kvalitativnih i kvantitativnih varijabli.
- Kojoj vrsti i podvrsti pripadaju varijable (obilježja):
 - poslovna djelatnost
 - zemlja podrijetla robe
 - cijena dionica na burzi
 - stupanj zadovoljstva kupaca u trgovini
 - broj članova kućanstva
 - količina padalina?
- Koje su od navedenih varijabli diskretne, a koje kontinuirane:
 - broj odsutnih s posla
 - veličina muške košulje
 - težina pakiranja kave dnevni utržak
 - broj stanovnika u općini ?

4. Statistički se skup sastoji od 13470 stanova sagrađenih u Republici Hrvatskoj 2011.god. Kojoj vrsti i podvrsti pripadaju varijable (obilježja):
- lokacija
 - vrsta vlasništva
 - broj soba
 - površina stana
 - cijena po m^2 ?

Rješenja

2) a)kvalitativno-nominalno-atributivno

b) kvalitativno-nominalno-geografski

c) kvantitativno-kontinuirano

d) kvalitativno-diskretno

3) a)D; b) K ; c)D ; d) D

4)a) kvalitativno-nominalno-geografski

b) kvalitativno-nominalno-atributivno

c) kvalitativno-diskretno

e) kvantitativno-kontinuirano

5. Statistički podaci i njihova svojstva

Statistički podaci jesu oni koji su prikupljeni prema utvrđenom nacrtu (planu) statističkog istraživanja, odnosno planu pokusa, statistički su promjenjivi i ima ih dovoljan broj.

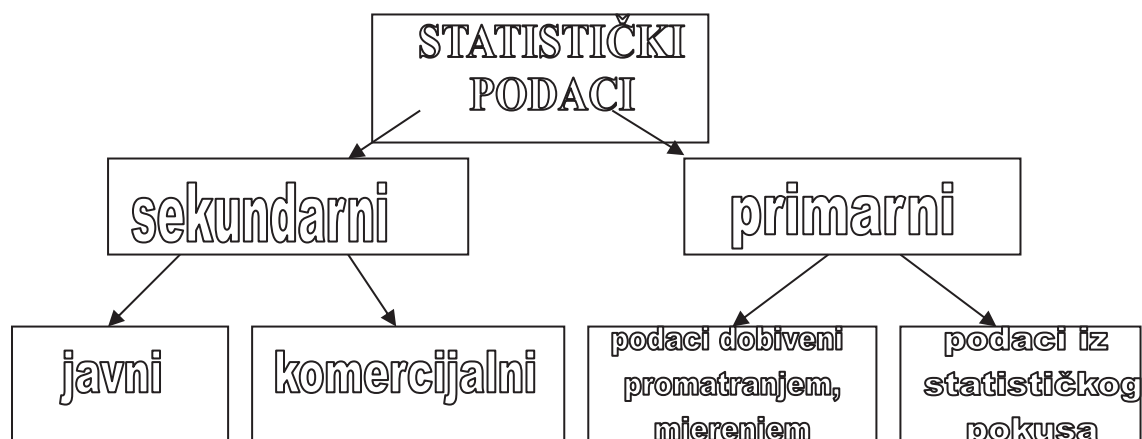
Razlikuju se podaci dobiveni opažanjem i podaci dobiveni statističkim pokusom. Podaci dobiveni opažanjem su oni podaci kakvi jesu i na koje istraživač nema utjecaja. Anketiraju li se turisti, tada je očito da anketar samo evidentira odgovor o zemlji podrijetla, duljini boravka i sl.

U statističkom pokusu na temelju nacrta pokusa istraživač određuje kontrolirane varijable i mjeri njihov utjecaj na neko svojstvo. Ako se mjeri utjecaj različitih količina umjetnih gnojiva na prinos jedne žitarice tada je količina gnojiva pod kontrolom istraživača.

Statistički podaci dobiveni postupkom prikupljanja nazivaju se *izvornima*. Podaci koji su rezultat provođenja računskih radnji nad izvornim podacima o jednoj ili više statističkih varijabli nazivaju se *izvedenima*. Izvorni su podaci o dobi stanovnika u popisu stanovništva, a izvedeni su podaci o gustoći naseljenosti.

5.1. Izvori podataka

Podatke dijelimo na sekundarne i primarne obzirom na izvor podataka.



5.2. Sekundarni podaci

Sekundarni podaci su podaci prikupljeni ili se sustavno prikupljaju u skladu s nekom svrhom i na određeni način, a njihov opseg i vrsta ne izvire neposredno iz potreba danog istraživanja.

(npr. podaci koje objavljuje Državni zavod za statistiku).

U pravilu su to broježani podaci uređeni i prikazani statističkim tablicama, lako su dostupni i ne zahtijevaju velike materijalne troškove. Nedostatak tih podataka je što su često nedostatni i potrebno je dobro procijeniti njihovu kvalitetu.

Prema izvorima, sekundarni podaci dijele se na interne i eksterne. Eksterni se još dijele na javne i komercijalne.

Interni podaci

Podaci pribavljeni unutar organizacije u kojoj se provodi istraživanje (npr. platne liste, poslovna izvješća i sl.)

Eksterni podaci

Podaci pribavljeni izvan istraživačke organizacije ili tvrtke. Ako su javni onda su dostupni svim korisnicima, bez naknade, a ako su komercijalni onda su dobiveni uz naknadu i pribavljeni prema zahtjevima korisnika.

Javni podaci su npr. podaci državnih statističkih ureda, tvrtki koje se bave istraživanjem banaka, osiguravajućih društava i sl.

Komercijalni podaci su podaci poslovnih organizacija kojima je djelatnost prikupljanje i prodaja podataka.

5.3.Primarni podaci

Primarni podaci prikupljaju se u skladu s predmetom i svrhom istraživanja, odnosno potrebama provođenja analize. Oni se pribavljaju utvrđivanjem oblika statističkih obilježja jedinica statističkih uzoraka, ili su rezultat izvođenja statističkog pokusa.

Podaci se mogu prikupljati za sve članove skupa ili samo za dio. Ako se podaci prikupljaju za svaki član skupa onda govorimo o **cenzusu ili iscrpnom obuhvatu jedinica skupa**, npr. popis stanovništva, registracija vozila i slično. Takvo prikupljanje je vrlo skupo i dugo traje.

Prikupljanje podataka može biti **jednokratno, periodično i tekuće**.

Jednokratno se provodi jedanput, periodično se provodi u vremenski jednakim točkama, a tekuće u jednakim vremenskim intervalima.

Primarni se podaci prikupljaju za određeni uzorak, a ne po jedinicama skupa. *Uzorak* je skup odabranih jedinki na kojima se provodi istraživanje. *Slučajni uzorak* se odabire tako da svaki element osnovnog skupa može biti uvršten u uzorak. Najčešće se koriste jednostavni slučajni uzorci.

Jednostavni slučajni uzorak bira se tako da iz skupa od N elemenata biramo n elemenata ($n < N$) tako da za svaki element postoji jednaka vjerojatnost da bude izabran. U praksi se to najčešće radi pomoću računalnog programa koji stvara slučajne brojeve ili pomoću tablica slučajnih brojeva.

5.4. Načini prikupljanja primarnih podataka

Podaci se mogu prikupljati na različite načine, ovisno o opsegu i potrebama istraživanja, raspoloživim tehničkim i financijskim sredstvima i slično.

Načini na koje prikupljamo podatke su ***mjerenje, brojenje, ocjenjivanje i opažanje.***

Mjeri se recimo trajanje telefonskih poziva, težine nekih proizvoda, visina učenika u razredu i slično.

Ocjenjuje se recimo znanje na ispitima, kvaliteta nekog proizvoda itd.

Opažanjem se uočavaju cijene nekih proizvoda, ponašanje članova neke skupine.

Prikupljanje podataka provodi jedna osoba, istraživačka skupina ili osoblje nekog poduzeća ovisno o kakvom je istraživanju riječ. Osobe koje prikupljaju podatke moraju biti obrazovane za taj posao tj. upoznate sa svim postupcima i sredstvima promatranja.

Dobiveni podaci se registriraju pomoću tehničkih uređaja, statističkih upitnika i drugih sredstava. Tehnički uređaji su mjerni uređaji, računala, optički čitači.

Statistički upitnici su upitnici ili anketni listići u kojima su navedena obilježja za jedinice o kojima se traže podaci.

Podaci se mogu prikupljati telefonski te putem interneta, što su vrlo brzi i relativno jeftini načini prikupljanja podataka za određeno istraživanje.

Pitanja u statističkim upitnicima moraju biti jasna, sažeta po mogućnosti s ponuđenim odgovorima. Upitnici bi trebali biti kratki jer je veliki broj pitanja uglavnom odbojan za ispitanike.

Primjer 5.3.1.

Zadatak je istražiti prehrambene navike učenika jedne srednje škole.

ANKETA – ZDRAVA PREHRANA

Zaokruži slovo ispred odgovora:

1. Da li doručkuješ prije odlaska u školu?

- a) Ne
- b) Da
- c) Ponekad

2. Vaš dnevni unos tekućine je:

- a) manji od litre
- b) veći od litre
- c) veći od 2 litre

3. Što pijete od tekućine?

- a) Gazirane sokove
- b) Negazirane sokove
- c) Vodu
- d) Sve navedeno pomalo

4. Što Vam je najvažnije kod hrane?

- a) Da je ukusna
- b) Da je zdrava
- c) Da je niskokalorična

5. Što uobičajeno jedete dok ste u školi?

- a) Pekarske proizvode
- b) Sendviče
- c) Grickalice

6. Po vlastitoj procjeni, kakva je tvoja sveukupna prehrana?

- a) Nezdrava
- b) Zdrava
- c) Ovisno o prilici, kombinacija zdrave i nezdrave
- d) Ne razmišljam uopće o tome

7. Da li bi se želio/la zdravije hraniti i time pozitivno utjecati na vlastito zdravlje?

- a) Da, ali mi takva prehrana nije dostupna
- b) Naravno, uvijek radije biram zdravu hranu umjesto nezdrave
- c) Ponekad, ali nikada ne bih izbacio/la „finu“ nezdravu hranu
- d) Ne, uopće me takva hrana ne zanima

5.5. Priprema podataka za statističku obradu

Sekundarni podaci se relativno brzo pripremaju jer su oni već uređeni za primarno istraživanje za koje su prikupljeni. Priprema primarnih podataka složeniji je posao.

Prvi dio je kontrola postupka prikupljanja podataka koja se vrši tokom prikupljanja podataka tako da se nadzire rad anketara, kontrolira se točnost i smislenost odgovora i uočene nedosljednosti u odgovorima.

Nakon toga podaci se unose u računalne programe za obradu podataka i pohranjuju.

Brzina unosa bitno se povećava *kodiranjem podataka*. Kodiranje (šifriranje) sastoji se od broječanog ili slovnog označavanja odgovora, odnosno oblika obilježja. Brojčane ili slovne oznake tvore *kodnu listu*, koja se utvrđuje u pripremnoj fazi prikupljanja podataka. Spol osobe može se kodirati ovako: za mušku osobu upisuje se broj 1, za žensku 0. Po pravilu kodiraju se svi kvalitativni podaci, a rjeđe podaci numeričkog obilježja.

Zadaci 5.

1. Napravite anketni upitnik da biste istražili koju marku tenisica nosi većina učenika u vašem razredu.
2. Napravite anketni upitnik kako biste saznali koliko vremena dnevno uče učenici u vašem razredu.
3. Što su primarni podaci? Koje su njihove prednosti, a koji nedostaci? Opišite korake pri pribavljanju primarnih podataka. Što se podrazumijeva pod pojmom census?
4. Opišite svojstvo slučajnog uzorka. Za koji uzorak se kaže da je jednostavni slučajni?
5. Za koje se podatke kaže da su iz sekundarnih izvora? Koje su prednosti, a koji su nedostaci sekundarnih podataka. Navedite primjere sekundarnih podataka.

Sažetak

Statistika – znanstvena metoda koja se bavi prikupljanjem, analiziranjem i tumačenjem podataka različite vrste.

Podaci – promotrena kvalitativna i kvantitativna svojstva objekata, stvari, osoba, procesa...

Svrha primjene statističkih metoda – donošenje suda o osobitosti promatranih pojava, ispitivanje različitih pretpostavki, predviđanje razine i stanja pojava.

Metode deskriptivne statistike – sastoje se u primjeni postupaka uređivanja, grupiranja, tabeliranja, grafičkog prikazivanja statističkih podataka; podaci se ne poopćavaju.

Statistički skup – sastoji se od jedinica kojima se ispituje jedno ili više svojstava koja od jedinice do jedinice očituju statističku promjenljivost

Opseg skupa – broj jedinica (prema opsegu skupa statistički skupovi se dijele na konačne i beskonačne)

Konačan skup – ima konačan broj elemenata

Beskonačan skup – ima beskonačno mnogo elemenata

Skup podataka ili osnovni skup – podaci o danoj varijabli za svaki element statističkog skupa

Populacija može biti:

- realna
- hipotetična
- konačna
- beskonačna

Uzorak – podskup statističkog skupa (u svakom statističkom istraživanju realni statistički skupovi se definiraju pojmovno, prostorno i vremenski)

pojmovno – određuje se pripadnost skupu

prostorno – određuje se kojem skupu pripadaju sve jedinice stat. skupa

vremenski – određuje se vremenski interval ili vremenska točka za koje su vezane sve jedinice skupa

Vrste i izvori statističkih podataka

Statistički podaci – rezultati mjerenja svojstava jedinica statističkih skupova, njihovih podskupova ili eksperimentalnih jedinica

Statistička varijabla ili obilježje – svojstvo koje oblikom ili stupnjem varira od jedinice do jedinice skupa; po njemu se elementi skupa razlikuju ili jedni drugima nalikuju

II UREĐIVANJE I PRIKAZIVANJE PODATAKA

1. Formiranje statističkih nizova, frekvencije

Nakon prikupljanja podataka, drugi korak u statističkim istraživanjima je uređivanje i prikazivanje podataka. Ovaj korak je nužan iz razloga što često prikupljeni podaci u izvornom obliku nisu pregledni i nije moguće donositi zaključke jer su podaci mnogobrojni. Uobičajeno se podatci navode prema nekom pravilu ili se pak grupiraju te njihovim uređivanjem nastaje **statistički niz**.

Prilikom grupiranja smještamo u grupe jedinice istih obilježja. Broj jedinica statističkog skupa koje imaju isto obilježje nazivamo **apsolutnom frekvencijom** (kraće: frekvencija f_i) ili **učestalost**. Pojam frekvencije već smo spomenuli i u prvom poglavlju, gdje smo opisali da frekvencije zapravo pokazuje koliko se puta neko svojstvo pojavljuje kod članova zadanog statističkog skupa. Zbroj svih frekvencija jednak je ukupnom broju N elemenata statističkog skupa. Frekvencije su brojevi između 0 i N . Prisjetimo se pojma frekvencije niza kroz sljedeći primjer.

Primjer 1.1.

Zaposlenici jedne tvrtke anketirani su o tome koji im je dan u tjednu najdraži. Dobiveni su sljedeći podaci:

ut sri čet pon pon sri pet pet čet sub
ned ned pon pet pet sri pet pet čet sub

Obilježje danog skupa je dan u tjednu, a vrijednosti su: ponedjeljak, utorak, srijeda, četvrtak, petak, subota, i nedjelja. Vrijednost *pon* pojavljuje se tri puta, vrijednost *ut* jednom, vrijednost *sri* tri puta itd. Podatke prikazujemo u **tablici frekvencija**.

Zadnji redak dodan je kao kontrolni redak.

Dani u tjednu	Frekvencija
<i>pon</i>	3
<i>ut</i>	1
<i>sri</i>	3
<i>čet</i>	3
<i>pet</i>	6
<i>sub</i>	2
<i>ned</i>	2
Zbroj	20

Iz prikazanih podataka vidimo da je anketirano **20** zaposlenika te da im je najdraži dan u tjednu petak.

Primjer 1.2

Podaci u tablici prikazuju statistički skup koji čini **40** osoba koje su u jednom danu kupile cipele u jednoj prodavaonici muških cipela. U lijevom stupcu tablice nalaze se vrijednosti obilježja elemenata tog skupa, veličina kupljenog para cipela, a u desnom stupcu frekvencije, broj kupaca.

Veličina cipela	Broj kupaca
36	2
37	1
38	3
39	2
40	6
41	13
42	6
43	4
44	2
45	1
Zbroj	40

Na temelju tablice frekvencija, poslovođa promatrane prodavaonice cipela može zaključiti da prilikom nove narudžbe cipela treba naručiti najviše pari veličine **41**. Sljedeće po količini narudžbe trebaju biti cipele veličina **40** i **42** itd.

Pretpostavimo da prodavaonica u sljedećoj narudžbi treba naručiti **600** pari cipela. Koliko među tih **600** pari treba biti onih s brojem **41**, koliko s brojem **40** itd.? Na to pitanje ne dobivamo odgovor iz dane tablice. Iz tog razloga se uvode relativne frekvencije vrijednosti promatranog obilježja.

Relativna frekvencija pojedine vrijednosti promatranog obilježja jednaka je količniku frekvencije te vrijednosti i ukupnog broja elemenata statističkog skupa N . Označit ćemo je s p_i . Zbroj svih relativnih frekvencija jednak je **1**.

Dakle, Ako su frekvencije $f_1, f_2, \dots, f_k$, relativne su frekvencije:

$$p_i = \frac{f_i}{N}, i = 1, 2, \dots, k$$

gdje je $f_1 + f_2 + \dots + f_k = N$.

Odredimo relativne frekvencije iz našeg primjera.

Veličina cipela	Broj kupaca	Relativne frekvencije
36	2	$\frac{2}{40} = 0.05$
37	1	$\frac{1}{40} = 0.025$
38	3	$\frac{3}{40} = 0.075$
39	2	$\frac{2}{40} = 0.05$
40	6	$\frac{6}{40} = 0.15$
41	13	$\frac{13}{40} = 0.325$
42	6	$\frac{6}{40} = 0.15$
43	4	$\frac{4}{40} = 0.1$
44	2	$\frac{2}{40} = 0.05$
45	1	$\frac{1}{40} = 0.025$
Zbroj	40	1

Sada je jasno da bi pri novoj narudžbi od **600** pari cipela poslovođa trebao naručiti:

$$600 \cdot 0.05 = 30 \text{ pari cipela broj } 36$$

$$600 \cdot 0.025 = 15 \text{ pari cipela broj } 37$$

$$600 \cdot 0.075 = 45 \text{ pari cipela broj } 38 \text{ itd.}$$

Uvođenje relativnih frekvencija omogućuje nam da uspoređujemo ista obilježja koja se promatraju na različitim skupovima. To pokazuje sljedeći primjer.

Primjer 1.3

U jednoj srednjoj školi sa **800** učenika promatrano je obilježje zaključna ocjena iz matematike, na skupu svih učenika škole i na skupu učenika **1.a** razreda.

Dobivene su sljedeće tablice.

Ocjena iz matematike na skupu svih učenika škole			
Vrijednosti obilježja	Frekvencije	Relativne frekvencije	Relativne frekvencije u postotcima
1	40	0.05	5
2	168	0.21	21
3	188	0.235	23.5
4	260	0.325	32.5
5	144	0.18	18
Zbroj	800	1	100

Ocjena iz matematike na skupu učenika 1.a razreda			
Vrijednosti obilježja	Frekvencije	Relativne frekvencije	Relativne frekvencije u postotcima
1	2	0.08	8
2	5	0.2	20
3	6	0.24	24
4	8	0.32	32
5	4	0.16	16
Zbroj	25	1	100

Relativnu frekvenciju (oznaka: P_i) zapisanu u obliku postotka zovemo **postotna relativna frekvencija**.

Na temelju frekvencija iz danih tablica ne možemo donijeti nikakav zaključak. Za vrijednost obilježja *ocjene odličan* je frekvencija u prvoj tablici **144**, a u drugoj **4**. No, frekvencija u

prvoj tablici dobivena je na skupu od **800** učenika, a u drugoj na skupu od **25** učenika. Zato te frekvencije ne možemo uspoređivati.

Uvođenje relativnih frekvencija omogućuje nam uspoređivanje. Iz tablica očitavamo: **18%** učenika cijele škole i **16%** učenika **1.a** razreda imalo je ocjenu **5** iz matematike. Možemo zaključiti da je uspjeh iz matematike na cijeloj školi bolji nego u **1. a** razredu.

Podatke kod kojih je frekvencija svake vrijednosti mala, potrebno je grupirati u grupe koje nazivamo **razredi**. Razred je interval oblika **$[a, b]$** (intervala mogu biti i otvoreni, ili poluotvoreni). Brojeve **a, b** zovemo **granice razreda**. Razredi mogu biti jednakih ili različitih veličina, ali se preporuča se da broj razreda ne premašuje **15**.

Broj jedinica koje čine neki razred označava njegovu frekvenciju. Ponekad je potrebno odrediti i **precizne granice**. Precizne granice ćemo odrediti tako da najprije odredimo razliku donje granice tekućeg razreda i gornje granice prethodnog razreda. Potom donju originalnu granicu umanjimo za polovinu spomenute razlike, a gornju granicu za polovinu razlike povećamo.

Veličina ili širina t – tog razreda je razlika između donje granice sljedećeg razreda i donje granice tekućeg razreda. Ako želimo odrediti podijeliti podatke u razrede jednakih širina, odredimo razliku najvećeg i najmanjeg podatka u grupi, podijelimo ju s brojem razreda i zaokružimo na više. Prikažimo grupiranje u **6** razreda jednake širine u sljedećem primjeru.

Primjer 1.4.

Učenicima 1.a razreda izmjerena je visina i dobiveni su ovi podaci iskazani u centimetrima:

182 153 164 173 184 175 180 155
201 177 180 183 186 188 182 178
169 168 173 159 152 162 163 185

Najmanja izmjerena visina je **152**, a najveća **201**.

Izračunajmo širinu razreda: $\frac{201-152}{6} = 8.16$. Dobiveni broj zaokružimo na **9**.

Budući da je širina svakog razreda **9**, to znači na se u prvom razredu nalaze vrijednosti: **152, 153, 154, 155, 156, 157, 158, 159** i **160**. I tako dalje za svaki razred. Napravimo tablicu frekvencija za razrede.

Visina učenika	Frekvencija
152 – 160	4
161 – 169	5
170 – 178	5
179 – 187	8
188 – 196	1
197 – 205	1

U ovom primjeru možemo uvesti i precizne granice razreda. Naime, razredi se moraju nastavljati jedan na drugoga. Zato se kao precizne granice prvoga razreda stavljaju **151.5** kao donja granica, i **160.5** gornja granica razreda itd. Uvođenje preciznih granica ima smisla jer smo kod mjerenja svaki rezultat koji je bio veći od **160.5** registrirali kao **161**, a svaki rezultat koji je bio manji od **160.5** registrirali smo kao **160**. Unesimo sada u tablicu i precizne granice razreda.

Visina učenika	Precizne granice razreda	Frekvencija
152 – 160	151.5 – 160.5	4
161 – 169	160.5 – 169.5	5
170 – 178	169.5 – 178.5	5
179 – 187	178.5 – 187.5	8
188 – 196	187.5 – 196.5	1
197 – 205	196.5 – 205.5	1

Ako nemamo razrede jednake širine, njihove frekvencije nisu usporedive pa je potrebno korigirati frekvencije. **Korigirane frekvencije** računamo dijeljenjem originalnih frekvencija sa širinama razreda.

Primjer 1.5.

Tablica prikazuje broj članova obveznih mirovinskih fondova u RH u veljači 2011. godine, prema dobi (u tisućama):

Starost, godine	Broj članova u tisućama	Širina razreda
15 – 24	187.84	10
25 – 34	570.68	10
35 – 44	523.49	10
45 – 64	286.54	20

Budući da su razredi različitih širina, potrebno je korigirati frekvencije dijeljenjem sa širinom razreda.

Starost, godine	Broj članova u tisućama	Širina razreda	Korigirane frekvencije
15 – 24	187.84	10	$\frac{187.84}{10} = 18.784$
25 – 34	570.68	10	$\frac{570.68}{10} = 57.068$
35 – 44	523.49	10	$\frac{523.49}{10} = 52.349$
45 – 64	286.54	20	$\frac{286.54}{20} = 14.327$

Ukoliko korigiramo relativne frekvencije, dobivamo **korigiranu relativnu frekvenciju** koja se ponekad naziva i **gustoća**.

Zadaci 1.

1. Učenici svih prvih razreda odgovarali su na pitanje: "Koliko imaš braće i sestara."

Dobiveni su ovi odgovori:

```

0 0 0 2 1 3 1 1 1 0 3 2 1 1 1
0 0 1 1 1 2 4 1 1 2 2 1 1 0 0

1 0 1 1 0 2 2 3 2 1 1 2 1 0 1
4 1 0 0 1 1 1 0 1 2 2 1 1 1 1

```

- Što je obilježje, a što su vrijednosti tog obilježja?
 - Napravi tablicu frekvencija.
 - Koliko djece nema ni brata ni sestru?
 - Koliko učenika ima barem jednog brata ili sestru?
 - Koliki dio učenika ima više od dvoje braće?
2. Provedite anketu u svom razredu:
- o tome koje kućne ljubimce imaju učenici
 - o sportovima kojima se bave učenici
 - o vrsti mobilne mreže kojom se učenici koriste

Podatke prikažite u tablicama frekvencije.

3. Tablica prikazuje broj učenika koji su završili srednju školu **2010.** godine u Republici Hrvatskoj.

Spol	Broj učenika
ženski	22 379
muški	22 431
ukupno	44 810

Izračunajte relativne i postotne za dane podatke.

4. Učenici četvrtog razreda jedne medicinske škole svoju praktičnu nastavu odrađuju u bolnici na različitim odjelima: interni(I), kardiologija(K), traumatologija(T), zarazni (Z), pedijatrija(P) i neurologija(N). Napravite tablični prikaz frekvencija, relativnih frekvencija i postotnih relativnih frekvencija zastupljenosti pojedinog odjela na temelju navedenih podataka.
5. U gradu radi **40** trgovina prehrambenim artiklima. U rujnu su te trgovine imale sljedeći promet iskazan u tisućama kuna:

23	10	15	50	66	77	34	39	46	52
110	98	81	72	65	62	89	37	121	96
51	49	107	39	47	53	66	98	105	84
28	38	73	82	104	80	71	69	44	58

Podatke grupiraj u **6** razreda i prikaži tablicom frekvencija.

6. Tablica prikazuje radnike poduzeća **XYZ** prema godinama starosti. Izračunaj veličine razreda i korigiraj frekvencije.

Godine starosti	Broj radnika
18 – 20	10
20 – 22	15
22 – 28	15
28 – 32	8

Rješenja

1. a. Obilježje: broj braće i sestara, vrijednosti obilježja: **0,1,2,3,4**

b.

c. **14**

d. **46**

e. $\frac{5}{60} = 8.3\%$

Broj braće i sestara	Frekvencije
0	14
1	30
2	11
3	3
4	2

3.

Spol	Broj učenika	Relativne frekvencije	Relativne postotne frekvencije
ženski	22 379	0.2497	24.97
muški	22 431	0.2503	25.03
ukupno	44 810	0.5	50

4.

	Frekvencije	Relativne frekvencije	Relativne postotne frekvencije
Interni	5	0.1786	17.86 %
Kardiologija	4	0.1429	14.29 %
Traumatologija	2	0.0714	7.14 %
Zarazni	4	0.1429	14.29 %
Pedijatrija	7	0.25	25 %
Neurologija	6	0.2143	21.43 %
Ukupno	28	1	100 %

5.

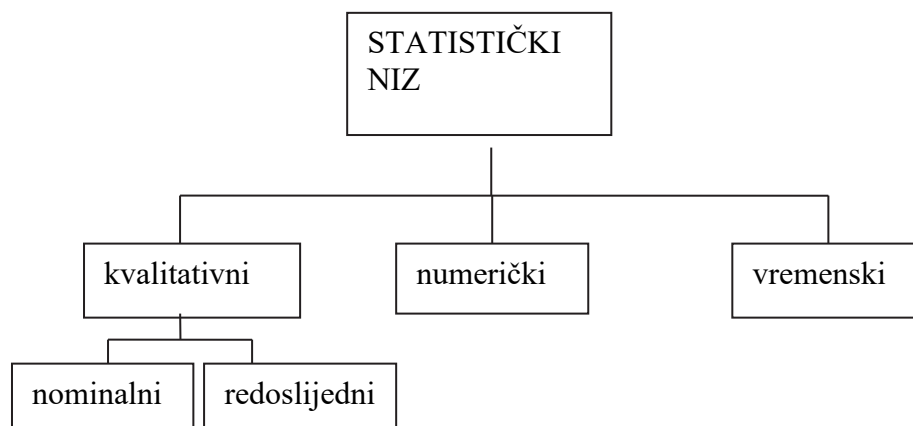
Razredi	Frekvencije
10 – 28	4
29 – 47	8
48 – 66	10
67 – 85	9
86 – 104	5
105 – 123	4

6.

Godine starosti	Broj radnika	Veličina razreda	Korigirane frekvencije
18 – 20	10	2	5
20 – 22	15	2	7.5
22 – 28	15	6	2.5
28 – 32	8	4	2

2. Vrste statističkih nizova

Nakon što se formiraju statistički nizovi, kako bismo ih mogli analizirati potrebno je utvrditi o kakvom se nizu radi.



2.1. Nizovi kvalitativnih podataka

Ako uređujemo podatke koji ovise o jednoj kvalitativnoj varijabli, onda govorimo o nizu kvalitativnih podataka. Kvalitativni statistički nizovi mogu biti *redoslijedni* i *nominalni*.

Grupiranjem podataka prema redoslijednom obilježju nastaju redoslijedni nizovi, dok se nominalni niz formira grupiranjem jedinica statističkog niza prema oblicima nominalnog obilježja (prisjetimo se nekih nominalnih obilježja: zanimanje, spol, nacionalnost, bračno stanje, marka automobila, mjesto boravka, rođenja).

Niz kvalitativnih podataka pregledno se prikazuje statističkom tabelom i grafikonom.

Primjer 2.1.1. (nominalni niz)

Statistički skup čine učenici koji su završili srednju školu u RH prije 2010. godine. Raščlanit ćemo ih na grupe prema obilježju spol.

Spol	Broj učenika f_i
Ženski	22 379
Muški	22 431
Ukupno	44 810

Primjer 2.1.2. (redoslijedni niz)

Prikažimo redoslijedni niz temeljem podataka o završnom uspjehu učenika srednjih škola na kraju školske godine iz izvora matične knjige evidencije škole.

Završni uspjeh
Dovoljan
Dobar
Vrlo dobar
Odličan

2.2. Numerički statistički niz

Numerički statistički niz formira se uređenjem kvantitativnih podataka, odnosno vrijednosti numeričke varijable. Način uređivanja ovisi o broju podataka i o tome jesu li podaci vrijednosti diskretne ili kontinuirane numeričke varijable. Podaci se mogu nizati prema veličini, grupirati prema pojedinačnim vrijednostima numeričke varijable ili po razredima.

Ako je podataka malo, radi se uređivanje podataka na temelju utvrđenog pravila (abecednim redom, prema rangu, učestalosti). Ako je podataka puno uređujemo ih grupiranjem.

Primjer 2.2.1. (numerički niz)

Prikazani su podaci o navršenim godinama radnog staža zaposlenih:

3, 7, 2, 9, 6, 7, 12, 15, 3, 8, 25, 17, 32, 1, 1

Broj podataka je 15, tj. mali je broj vrijednosti.

Uređeni podaci su slijedeći:

x_i	1	1	2	3	6	7	7	8	9	12	15	17	25	32
-------	---	---	---	---	---	---	---	---	---	----	----	----	----	----

Postupnim zbrajanjem originalnih frekvencija (apsolutnih ili relativnih) numeričkog niza nastaju **kumulativni niz**. Prva frekvencija kumulativnog niza jednaka je prvoj frekvenciji izvornog niza. Druga frekvencija kumulativnog niza jednaka je zbroju prve i druge frekvencije polaznog niza. Posljednja frekvencija tog niza jednaka je zbroju svih frekvencija izvornog niza. Simbolički se formiranje kumulativnog niza može opisati na sljedeći način.

Kumulativni niz numeričkog niza s frekvencijama $f_1, f_2, \dots, f_k$ dan je izrazom:

$$S_1 = f_1, S_2 = f_1 + f_2, \dots, S_k = f_1 + f_2 + \dots + f_k$$

Kumulativni niz jednako se formira i na temelju relativnih frekvencija i postotnih frekvencija.

Iz načina formiranja kumulativnog niza slijedi i značenje njegovih članova. i -ti član kumulativnog niza pokazuje koliko je jedinica kojima je vrijednost numeričke varijable

jednaka onoj vrijednosti varijable čija je izvorna frekvencija posljednja ušla u kumulativni niz ili je manja od nje.

Ako se kumulativni niz formira na temelju postotnih frekvencija onda i -ti član kumulativnog niza pokazuje koliki je *postotak jedinica* kojima je vrijednost numeričke varijable jednaka onoj vrijednosti varijable čija je frekvencija posljednja ušla u niz ili je manja od nje.

Primjer 2.2.2. (kumulativni niz)

Prikazani su podaci o završenim stanovima za stanovanje prema broju soba (jednosobni, dvosobni...) u Hrvatskoj 2009. godine.

Broj soba	Broj stanova	Kumulativni niz (na temelju frekvencija)	Struktura u %	Kumulativni niz (na postotnih frekvencija)
1	2 435	2 435	13.21	13.21
2	5 845	8 280	31.70	44.91
3	4 955	13 235	26.88	71.79
4	2 997	16 232	16.26	88.05
5-8	2 204	18 436	11.95	100

Pogledajmo kako je formiran kumulativni niz na temelju frekvencija:

$$S_1 = f_1 = 2\,435, S_2 = f_1 + f_2 = 2\,435 + 5\,845 = 8\,280, \dots,$$

$$S_5 = f_1 + f_2 + \dots + f_5 = 18\,436 = N$$

Postupnim zbrajanjem postotnih frekvencija nastao je kumulativni niz prikazan u zadnjem stupcu.

Za primjenu je važno tumačenje članova kumulativnog niza. Treći član kumulativnog niza na temelju frekvencija jest 13 235. On pokazuje da je broj stanova s 1, 2 i 3 sobe jednak 13 235.

Drugim riječima, 13 235 je broj završenih stanova s 3 i *manje soba*. Pripadajući član kumulativnog niza na temelju postotnih frekvencija je 71.79%. Prema tome, od ukupnog broja završenih stanova s 3 i manje soba ih je 71.79%.

2.3. Vremenski niz

Vremenski niz čine kronološki uređeni podaci.

Primjer 2.3. (vremenski niz)

U tablici su prikazani podaci o ukupnom broju zaposlenih u Hrvatskoj u razdoblju 2005.-2011., te o broju zaposlenih žena (u tisućama):

Godina	2005.	2006.	2007.	2008.	2009.	2010.	2011.
Ukupno	1 420	1 468	1 517	1 555	1 499	1 432	1 411
Od toga žena	641	661	684	701	679	657	651
Postotak žena	45.1	45.0	45.1	45.1	45.3	45.9	46.1

3. Tabeliranje

Tabelarno prikazivanje statističkih podataka ima za cilj da na jasan i pregledan način prikaže rezultate prikupljanja i grupiranja statističkih podataka. Svaka statistička tablica mora imati naslov, tekstualni dio, brojčani dio i izvor podataka.

Pogledajmo jedan primjer složenije tablice:

Tablica 1. Naslov tablice

		ZAGLAVLJE				
		Oznake stupca				
0		1	2	...	<i>n</i>	
Predstupac	Oznake					Zbirni stupac
	retka		Brojčani	dio	tablice	
		Polje				
					Polje	
Zbirni redak						

Izvor podataka...

Zaglavljje je prvi redak koji služi za tumačenje brojeva u stupcima. Predstupac je prvi stupac koji služi za tumačenje brojeva koji se nalaze u redcima. Zbirni redak služi za unošenje zbroja danih u stupcima, a zbirni stupac za unošenje zbrojeva brojeva u redcima tablice.

Brojčani dio tablice sastoji se od polja u koja se unose rezultati grupiranja statističkih podataka u obliku apsolutnih (ili relativnih) frekvencija. Ispod tablice navodi se izvor

podataka. Niti jedno polje u brojčanom dijelu tablice ne smije ostati prazno. U svako polje upisuje se broj ili drugi znak koji je uobičajen u statističkoj praksi.

Evo primjera nekih oznaka i njihovo značenje:

- * ispravljen podatak
- 1) oznaka za napomenu ispod tablice
- () nepotpun podatak
 - nema pojave
- ... ne raspolaže se podatkom
- 0 podatak je manji od 0.5 upotrijebljene jedinice mjere
- 00 podatak je manji od 0.05 upotrijebljene jedinice mjere
- Ø prosjek

Oblik statističke tablice ovisi o tome promatramo li jedan ili više statističkih skupova prema jednom ili više obilježja. Razlikuju se:

- a. jednostavne statističke tablice
- b. složene statističke tablice koje mogu biti skupne i kombinirane

Primjer 3.1.

Tablica 2. Noćenja turista xxxx godine u izabranim mjestima (u 000)

Mjesto	Broj turista
Zagreb	725
Kupališna mjesta	612
Primorska mjesta	8022
Ukupno	9359

Izvor: Statistički ljetopis xxxx godine, str. xx.

Ovo je jednostavna statistička tablica u kojoj je prikazano jedno statističko obilježje.

Primjer 3.2.

Tablica 3. Status studiranja prema vrsti škola školske godine xxxx/xx

škole	izvanredni status	redovni status	ukupno
srednje škole	310	894	1 204
veleučilišta	1 080	5 641	6 721
fakulteti	1 709	17 594	19 303
ukupno	3 099	24 129	27 228

Izvor : Statistički ljetopis xxxx godine, str. xxx.

Ovo je kombinirana statistička tablica koja prikazuje statističke nizove koji su nastali grupiranjem statističkog skupa učenika i studenata prema statusu studiranja i vrsti škole.

4. Grafičko prikazivanje kvalitativnih nizova

Pri obradi statističkih podataka, radi lakšeg uočavanja bitnih osobina promatrane pojave, služimo se i geometrijskim zorom, tj. statističke nizove prikazujemo grafički koristeći:

stupčasti dijagram
 kružni dijagram
 dijagram točaka
 dijagram stablo-list
 histogram
 poligon frekvencija
 kumulanta

4.1. Stupčasti dijagram (eng. BarChart)

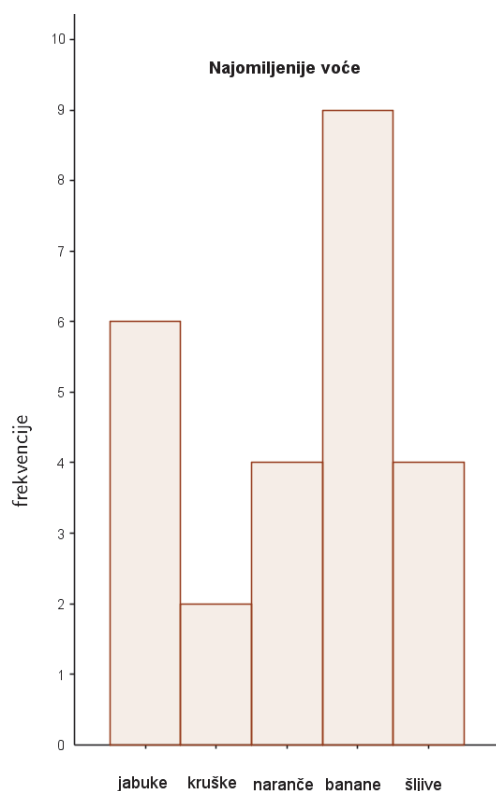
Stupčasti dijagram je površinski grafikon koji predočava statistički skup pomoću pravokutnika koji imaju jednaku osnovicu a čija visina predstavlja frekvencije. Pravokutnike smještamo u pravokutnom koordinatnom sustavu, pri čemu osnove smještamo na os apscisa, a frekvencije na os ordinata.

Primjer 4.1.1. (jednostavni stupčasti dijagram)

Nastavnici škole anketirani su o najomiljenijoj vrsti voća. Dobiven je statistički skup kojem je obilježje "najomiljenije voće", a vrijednosti tog obilježja su: jabuke, kruške, naranče, banane, i šljive. Njihovi odgovori, frekvencije, te izračunata relativna frekvencija dani su u tablici.

Najomiljenija vrsta voća	Frekvencija	Relativna frekvencija
jabuke	6	0.24
kruške	2	0.08
naranče	4	0.16
banane	9	0.36
šljive	4	0.16
Zbroj	25	1

Podatke iz te tablice uobičajeno je grafički prikazati na sljedeći način. Nacrtamo pravokutni koordinatni sustav u ravnini (napomenimo da duljine jedničnih dužina na koordinatnom osima najčešće nisu međusobno jednake). Na os apscisa naneseemo vrijednosti obilježja, a na os ordinata naneseemo frekvencije tih vrijednosti. Zatim nacrtamo pravokutnike kojima su duljine visina jednake odgovarajućim frekvencijama.



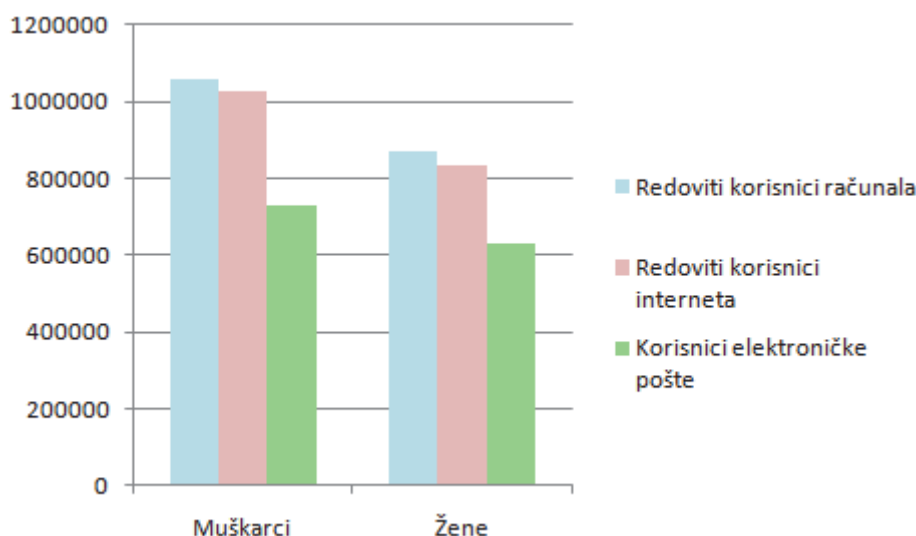
Uspoređivanje dva ili više statističkih nizova provodi se pomoću dvostrukih ili višestrukih stupaca. Crtaju se jednako kao jednostavni stupčasti dijagrami, pri čemu se različiti nizovi označavaju drugačijim bojama. Primjer višestrukih stupaca prikazan je u sljedećem primjeru.

Primjer 4.1.2. (višestruki stupčasti dijagram)

Uspoređivanje dva ili više nizova podataka provodi se pomoću dvostrukih ili višestrukih stupaca. Grafikon višestrukih stupaca crtamo u pravokutnom koordinatnom sustavu, pri čemu su stupci jednakih osnovica, dok im je visina određena frekvencijama. Stupce crtamo u skupinama po dva ili više, ovisno o tome koliko se nizova uspoređuje, pri čemu se preporuča usporediti manji broj nizova kako bi grafikon bio pregledan.

Temeljem podataka u tablici o korisnicima interneta i računala u Hrvatskoj u **2010.** godini izrađen je grafikon trostrukih stupaca.

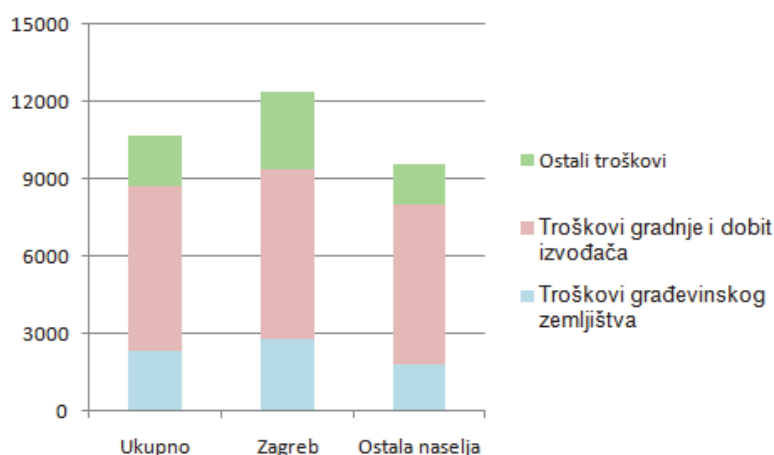
Korisnici	Muškarci	Žene
Redoviti korisnici računala	1 060 218	869 785
Redoviti korisnici interneta	1 028 250	834 722
Korisnici elektroničke pošte	728 138	630 568



Promatranjem slike možemo zaključiti kako je veći broj korisnika računala, interneta i elektroničke pošte muškog spola. To se može, primjerice, objasniti većom stopom zaposlenosti muškaraca koji na radnom mjestu koriste računala.

Primjer 4.1.3. (razdijeljeni ili strukturni stupci)

Razdijeljeni ili strukturni stupci prikazuju statistički niz čije ga frekvencije rastavljaju na dva ili više dijelova. Visina svakog stupca predoduje opseg pojedinačnog skupa koji analiziramo. Stupce možemo crtati tako da na os ordinata nanosimo frekvencije (ili relativne frekvencije). Pomoću ovih stupaca možemo uspoređivati strukturu međusobno srodnih skupova podataka. Na slici su prikazani razdijeljeni stupci na primjeru prosječne cijene kvadratnog metra prodanog novoizgrađenog stana u **2010.** godini, pri čemu se vrši usporedba na razini Hrvatske, grada Zagreba i ostalih naselja. Može se uočiti kako je prosječna cijena kvadratnog metra stana najveća u Zagrebu, ali je struktura troškova koji čine tu cijenu slična u Zagrebu i u ostalim naseljima. Međutim, troškovi građevinskog zemljišta su gotovo dvostruko veći u Zagrebu u odnosu na ostala naselja, što vjerojatno doprinosi većoj cijeni kvadratnog metra stana.



4.2. Kružni dijagram

Kružni dijagram služi za isticanje odnosa frekvencija, uspoređivanje opsega dva ili više statističkih nizova. Polumjer kruga određuje se proizvoljno, pri čemu isječki kruga moraju biti proporcionalni frekvencijama niza. S obzirom da puni krug ima 360° potrebno je prije crtanja, izračunati veličinu kuta svakog kružnog isječka. Stupnjevi isječka dani su izrazom:

$$s_i = \frac{f_i}{N} 360.$$

gdje je f_i frekvencija pojedinog obilježja, a N ukupan broj elemenata statističkog skupa.

(Napomena: količnikom $\frac{f_i}{N}$ računamo relativne frekvencije, pa ukoliko ih već imamo poznate u tablici dovoljno je stupnjeve isječka izračunati množeći ih s 360° .)

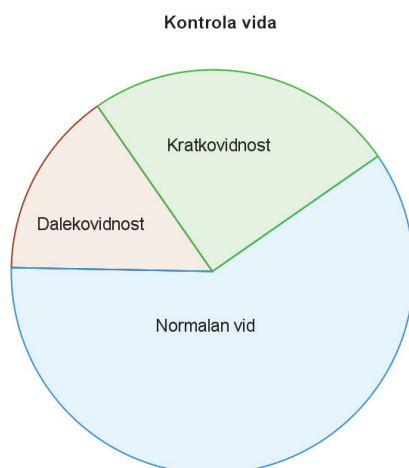
Primjer 4.2.

60 djelatnika jedne firme poslano je na kontrolu vida. Nakon pregleda kod očnog liječnika dobivena je tablica razdiobe promatranog obilježja (vid) na tom skupu od 60 osoba.

Obilježje (vid)	Frekvencije	Relativne frekvencije	Postotne relativne frekvencije
Normalan vid	36	0.6	60
Kratkovidnost	15	0.25	25
Dalekovidnost	9	0.15	15
Zbroj	60	1	100

Tablicu grafički prikazujemo pomoću kružnog dijagrama na sljedeći način.

Iz tablice očitavamo da 60% osoba iz danog skupa ima normalan vid. Puni kut ima 360° . Zato je 60% od punog kuta jednako $360^\circ \cdot 0.6 = 216^\circ$. Slično je 25% od punog kuta jednako $360^\circ \cdot 0.25 = 90^\circ$, a 15% od punog kuta jednako je $360^\circ \cdot 0.15 = 54^\circ$. Sad nacrtamo krug s tri kružna isječka koji su određeni središnjim kutovima od $216^\circ, 90^\circ$, i 54° .



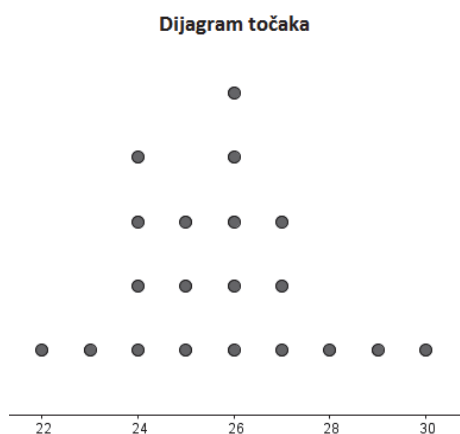
4.3. Dijagram točaka (eng. DotPlot)

Kod nizova s manjim brojem podataka koristimo dijagram s točkama. Dijagram s točkama crtamo tako da na vodoravno položenoj dužini naznačimo mjerilo nekog obilježja. Iznad vodoravne linije ucrtavamo točke nad onim vrijednostima na koje se odnose. Prikažimo dijagram točaka za podatke u sljedećem primjeru.

Primjer 4.3.

Na satu tjelesne i zdravstvene kulture mjeri se vrijeme izvođenja zadane vježbe. Podaci dobiveni u 20 nezavisnih mjerenja su (u sekundama):

24 28 22 26 24 27 26 25 26 23 30 26 29 25 27 24 26 25 24 27



4.4. Dijagram stablo-list (eng. StemPlot)

Dijagram stablo-list koristan je za prikaz podataka u kojima je frekvencija svake vrijednosti mala (**1** ili **2**). Potrebno je najprije vrijednosti poredati po veličini te se zatim znamenke svake od vrijednosti niza podijele na dva dijela. Prvi dio znamenke nazivamo stablo, a ostatak list. Stablo od lista odvajamo okomicom. U stupcu prvo nižemo vrijednosti stabla od najmanje do najveće, a zatim uz svako stablo predložimo odgovarajuće listove.

Primjer 4.4.

Ordinaciju je tijekom tjedna posjetilo **60** pacijenata. Njihove godine starosti su sljedeće:

29	35	46	47	49	65	31	65	66	70
71	21	73	37	55	43	39	66	78	64
38	23	65	37	54	45	44	61	29	51
72	48	58	45	56	80	49	63	30	53
54	52	46	61	60	37	51	71	31	68
60	61	35	59	57	42	53	43	45	64

Prikažimo podatke u stablo-list dijagramu.

```

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

```

Ključ : 3|1 znači 31

U prvi redak zapisujemo sve brojeve kojima je prva znamenka **2**, u drugi redak sve brojeve kojima je prva znamenka **3**, itd., a drugu znamenku zapisujemo nakon vertikalne crte. Postupak ponavljamo dok ne iscrpimo sve mogućnosti. Uz dijagram zapisujemo i ključ u kojem je opisano značenje zapisa.

4.5. Histogram

Histogram je površinski grafikon kojeg konstruiramo pomoću stupaca (pravokutnika) koji se oslanjaju jedan na drugi. Visina tih stupaca ovisi o veličini frekvencije pojedinog obilježja. Konstruiramo ga u pravokutnom koordinatnom sustavu u kojemu na os apscisa nanosimo brojčane vrijednosti obilježja, a na okomite frekvencije.

Primjer 4.5.

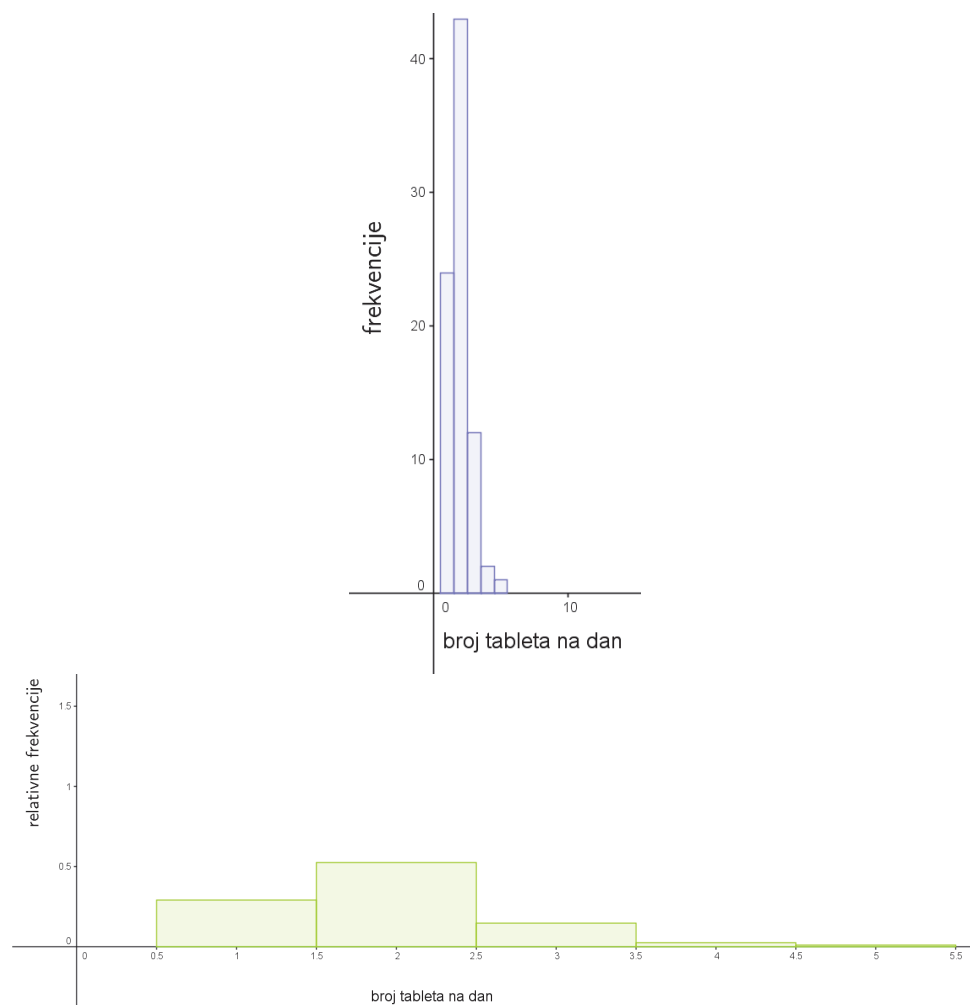
Ispitujući količinu lijekova koje dnevno uzimaju pacijenti, došlo se do sljedećih podataka o broju lijekova u obliku tableta koje svaki dan uzimaju pacijenti bolničkog odjela:

broj tableta koje se uzimaju svakodnevno	broj pacijenata koji uzimaju taj broj lijekova
1	24
2	43
3	12
4	2
5	1

Izračunajmo prvo relativne frekvencije tako da podijelimo frekvencije s ukupnim brojem podataka. Istraživanje je provedeno na ukupno **82** pacijenta. Vrijednosti obilježja su brojevi: **1,2,3,4,5** i za svaku vrijednost se crta stupac. Stupci se crtaju spojeno i to ovako: na horizontalnoj osi odredimo točke **1,2,3,4,5**. Tada baza stupca počinje **0.5** prije i završava **0.5** poslije, a to su precizne granice razreda.

broj tableta	precizne granice razreda	frekvencija	relativna frekvencija
1	0.5 – 1.5	24	$\frac{24}{82}$
2	1.5 – 2.5	43	$\frac{43}{82}$
3	2.5 – 3.5	12	$\frac{12}{82}$
4	3.5 – 4.5	2	$\frac{2}{82}$
5	4.5 – 5.5	1	$\frac{1}{82}$

Na vertikalnu os nanesimo frekvencije ili ako želimo relativne frekvencije.



4.6. Poligon frekvencija (eng. FrequencyPolygon)

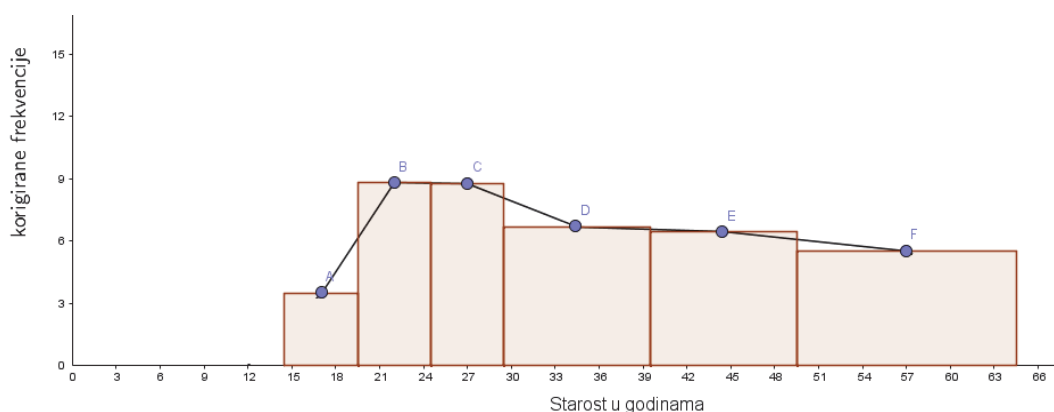
Poligon frekvencija je linijski grafikon koji također crtamo u pravokutnom koordinatnom sustavu, s istim mjerilima na okomitoj i horizontalnoj osi kao histogram. Ovaj grafikon nastaje tako da na sredini dužina koje predočavaju vrijednosti numeričkog obilježja ucrtamo

točke čije koordinate odgovaraju brojčanim vrijednostima obilježja i pripadajućim frekvencijama te ih potom spojimo linijama. Ako je prethodno nacrtan histogram, dovoljno je polovice vrhova stupaca (tj. sredine razreda- aritmetička sredina donje i gornje granice razreda) spojiti linijama da bi se kreirao poligon frekvencija.

Primjer 4.6.

Prikazat ćemo na istoj slici histogram i poligon frekvencija temeljem podataka o nezaposlenim osobama prijavljenim Hrvatskom zavodu za zapošljavanje, na dan **31. prosinca 2010.** godine. Izračunajmo širine pojedinog razreda, precizne granice, korigirane frekvencije, i sredine razreda.

Starost	Broj osoba	Precizne granice	Širina razreda	Korigirane frekvencije	Sredina razreda
15 – 19	17 455	14.5 – 19.5	5	3 491	17
20 – 24	44 031	19.5 – 24.5	5	8 806.2	22
25 – 29	43 932	24.5 – 29.5	5	8 786.4	27
30 – 39	66 955	29.5 – 39.5	10	6 695.5	34.5
40 – 49	64 512	39.5 – 49.5	10	6 451.2	44.5
50 i više	82 960	49.5 – 64.5	15	5 530.67	57



4.7. Kumulanta

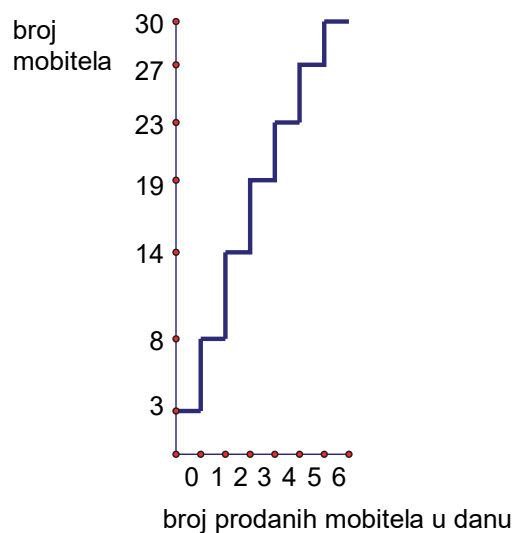
Kumulantom grafički prikazujemo kumulativne nizove. To je stepenasti linijski grafikonom koji se crta u pravokutnom koordinatnom sustavu, pri čemu na horizontalnoj osi nanosimo mjerilo za vrijednosti obilježja, a na okomitu os vrijednosti kumulativnog niza.

Primjer 4.7.

Tablica prikazuje broj prodanih mobitela u prodavaonici tijekom 30 dana.

Broj prodanih mobitela u danu	Broj mobitela	Kumulativni niz
0	3	3
1	5	$3+5=8$
2	6	$3+5+6=14$
3	5	$3+5+6+5=19$
4	4	$3+5+6+5+4=23$
5	4	$3+5+6+5+4+4=27$
6	3	$3+5+6+5+4+4+3=30$

U trećem stupcu izračunati su članovi kumulativnog niza. Nacrtajmo sada kumulantu.

**Zadaci4.**

1. Uzroci smrti za **700** osoba dani su u tablici. Prikaži podatke stupčastim dijagramom.

uzrok smrti	broj umrlih (frekvencija)
zatajenje rada srca	264
moždani udar	189
maligna bolest	79
plućne bolesti	81
ostalo	81

2. Tijekom akcije "Brinemo se za svoje zdravlje" medicinsko je osoblje na jednom gradskom trgu mjerilo tlak građanima. Podatke koje su dobili zapisali su u ovakvoj tablici frekvencija. Prikaži podatke pomoću višestrukog stupčastog dijagrama.

	normalan tlak	povišen tlak	ukupno
žene	36	44	80
muškarci	45	25	70
ukupno	81	69	150

3. Prošli tjedan u Zavodu za transfuziju 50 je osoba dobrovoljno darivalo krv. Svakom je darivatelju određena krvna grupa i dobiveni su sljedeći podaci:

tip krvne grupe	broj osoba koje imaju taj tip krvne grupe
0	27
A	11
B	9
AB	3

Izračunaj relativne frekvencije i prikaži podatke kružnim dijagramom.

4. Prikaži dijagramom s točkama podatke o visini u cm 10 učenica jednog razreda:

153 163 165 166 166 171 178 178 180 181.

5. Mjereno je vrijeme čekanja (u minutama) pojedinih stranaka u redu u jednoj poslovnoj banci:

6 8 10 11 12 20 22 25 34 37 42 55

Prikaži podatke Stablo-list dijagramom.

6. U Statističkom ljetopisu Republike Hrvatske za **2003.** godinu nalazi se ovakva tablica:

navršene godine života	broj osoba
15 – 19	25 617
20 – 24	61 802
25 – 29	49 661
30 – 39	86 440
40 – 49	85 322
50 i više	57 320

Izračunajte precizne granice razreda, širine razreda, sredine razreda te korigirajte frekvencije ako je potrebno. Prikažite podatke histogramom i poligonom frekvencija.

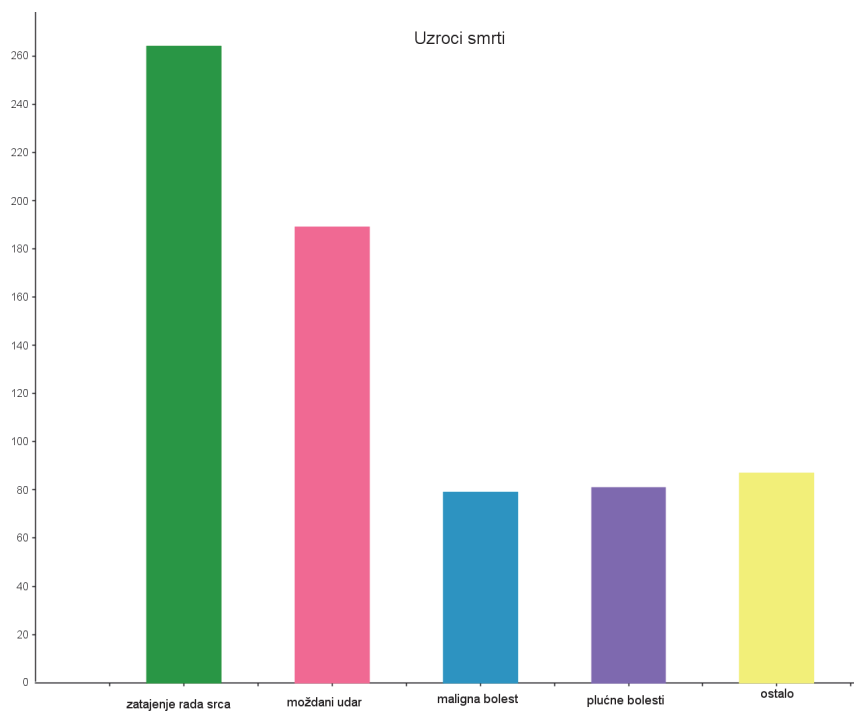
7. Zrakoplovna kompanija M-AIR ispituje učestalost otkaza rezervacija na liniji Zagreb – London, let AM457 u srpnju i kolovozu 2012. godine. Analiza je pokazala sljedeće:

Broj otkaza rezervacija	Broj letova
0	9
1	15
2	11
3	8
4	4
5	3

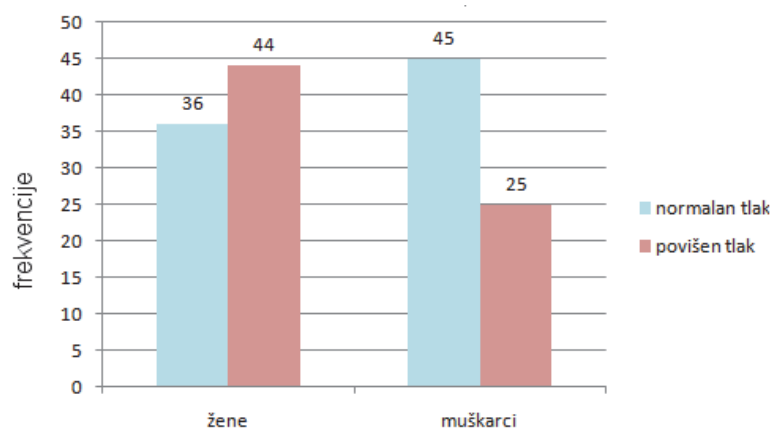
Sastavite kumulativni niz i prikažite ga grafički.

Rješenja 4.

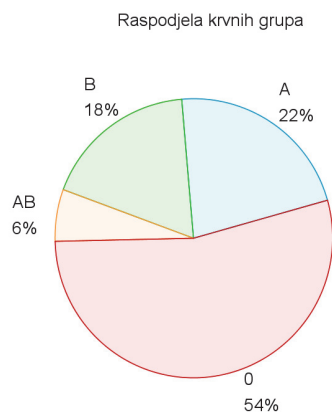
1.



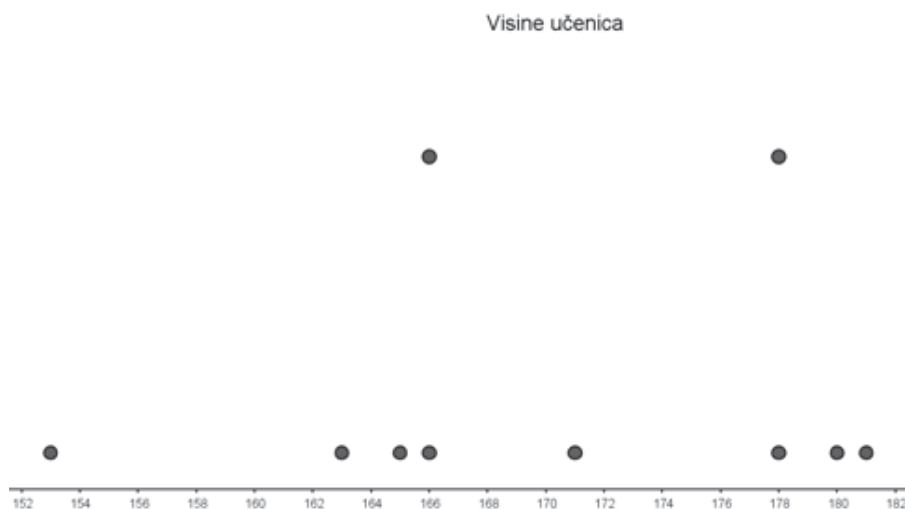
2.



3.



4.



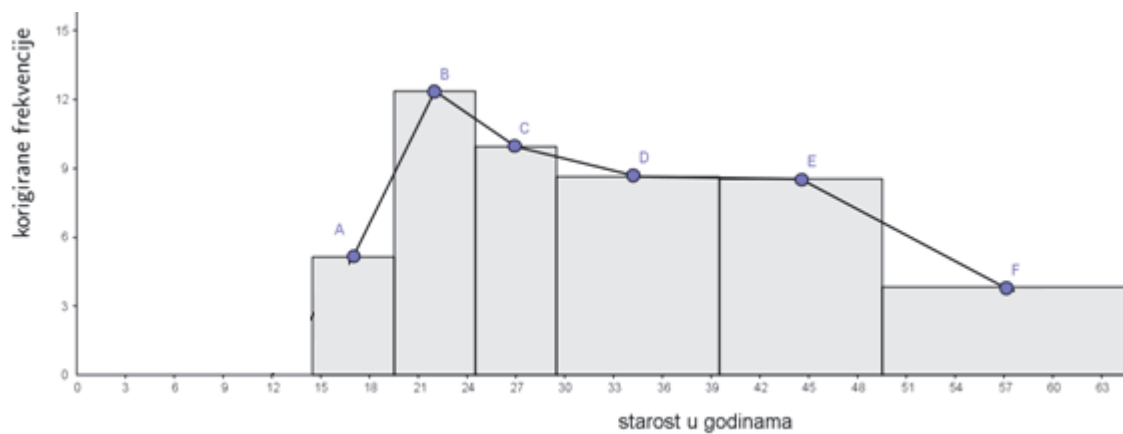
5.

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

Ključ : 3|1 znači 31

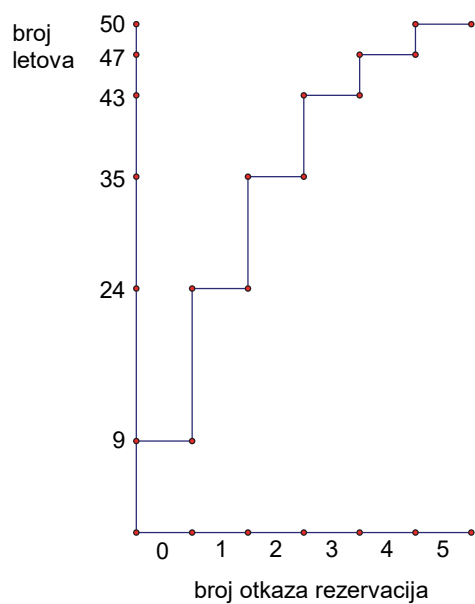
6.

navršene godine života	broj osoba	precizne granice	širina razreda	korigirane frekvencije	sredina razreda
15 – 19	25 617	14.5 – 19.5	5	5123.4	17
20 – 24	61 802	19.5 – 24.5	5	12 360.4	22
25 – 29	49 661	24.5 – 29.5	5	9 932.2	27
30 – 39	86 440	29.5 – 39.5	10	8 644	34.5
40 – 49	85 322	39.5 – 49.5	10	8 532.2	44.5
50 i više	57 320	49.5 – 64.5	15	3 821.33	57



7.

Broj otkaza rezervacija	Broj letova	Kumulativni niz
0	9	9
1	15	24
2	11	35
3	8	43
4	4	47
5	3	50



Sažetak

Statistički niz - nastaje uređivanjem podataka

Frekvencija- broj jedinica statističkog skupa koje imaju isto obilježje

Relativna frekvencija- količnik frekvencije i ukupnog broja elemenata statističkog skupa

Grupiranje- postupak kojim se elementi statističkog skupa razvrstavaju u razrede jednakih ili različitih širina

Korigirana frekvencija- količnik frekvencije i širine razreda

Niz kvalitativnih podataka- nastaje uređivanjem podataka koji ovise o jednoj kvalitativnoj varijabli

Vremenski niz – nastaje uređivanjem podataka koji čine vremenski niz

Numerički statistički niz – nastaje uređivanjem kvantitativnih podataka

Kumulativni niz- nastaje postupnim zbrajanjem frekvencija izvornog niza

Tabeliranje- tabelarno prikazivanje statističkih podataka

Statistička tablica- naslov, tekstualni dio, brojčani dio i izvor podataka

Grafičko prikazivanje podataka:

- *stupčasti i kružni* dijagram- primjereni za nizove kvalitativnih podataka
- *dijagram točaka, dijagram stablo-list*- primjereni za numeričke nizove s malim brojem podataka
- *histogram, poligon frekvencija*- primjereni za numeričke nizove s grupiranim podacima
- *kumulanta*-prikaz kumulativnog niza

III SREDNJE VRIJEDNOSTI STATISTIČKIH NIZOVA

Statističku analizu podataka uobičajeno započinjemo izračunom mjera srednjih vrijednosti ili centralne tendencije nizova. Razlikujemo potpune i položajne srednje vrijednosti.

Potpune srednje vrijednosti su aritmetička, geometrijska i harmonijska sredina.

U izračunu ovih srednjih vrijednosti sudjeluju sve vrijednosti odnosno podaci statističkog skupa.

Položajne srednje vrijednosti su mod i medijan i njihova vrijednost određena je mjestom (položajem) unutar numeričkog niza.

1. Aritmetička sredina (prosjeck)

To je osnovna i najvažnija mjera centralne tendencije. Svrha aritmetičke sredine je da se skup podataka zamijeni jednim brojem.

Definicija aritmetičke sredine ovisi o tome radi li se o negrupiranim ili grupiranim podacima.

Ako se radi o **negrupiranim podacima** izračunava se kao omjer zbroja svih vrijednosti numeričkog niza $x_1, x_2, \dots, x_n$ i broja članova niza n .

$$A_n(x) = \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Primjer 1.1.

Izračunajmo prosječnu težinu 6 nasumično odabranih učenika. Težina je dana u kilogramima, te su podaci poredani po veličini:

51 54 54 57 62 69

S obzirom da se ovdje radi o negrupiranim podacima, koristit ćemo formulu (1):

$$\bar{x} = \frac{51 + 54 + 54 + 57 + 62 + 69}{6} = \frac{347}{6} = 57.8\dot{3} \text{ kg}$$

Računamo li aritmetičku sredinu za **grupirane** podatke, tada se ona naziva **vagana, težinska ili ponderirana aritmetička sredina**.

Ako numerička varijabla poprima različite vrijednosti $x_1, x_2, \dots, x_n$ s frekvencijama $f_1, f_2, \dots, f_n$

(broj x_1 ponavlja se f_1 puta itd.) izračun aritmetičke sredine glasi:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{N} = \frac{\sum_{i=1}^n f_i x_i}{N} = \frac{1}{N} \sum_{i=1}^n f_i x_i, \quad N = \sum_{i=1}^n f_i$$

Primjer 1.2.

Kontrola proizvoda vrši se uzimanjem uzorka od 20 proizvoda i zapisivanjem broja neispravnih među njima. Pregledano je 18 uzoraka i dobiveni su podaci:

3 0 0 3 1 0 2 1 0
0 4 0 0 2 2 0 0 0

x_i	f_i
0	10
1	2
2	3
3	2
4	1

n (ukupni broj razreda) = 5

N (ukupan broj podataka) = 18

x_i su vrijednosti obilježja (modaliteti)

f_i su apsolutne frekvencije

Aritmetička sredina je:

$$\bar{x} = \frac{1}{18} (0 \cdot 10 + 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 2 + 4 \cdot 1) = \frac{1}{18} \cdot 18 = 1$$

Treba naglasiti da $f_1, f_2, \dots, f_n$ predstavljaju težine (pondere) od kuda i naziv ponderirana težina.

Primjer 1.3.

Izračunajmo vaganu aritmetičku sredinu na primjeru visine uzorka učenika u jednoj srednjoj školi:

Visina u cm	Razredna sredina	Broj učenika	Podtotali
X_i	x_i	f_i	$f_i x_i$
162-165	163,5	6	981
166-169	167,5	4	670
170-173	171,5	6	1029
174-177	175,5	0	0
178-181	179,5	2	359
182-185	183,5	2	367
Ukupno		20	3406

$$\bar{x} = \frac{1}{20} (6 \cdot 163,5 + 4 \cdot 167,5 + 6 \cdot 171,5 + 0 \cdot 175,5 + 2 \cdot 179,5 + 2 \cdot 183,5) = 170,3$$

Općenito; ako je širina razreda konstantna i iznosi c , sredine razreda mogu se pisati u obliku

$$x_i = x_1 + (i-1) \cdot c \quad i = 1, \dots, n$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{1}{N} \sum_{i=1}^n f_i [x_1 + (i-1) \cdot c] = \frac{1}{N} \sum_{i=1}^n f_i x_1 + \frac{1}{N} \sum_{i=1}^n f_i (i-1) \cdot c = \frac{x_1}{N} \sum_{i=1}^n f_i + \frac{c}{N} \sum_{i=1}^n f_i (i-1) = x_1 + \frac{c}{N} \sum_{i=1}^n f_i (i-1)$$

U našem primjeru je

$$x_i = x_1 + (i-1) \cdot 4 \quad i = 1, \dots, 6$$

$$x_1 = 163,5 + 0 \cdot 4$$

$$x_2 = 163,5 + 1 \cdot 4$$

$$x_3 = 163,5 + 2 \cdot 4$$

$$x_4 = 163,5 + 3 \cdot 4$$

$$x_5 = 163,5 + 4 \cdot 4$$

$$x_6 = 163,5 + 5 \cdot 4$$

$$\bar{x} = 163,5 + \frac{4}{20} (6 \cdot 0 + 4 \cdot (2-1) + 6 \cdot 2 + 0 \cdot 3 + 2 \cdot 4 + 2 \cdot 5) = 170,3$$

Možemo zaključiti kako je u odabranom uzorku srednje škole prosječna visina 170,3cm.

Za aritmetičku sredinu je karakteristično da je suma svih odstupanja jednaka 0.

$$\sum_{i=1}^n f_i(x_i - \bar{x}) = 0$$

$$\sum_{i=1}^n f_i(x_i - \bar{x}) = \sum_{i=1}^n f_i x_i - \sum_{i=1}^n f_i \cdot \bar{x} = \frac{N}{N} \sum_{i=1}^n f_i x_i - \bar{x} \sum_{i=1}^n f_i = N \cdot \bar{x} - \bar{x} \cdot N = 0 \quad (\bar{x} = \frac{1}{n} \sum_{i=1}^n f_i x_i)$$

Primjer 1.4

Prikažimo izračun vagane aritmetičke sredine na primjeru broja djece po obitelji u jednome gradu, kao primjer numeričke diskretne varijable. Podatci o broju djece po obitelji dani su u sljedećoj tablici:

Broj djece po obitelji	Broj obitelji
0	502
1	2.194
2	3.466
3	3.827
4	5.326
5	3.832
6	3.060
7	916
Ukupno	22.123

Kako bismo mogli izračunati aritmetičku sredinu, na istom primjeru možemo prikazati izračun korištenjem frekvencija f_i kao i relativnih frekvencija p_i ili relativnih postotnih frekvencija P_i .

Broj djece po obitelji	Broj obitelji	Relativna frekvencija		Podtotali		
x_i	f_i	p_i	P_i	$f_i x_i$	$p_i x_i$	$P_i x_i$
0	502	0,023	2,27	0	0	0
1	2.194	0,099	9,92	2,194	0,099	9,917
2	3.466	0,157	15,67	6,932	0,313	31,334
3	3.827	0,173	17,30	11,481	0,519	51,896
4	5.326	0,241	24,07	21,304	0,963	96,298
5	3.832	0,173	17,32	19,160	0,866	86,607
6	3.060	0,093	9,31	12,360	0,559	55,869
7	916	0,041	4,14	6,412	0,290	28,983
Ukupno	22. 123	1	100	79,843	3,609	360,9

Primjerice, izračun relativne frekvencije p_1 i P_1 izvršili smo na sljedeći način:

$$p_1 = \frac{502}{22.123} = 0,023$$

$$P_1 = \frac{502}{22.123} \cdot 100 = 2,27$$

Izračun aritmetičke sredine dan je kao:

$$\bar{x} = \frac{\sum_{i=1}^N f_i x_i}{N} = \frac{502 \cdot 0 + 2.194 \cdot 1 + \dots + 916 \cdot 7}{502 + 2.194 + \dots + 916} = \frac{79.843}{22.123} = 3,609$$

$$\bar{x} = \sum_{i=1}^N p_i x_i = 0 + 0,099 + \dots + 0,290 = 3,609$$

$$\bar{x} = \frac{\sum_{i=1}^N P_i x_i}{100} = \frac{0 + 9,917 + \dots + 28,983}{100} = 3,609.$$

Interpretacija aritmetičke sredine u ovom slučaju glasi: prosječan broj djece po obitelji u jednome gradu iznosi 3,609 djece. Do identičnog rezultata došli smo na tri načina, koristeći apsolutne frekvencije, kao i relativne i relativne postotne.

2. Geometrijska i harmonijska sredina

Geometrijska sredina

- a. **za negrupirani niz** $x_1, x_2, x_3, \dots, x_n$ geometrijska sredina definira se kao n -ti korijen iz umnoška vrijednosti varijable

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$$

$$G_n(x) = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

- b. **za grupirani niz** je $G_n = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$

Geometrijska sredina ima smisla za pozitivne vrijednosti numeričke varijable tj. $x_i > 0$, za $i = 1, 2, \dots, k$. Može se koristiti kao mjera prosječne brzine promjene nekih pojava u vremenu.

Neka je $x_1, x_2, \dots, x_n$ vremenski statistički skup i x_i označava stanje prirodne pojave u određenom vremenskom intervalu.

Tada:

$$y_1 = \frac{x_2}{x_1}, y_2 = \frac{x_3}{x_2}, \dots, y_{n-1} = \frac{x_n}{x_{n-1}}$$

predstavlja brzine promjena te pojave u pojedinim vremenskim intervalima, a njihova geometrijska sredina

$$y_G = \sqrt[n-1]{y_1 \cdot y_2 \cdot \dots \cdot y_{n-1}} = \sqrt[n-1]{\frac{x_n}{x_1}}$$

predstavlja srednju brzinu promjene u cijelom vremenskom periodu u kojem se promatra pojava.

Primjer 2.1.

Promet robe u jednom skladištu u prvoj godini iznosi 4000 t, u drugoj 18000 t, a u trećoj 36000 t. Vidimo da je promet u drugoj godini 4,5 puta veći nego u prvoj, a u trećoj 2 puta veći nego u drugoj godini. Prosječni porast prometa u promatranom razdoblju iznosi:

$$y_G = \sqrt{y_1 \cdot y_2} = \sqrt{\frac{18000}{4000} \cdot \frac{36000}{18000}} = \sqrt{9} = 3$$

Ovaj rezultat odgovara stvarnom stanju, jer je promet od prve do treće godine povećan 9 puta.

Aritmetička sredina

$$\frac{4,5 + 2}{2} = 3,25$$

daje prosječan rast prometa od 3,25 koji očito ne izražava prosječnu promjenu prometa dobro kao geometrijska sredina.

Harmonijska sredina

- a. **za negrupirani niz** $x_1, x_2, x_3, \dots, x_n$ ($x_i \neq 0$) harmonijsku sredinu H_n računamo po formuli:

$$H_n(x) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

b. za grupirani niz je:

$$H_n(x) = \frac{n}{\sum_{i=1}^k \frac{f_i}{x_i}}$$

Harmonijsku sredinu valja upotrebljavati kad želimo dobiti npr. prosjek nekih odnosa, prosječnu brzinu...

Primjer 2.2.

Četiri radnika obavljaju istovrsnu uslugu. Radnik A izvrši tu uslugu za 10 minuta, radnik B za 6 minuta, radnik C za 5 minuta, a radnik D za 4 minute.

a. Ako polazimo od aritmetičke sredine onda i -ti radnik izvrši $\frac{\bar{x}}{x_i}$ usluge, gdje je

$$\bar{x} = \frac{25}{4} = 6,25 \text{ min prosječno vrijeme za pružanje te usluge.}$$

$$\text{A izvrši } \frac{6,25}{10} = 0,625 \text{ usluge}$$

$$\text{B izvrši } \frac{6,25}{6} = 1,0417 \text{ usluge}$$

$$\text{C izvrši } \frac{6,25}{5} = 1,25 \text{ usluge}$$

$$\text{D izvrši } \frac{6,25}{4} = 1,5625 \text{ usluge}$$

Kada zbrojimo dobivamo 4,47 usluge (umjesto 4).

$$\text{b. Harmonijska sredina iznosi: } x_H = \frac{4}{\frac{1}{10} + \frac{1}{6} + \frac{1}{5} + \frac{1}{4}} = 5,5814.$$

Sada i -ti radnik izvrši $\frac{x_H}{x_i}$ usluge:

$$\text{A izvrši } \frac{5,5814}{10} = 0,5581 \text{ usluge}$$

$$\text{B izvrši } 0,9302 \text{ usluge}$$

$$\text{C izvrši } 1,1163 \text{ usluge}$$

$$\text{D izvrši } 1,3954 \text{ usluge.}$$

Zbroj svih usluga sada iznosi 4.

Primjer 2.3.

Tri kućanice su u anketi iskazale koliko u njihovoj obitelji traje staklenka marmelade od 1 kg: 5 dana, 10 dana i 15 dana. Koliko prosječno traje staklenka marmelade?

Ne traje 10 dana, jer bi za 30 dana ukupno sva tri domaćinstva potrošila $6+3+2=11$ staklenki, pa 1 kg traje za sva tri domaćinstva $\frac{30}{11} = 2,72$ dana.

Jednom bi domaćinstvu kilogram prosječno trajao $3 \cdot 2,72 = 8,2$ dana, što se i dobiva računanjem harmonijske sredine:

$$x_H = \frac{3}{\frac{1}{5} + \frac{1}{10} + \frac{1}{15}} = 8,2$$

Može se pokazati da vrijedi $H \leq G \leq A$

Teorem

Ako su $x_1 > 0, x_2 > 0, \dots, x_n > 0$ tada vrijedi:

$$\min\{x_1, x_2, \dots, x_n\} \leq H_n(x) \leq G_n(x) \leq A_n(x) \leq \max\{x_1, x_2, \dots, x_n\}$$

Primjer 2.4.

Odredimo srednje vrijednosti za brojeve 2, 4, 8.

$$\bar{x} = \frac{2+4+8}{3} = 4,67$$

$$x_G = \sqrt[3]{2 \cdot 4 \cdot 8} = 4$$

$$x_H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = 3,43$$

Vidimo da je $3,43 \leq 4 \leq 4,67$.

3. Mod. Medijan.

3.1. Mod

Mod i medijan nazivamo položajne mjere centralne tendencije jer ovise o samom položaju podataka u statističkom nizu.

Mod je vrijednost obilježja s najvećom frekvencijom (najčešće se pojavljuje u nizu) i on postoji ako u nizu podataka postoje barem dva jednaka podatka.

Mod nije jedinstven broj za niz podataka, tj. skup može imati više modova, ali ne mora imati nijedan.

Primjer 3.1.1

- a. 66,67,67,67,72,72,72,80,80

Ovaj niz ima dva moda $M_{o1} = 67$ i $M_{o2} = 72$ (skup se zove bimodalni).

- b. 34,56,66,77,98,99

Ovaj niz nema moda.

Ako razmatramo distribuciju frekvencija, modalni razred je onaj koji ima najveću frekvenciju ako je veličina razreda jednaka, odnosno onaj koji ima najveću korigiranu frekvenciju ako je veličina razreda nejednaka. Mod se izračunava pomoću korigiranih frekvencija ovako:

$$M_o = L_1 + \frac{b-a}{(b-a)+(b-c)} \cdot i \quad (1)$$

gdje je M_o mod, L_1 donja granica modalnog razreda, b najveća (korigirana) frekvencija (frekvencija modalnog razreda), a (korigirana) frekvencija ispred frekvencije modalnog razreda, c (korigirana) frekvencija iza frekvencija modalnog razreda te je i veličina modalnog razreda.

Primjer 3.1.2. (niz kvalitativnih podataka)

Prikupljeni su sljedeći podaci o broju parova i njihovoj djeci:

broj djece/ x_i	0	1	2	3	4	5	6	7	8	9	10	11	12
broj parova/ f_i	70	90	108	86	70	47	30	20	15	5	4	3	2

Odredimo M_o i $\bar{x}$.

Najveća frekvencija je 108, te je $M_o = 2$. Aritmetička sredina je $\bar{x}$ je 3,03.

Pri gradnji stanova za prosječnu obitelj s troje djece učinila bi se pogreška.

Primjer 3.1.3. (grupirani podaci-numerička kontinuirana varijabla)

Radnici poduzeća X prema godinama starosti dani su u tablici:

Godine starosti x_i	Broj radnika f_i	Sredina razreda $\bar{x}$	Veličina razreda i	Korigirane frekvencije $f_{ci} = \frac{f_i}{i}$	
18-20	10	19	2	5	a
20-22	15	21	2	7,5	b
22-28	15	25	6	2,5	c
28-32	8	30	4	2	
Σ	48				

Najveća korigirana frekvencija je $f_{ci}=7,5$ za godine starosti 20-22.

Zato je u tom retku pridodana oznaka **b**.

$$M_o = L_1 + \frac{b-a}{(b-a)+(b-c)} \cdot i = 20 + \frac{7,5-5}{(7,5-5)+(7,5-2,5)} \cdot 2 = 20,66 \approx 21$$

"Najčešća" dob radnika poduzeća X je 21 godina, dok je $\bar{x}=23,33$.

Svojstva moda:

- određen je svojim položajem u nizu i na njega ne utječu ni izrazito velike, ni izrazito male vrijednosti obilježja kao što je slučaj kod aritmetičke sredine
- ima nedostatak da je osjetljiv na način grupiranja razreda (određivanje broja i širine razreda).

Ako neki niz sadrži samo jedan mod naziva se **unimodalan niz**, dok nizove s više od jednim modom nazivamo **multimodalni**.

3.2. Medijan

Medijan je položajna srednja vrijednost koja statistički niz dijeli na **dva jednakobrojna dijela**.

Niz je najprije potrebno urediti po veličini od najmanje do najveće vrijednosti, a zatim odrediti medijalnu vrijednost. Ako se radi o neparnom broju podataka u nizu, medijalna je vrijednost ona koja pripada središnjem članu tog niza. U protivnom, ako je niz sastavljen od parnog broja vrijednosti, medijan je dan kao poluzbroj vrijednosti središnja dva člana niza.

Za medijan vrijedi da je 50% podataka manje ili jednako toj vrijednosti, a 50% podataka je veće ili jednako njemu.

Primjer 3.2.1.

Odredimo medijan nizova:

- a. 35,38,45,47,59,62,70

Niz ima 7 brojeva. Sredina niza je četvrti broj. Dakle, medijan je 47.

- b. 35,38,45,47,59,62

Ovaj niz ima 6 članova i sredina se nalazi između trećeg i četvrtog člana.

Stoga je medijan : $M_e = \frac{45 + 47}{2} = 46$.

Zaključimo:

Kad imamo niz negrupiranih podataka s **neparnim brojem članova** tada je medijan jednak r -tom članu niza $M_e = x_r$ gdje je $r = \frac{N+1}{2}$ redni broj podatka a N broj članova niza.

Kad imamo niz negrupiranih podataka s **parnim brojem članova** tada je $M_e = \frac{x_{r_1} + x_{r_2}}{2}$,

gdje x_{r_1} i x_{r_2} predstavljaju dva središnja člana niza, pri čemu je $r_1 = \frac{1}{2}N$ i $r_2 = \frac{1}{2}N + 1$.

(rj: $M_e = 85\text{km}$ (mjesto na pola puta između 60-og i 110-og km.))

Primjer 3.2.6.

Domaćinstvo prema broju članova na nekom području raspoređena su kako slijedi

Broj članova	Broj domaćinstva
1	30
2	60
3	70
4	50
5	20
6-9	40

Proširena tablica izgleda ovako:

x_i	f_i	korigirane frekvencije	kumulativni niz
1	30	30	30
2	60	60	90
3	70	70	160
4	50	50	210
5	20	20	230
6-9	40	40:10=4	270
Σ	$N=270$		

- a. Izračunajmo koliko iznosi broj članova koji dijeli razdiobu na dva dijela.

Određujemo medijan:

redni broj središnjih elemenata za $N=270$ (paran) jednak je:

$$\frac{N}{2} = \frac{270}{2} = 135, \quad \frac{N}{2} + 1 = 136,$$

pa slijedi da je $M_e = 3$.

Znači polovina domaćinstva ima 3 člana ili manje, dok polovina ima 3 člana ili više.

- b. Izračunajmo najčešći broj članova u jednom domaćinstvu.

Najveća korigirana frekvencija je 70, pa je $M_o = 3$.

Znači najčešći broj članova domaćinstva je 3.

U slučaju grupiranih podataka najprije je potrebno izračunati frekvencije kumulativnog niza "manje od", u kojem se traži središnji član. Potrebno je pronaći razred u kojem je sadržana

frekvencija koja sadrži $\frac{N}{2}$, te ga nazivamo **medijalni razred**. Ako su veličine svih razreda nekog obilježja jednake 1, vrijednost obilježja koja sadrži medijalni razred je medijan.

U slučaju razrednih veličina različitih od 1, medijan računamo kao:

$$M_e = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^m f_i}{f_{med}} \cdot i, \quad (2)$$

gdje su:

L_1 donja granica medijalnog razreda,

$\frac{N}{2}$ polovica članova niza,

$\sum_{i=1}^m f_i$ zbroj svih frekvencija do medijalnog razreda (ne uključujući medijalni razred),

f_{med} apsolutna frekvencija medijalnog razreda, te

i veličina medijalnog razreda.

Primjer 3.2.7.

Izračunajmo mod i medijan na primjeru prodaje modemske žice u jednoj trgovačkom centru tijekom 2012.godine. U tablici prikazana je distribucija prodanih žica prema duljini.

Duljina modemske žice u mm	Broj prodanih žica	Veličina razreda	Korigirane frekvencije	Kumulativni niz
X_i	f_i	i_i	f_{ci}	k_i
100-200	150	100	1,5	150
200-300	200	100	2	350
300-400	300	100	3	650
400-550	70	150	0,47	720
550-700	40	150	0,27	760
Ukupno	760	-	-	-

Mod računamo po formuli (1):

$$M_o = 300 + \frac{3 - 2}{(3 - 2) + (3 - 0,47)} \cdot 100 = 328,33$$

Najčešća duljina modemske žice prodane u trgovačkom centru iznosi 328,33 mm.

Medijan računamo formulom (2):

$$M_e = 300 + \frac{380 - 350}{300} \cdot 100 = 310$$

Od ukupnog broja prodanih modernih žica u trgovačkom centru tijekom 2012.godine, polovica prodanih žica bilo je duljine 310 mm ili kraće, a druga polovica 310 mm ili dulje.

Svojstva medijana:

- jedinstven je i $x_{\min} \leq M_e \leq x_{\max}$
- na njegovu vrijednost ne utječu ekstremne vrijednosti obilježja
- primjenjuje se u slučaju izrazito asimetričnih razdioba

Pregled mogućnosti izbora srednjih vrijednosti s obzirom na vrstu varijable (obilježja)

Vrsta varijable (obilježja)	Sredina				
	M_0	M_e	$\bar{x}$	G	H
nominalna	da	ne	ne	ne	ne
redoslijedna	da	da	ne	ne	ne
numerička	da	da	da	da	da

Zadaci 3.

- Nabrojite srednje vrijednosti statističkih nizova.
- Za koje srednje vrijednosti kažemo da su potpune mjere i zašto, te za koje kažemo da su položajne mjere i zašto?
- Definirajte aritmetičku sredinu. Kako ju računamo za negrupirane, a kako za grupirane podatke?
- Definirajte mod. Kako mod računamo za grupirane, a kako za negrupirane podatke?
- Definirajte medijan. Što je potrebno učiniti s nizom podataka prije izračuna medijana? Kako računamo medijan za grupirane podatke, a kako za negrupirane podatke?
- Izračunaj aritmetičku sredinu i medijan niza:
 - 2,4,8,8,8,10,14;
 - 3,3,2.1,8.3,9.7,10,10;
 - 12,13,14,14,15,18,19,19.
- Aritmetička sredina 10 podataka iznosi 91. Ako iz skupa podataka izbacimo podatak 100, kolika je aritmetička sredina preostalih podataka?
- Aritmetička sredina 15 podataka je 48, a ako iz tog skupa izuzmemo jedan podatak, aritmetička sredina preostalih podataka je 50. Koji je podatak izuzet?

9. Autobus je 3 sata vozio prosječnom brzinom od 85 km/h, a zatim 2 sata prosječnom brzinom od 90 km/h. Kolika je bila prosječna brzina autobusa tijekom tih 5 sati putovanja?
10. Analizirajmo prosječno vrijeme čekanja u redu u jednoj poslovnoj banci, te su dani podaci o čekanju pojedinih stranaka u minutama:

6 8 10 12 20 22 25 34 37 42 55

Izračunajte aritmetičku, geometrijsku i harmonijsku sredinu, a potom mod i medijan. Interpretirajte svaku izračunatu vrijednost.

11. Tijekom sedam dana u lipnju 2010. godine izmjerene su sljedeće temperature u Krapini (u stupnjevima Celzijevim): 25, 22, 29, 28, 27 i 30. Izračunajte aritmetičku, geometrijsku i harmonijsku sredinu te ih interpretirajte.
12. Izračunajte mod i medijan za podatke u prethodnom zadatku te ih interpretirajte.
13. Izračunajte prosječan broj prodanih mobitela u jednoj prodavaonici temeljem podataka danih u tablici. Potom izračunajte mod i medijan i sve interpretirajte.

Broj prodanih mobitela u danu	Broj dana
x_i	f_i
0	3
1	5
2	6
3	5
4	4
5	4
6	3
Ukupno	30

Rješenja 3.

6. a) $\bar{x} = 7.9$, $M_e = 8$

b) $\bar{x} = 6.59$, $M_e = 8.3$

c) $\bar{x} = 15.5$, $M_e = 14.5$

7. $91 \cdot 10 - 100 = 810, \bar{x} = \frac{810}{9} = 90$

8. $15 \cdot 48 = 14 \cdot 50 + y, y = 20$

9. $\bar{x} = \frac{3 \cdot 85 + 2 \cdot 90}{5} = 87 \text{ km/h}$

$$10. \bar{x} = \frac{6+8+\dots+55}{12} = 23,5 \text{ min}$$

$$G = \sqrt[12]{6 \cdot 8 \cdot \dots \cdot 55} = 18,86 \text{ min}$$

$$H = \frac{12}{\frac{1}{6} + \frac{1}{8} + \dots + \frac{1}{55}} = 15 \text{ min}$$

Mod ne postoji u nizu. Za izračun medijana potrebno je niz urediti po veličini. Paran je broj članova u nizu pa je $M_e = \frac{20+22}{2} = 21$ minuta (polovica stranaka čekalo je u redu 21 min ili kraće, a druga polovica 21 min ili dulje).

$$11. \bar{x} = 27,14 \text{ C}; \quad G = 27,01 \text{ C}; \quad H = 26,87 \text{ C}$$

12. $M_o = 29^\circ \text{C}$, $M_e = 28^\circ \text{C}$ (najčešća temperatura zraka bila je 29; polovica dana imalo je temperaturu 28 ili niže, a polovica 28 ili više).

13.

Broj prodanih mobitela u danu	Broj dana	Podtotali	Kumulativni niz
x_i	f_i	$f_i x_i$	k_i
0	3	0	3
1	5	5	8
2	6	12	14
3	5	15	19
4	4	16	23
5	4	20	27
6	3	18	30
Ukupno	30	86	-

$$\bar{x} = \frac{86}{30} = 2,87, \quad M_e = 3, \quad M_o = 3$$

Sažetak

Aritmetička sredina – najvažnija i najraširenija srednja vrijednost; određuje se tako da se zbroje vrijednosti numeričke varijable i podijele s njihovim brojem

Svojstva:

- zbroj odstupanja vrijednosti numeričke varijable od njezine aritmetičke sredine jednak je nuli
- zbroj kvadrata odstupanja vrijednosti numeričke varijable od njezine sredine minimalan je

- aritmetička sredina nalazi se između najmanje i najveće vrijednosti niza za koji je izračunana

Aritmetička sredina aritmetičkih sredina – izračunava se kao vagana sredina u kojoj se za pondere uzima broj podataka za koje su računane pojedine sredine.

Aritmetička sredina relativnih brojeva koordinacije i aritmetička sredina postotaka – izračunavaju se kao vagane sredine u kojima su ponderi osnove tih brojeva

Geometrijska sredina

Geometrijska sredina n vrijednosti numeričke varijable x :

- za negrupirani niz** $x_1, x_2, x_3, \dots, x_n$ definira se kao n -ti korijen iz umnoška vrijednosti varijable $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$

$$G_n(x) = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

- za grupirani niz** je $G_n = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$

Harmonijska sredina

Harmonijska sredina n vrijednosti numeričke varijable x recipročna je vrijednosti aritmetičke sredine njezinih recipročnih vrijednosti

- za negrupirani niz** $x_1, x_2, x_3, \dots, x_n$ ($x_i \neq 0$) harmonijsku sredinu H_n računamo po formuli:

$$H_n(x) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

- za grupirani niz** je

$$H_n(x) = \frac{n}{\sum_{i=1}^k \frac{f_i}{x_i}}$$

Mod

- položajna srednja vrijednost; najčešća vrijednost ili modalitet koji se pojavljuje u nizu; postoji ako su u nizu barem dva jednaka podatka
- mod distribucije frekvencija s razredima aproksimira se pomoću izraza:

$$M_o = L_1 + \frac{b-a}{(b-a)+(b-c)} \cdot i ,$$

gdje je:

M_o mod,

L_1 donja granica modalnog razreda,

b najveća (korigirana) frekvencija (frekvencija modalnog razreda),

a (korigirana) frekvencija ispred frekvencije modalnog razreda,

c (korigirana) frekvencija iza frekvencija modalnog razreda, te

i veličina modalnog razreda.

Medijan

- položajna srednja vrijednost koja numerički niz uređen po veličini dijeli na dva jednaka dijela
- ako je broj podataka neparan, medijan je vrijednost središnjeg člana uređenog po veličini, a ako je broj podatak paran, medijan je jednak poluzbroju vrijednosti varijable središnjih dvaju članova uređenog niza
- medijan u distribuciji frekvencija s razredima aproksimira se pomoću izraza:

$$M_e = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^m f_i}{f_{med}} \cdot i ,$$

gdje je:

L_1 donja granica medijalnog razreda,

$\frac{N}{2}$ polovica članova niza,

$\sum_{i=1}^m f_i$ zbroj svih frekvencija do medijalnog razreda (ne uključujući medijalni razred),

f_{med} apsolutna frekvencija medijalnog razreda, te

i veličina medijalnog razreda (medijalni je onaj razred čija kumulativna frekvencija prvi put uključuje vrijednost $N/2$).

IV MJERE RASPRŠENOSTI (DISPERZIJE)

Disperzija je pojam za raspršenost članova numeričkog niza od neke srednje vrijednosti.

Mjere disperzije su veličine kojim se utvrđuje veličina raspršenosti članova numeričkog niza od neke srednje vrijednosti, odnosno utvrđuje reprezentativnost srednjih vrijednosti. Mala vrijednost pokazatelja disperzije znači da je izračunata srednja vrijednost dobar reprezentant skupa podataka i obrnuto.

Najpoznatije mjere disperzije dijelimo u dvije skupine, apsolutne mjere disperzije i relativne mjere disperzije

1.1. Apsolutne mjere disperzije

Apsolutne mjere disperzije su izražene u istim jedinicama mjere kao i numeričko obilježje.

To su:

- a. raspon varijacije
 - b. interkvartil
 - c. varijanca
 - d. standardna devijacija
- a. **Raspon varijacije** (R) je razlika između najveće i najmanje vrijednosti numeričkog obilježja:

$$R_x = x_{\max} - x_{\min}$$

Raspon uvijek izražavamo u mjernim jedinicama varijable a koristi se u praćenju cijena roba i usluga, tečaja valuta i slično. To je najjednostavnija mjera raspršenosti koja ovisi samo o dvije vrijednosti (najvećoj i najmanjoj vrijednosti niza), a budući da to često nisu tipične vrijednosti, računanje raspona se preoblikuje tako da se pri računanju izostavi dio podataka na početku i na kraju niza vrijednosti uređenog po veličini.

Izostave li se prva i posljednja četvrtina članova uređenog niza, ostaje središnjih 50% podataka i taj raspon varijacije je mjera raspršenosti koja se naziva interkvartil.

b. **Interkvartil** ($Q_3 - Q_1$)

Interkvartil je razlika između gornjeg Q_3 i donjeg kvartila Q_1 .

To je mjera disperzije kojom se utvrđuje reprezentativnost medijana kao srednje vrijednosti.

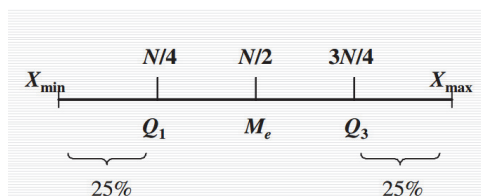
Interkvartil obuhvaća 50 % podataka, a isključuje 25 % prvih i 25 % posljednjih koji su često atipični. Da bismo razumjeli pojam interkvartila moramo objasniti kvantile i kvartile.

Kvantili

Kvantili označavaju vrijednosti numeričkog obilježja koje niz (uređen prema veličini), dijele na q jednakih dijelova. Kvantili dobivaju naziv prema broju dijelova.

Ako je:

- $q=4$ onda imamo kvartile (donji, srednji i gornji)
- $q=10$ onda imamo decile
- $q=100$ onda imamo centile ili percentile. **Percentili** dijele distribuciju rezultata na sto jednakih dijelova i svaki dio obuhvaća 1 % rezultata. Prvi centil obuhvaća 1% rezultata, drugi 2 % rezultata itd.



Donji kvartil (Q_1) (kažemo još prvi kvartil ili 25%) je vrijednost koja se nalazi na središnjem mjestu polovice podataka nižih od medijana.

$$Q_1 = x_{\left[\frac{N}{4}\right]+1}$$

gdje je $\left[\frac{N}{4}\right]$ cijeli dio tog razlomka (npr. $\left[\frac{5}{4}\right] = \lfloor 1.25 \rfloor = 1$; $\left[\frac{19}{4}\right] = \lfloor 4.75 \rfloor = 4$).

Kad je broj članova niza N djeljiv s četiri bez ostatka, donji je kvartil:

$$Q_1 = \frac{1}{2} \left[x_{\frac{N}{4}} + x_{\frac{N}{4}+1} \right]$$

(poluzbroj podataka s rednim brojevima $N/4$ i $N/4+1$).

Srednji (drugi) kvartil je medijan $Q_2 = M_e$

Gornji kvartil (Q_3) (treći kvartil ili 75%) je vrijednost koja se nalazi na središnjem mjestu polovice podataka viših od medijana tj. vrijednost varijable s rednim brojem $\left\lfloor \frac{3N}{4} \right\rfloor + 1$

$$Q_3 = x_{\left\lfloor \frac{3N}{4} \right\rfloor + 1}, \text{ gdje } \frac{3N}{4} \notin Z$$

$$Q_3 = \frac{1}{2} \left[x_{\frac{3N}{4}} + x_{\frac{3N}{4} + 1} \right] \text{ kada je } \frac{3N}{4} \in Z$$

Iz prethodnog proizlazi da je raspon između donjeg i gornjeg kvartila interval unutar kojeg se kreće vrijednost numeričkog obilježja za 50% jedinica u nizu i da se između Q_1 i Q_3 nalazi medijan. Što je raspon između donjeg i gornjeg kvartila manji, medijan je reprezentativniji jer je zgusnutost oko medijana veća, i obrnuto.

U distribuciji frekvencija po razredima, vrijednosti kvartila potrebnih za računanje interkvartila aproksimiraju se slično kao medijan.

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_1}{f_{k \text{ var}}} \cdot i$$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum f_1}{f_{k \text{ var}}} \cdot i$$

.

U danim formulama oznake su:

N je zbroj frekvencija (apsolutnih ili relativnih)

$F_{k \text{ var}}$ je frekvencija kvartilnog razreda – razreda čija kumulativna frekvencija prvi put

uključuje vrijednosti $N/4$ odnosno $3N/4$

L_1 je donja granica kvartilnog razreda

$\sum f_1$ je zbroj frekvencija do kvartilnog razreda,

i je veličina kvartilnog razreda.

Primjer 1.1.1.

Tijekom jednog tjedna prodaja određene knjige u svim knjižarama u gradu zabilježena je na slijedeći način:

1 3 8 9 6 11 12 2 3 4 7 8 7 3.

Razvrstajmo ovaj negrupirani niz i odredimo raspon varijacije, kvartile i interkvartil.

Razvrstajmo niz:

1 2 3 3 3 4 6 7 7 8 8 9 11 12

Vidimo da je $N=14$.

Izračunajmo donji kvartil:

$$Q_1 = x_{\left[\frac{14}{4}\right]+1} = x_{[3.5]+1} = x_{3+1} = x_4 = 3$$

Donji kvartil pokazuje da je prvih 25% knjižara prodalo manje od 3 knjige, a preostalih 75% više od 3 knjige.

Izračunajmo gornji kvartil:

$$Q_3 = x_{\left[\frac{3 \cdot 14}{4}\right]+1} = x_{[10.5]+1} = x_{11} = 8$$

Gornji kvartil pokazuje da je prvih 75% knjižara prodalo manje od 8 knjiga, a preostalih 25% više od 8 knjiga.

Sada možemo izračunati interkvartil:

$$I_Q = 8 - 3 = 5$$

koji pokazuje da je prodanost knjige u središnjih 50% knjižara između 8 i 3, odnosno u rasponu od 5 knjiga. Raspon varijacije je za taj niz $12-1=11$.

Vidi se velika razlika između interkvartila i raspona varijacije.

Primjer 1.1.2.

Struktura plaća u jednom poduzeću 2002.godine izražena u postocima dana je u tablici.

Izračunaj raspon varijacije i interkvartil.

Plaća u kunama	1800.5-2700.5	2700.5-3000.5	3000.5-3500.5	3500.5-4000.5	4000.5-4500.5	4500.5-5000.5	5000.5-7000.5	7000.5-i više
Postotak zaposlenika	4.2	5.5	14.3	20.3	28.5	9.6	15.1	2.5

Napravimo tablicu sa kumulativnim nizom i veličinom razreda.

Plaća	Struktura u %	Kumulativni niz	Veličina razreda
1800.5-2700.5	4.2	4.2	900
2700.5-3000.5	5.5	9.7	300
3000.5-3500.5	14.3	24.0	500
3500.5-4000.5	20.3	44.3	500
4000.5-4500.5	28.5	72.8	500
4500.5-5000.5	9.6	82.4	500
5000.5-7000.5	15.1	97.5	2000
7000.5-i više	2.5	100.0	(8000)
Ukupno	100.0	-	

Zbroj frekvencija je $N=100$; $\frac{N}{4} = 25$.

Kvartilni razred je 3500.5-4000.5.

Donja granica mu je $L_1 = 3500.5$, frekvencija $F_{k\text{ var}} = 20.3$.

Zbroj frekvencija do kvartilnog razreda 24.0, a veličina razreda 500.

Stoga je po formuli donji kvartil $Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_1}{f_{k\text{ var}}} \cdot i = 3500,5 + \frac{25 - 24}{20.3} \cdot 500 = 3525,13$

Na sličan način dobivamo gornji kvartil koji iznosi 5 600.50.

Interkvartil $I_Q = Q_3 - Q_1$ iznosi 2 075.37.

Donji kvartil pokazuje da je plaća prvih 25% zaposlenih (zaokruženo) 3525 kuna, i manja, a da je plaća ostalih 75% veća od navedenog iznosa.

Gornji je kvartil 5600 kuna, dakle, prvih 75% zaposlenih ima plaću 5600 kn i manje, a preostalih 25% ima plaću veću od spomenute.

Inerkvartil pokazuje da je raspon plaća središnjih 50% zaposlenih (zaokruženo) 2075 kuna.

Raspršenost je umjerena.

- c. **Varijanca** (σ^2) je aritmetička sredina kvadrata odstupanja vrijednosti numeričkog obilježja od njihove aritmetičke sredine.

Varijancu niza N pojedinačnih vrijednosti numeričke varijable $X : x_1, x_2, \dots, x_N$ sa sredinom $\bar{x}$ računamo po formuli:

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

a varijancu distribucije frekvencije (f_i, x_i) , $i = 1, 2, \dots, k$:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k f_i (x_i - \bar{x})^2 \text{ gdje je } N = \sum_{i=1}^k f_i .$$

Varijanca je najvažnija mjera raspršenosti u statistici, a zatim je iz nje izvedena standardna devijacija i koeficijent varijacije. Varijanca se određuje na temelju svih podataka. Ako su vrijednosti niza međusobno jednake onda nema varijabilnosti tj. aritmetička sredina jednaka je vrijednosti članova niza te je odstupanje od nje jednako nuli. Što je raspršenost veća, veće je odstupanje vrijednosti varijable od aritmetičke sredine.

Zbog toga se mjerenje raspršenosti može temeljiti na razlikama $x_i - \bar{x}$, $i = 1, \dots, N$.

Broj članova niza, pa onda i razlika, je velik te je teško donijeti zaključak o stupnju varijabilnosti niza. Zbog toga se utvrđuje broj kojim se to izražava, a to je upravo varijanca.

d. **Standardna devijacija (σ)** je pozitivan drugi korijen iz varijance.

Standardna devijacija numeričkog niza je:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2}{N}} \quad (1),$$

a standardna devijacija distribucije frekvencija (f_i, x_i) , $i = 1, 2, \dots, k$ je

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^k f_i (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^k f_i x_i^2 - \frac{1}{N} \left(\sum_{i=1}^k f_i x_i \right)^2}{N}}, N = \sum_{i=1}^k f_i \quad (2) .$$

Standardna devijacija nam daje prosječno odstupanje pojedinačnih vrijednosti numeričkog obilježja od aritmetičke sredine.

Ako je odstupanje maleno to ukazuje na malu raspršenost, odnosno disperziju članova numeričkog niza od aritmetičke sredine iz čega slijedi dobra reprezentativnost aritmetičke sredine. U obrnutom slučaju kad je disperzija velika, reprezentativnost aritmetičke sredine je slaba.

Umjesto frekvencija f_i u računanju varijance, odnosno standardne devijacije jednako se mogu primijeniti relativne frekvencije p_i ili postotne frekvencije P_i .

Za distribuciju frekvencija s razredima u izrazima (1) i (2) vrijednosti x_i predočuju razredne sredine. Ako se niz podataka shvaća kao slučajni uzorak, onda se u formulama umjesto s N dijeli s $N-1$.

Zbog toga varijanca sadrži pogrešku koja se obično zanemaruje.

Primjer 1.1.3.

Izračunajmo varijancu i standardnu devijaciju za podatke iz tablice o broju prodanih mobitela.

Broj prodanih mobitela u danu u danu u jednoj trgovini u 30 radnih dana.	Broj dana
X_i	f_i
0	3
1	5
2	6
3	5
4	4
5	4
6	3
Ukupno	30

Napravimo prvo tablicu:

Broj prodanih mobitela u danu	Broj dana	Podtotali	Odstupanje od prosjeka	Kvadrati odstupanja	
X_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
0	3	0	-2.87	8.24	24.71
1	5	5	-1.87	3.50	17.48
2	6	12	-0.87	0.76	4.54
3	5	15	0.13	0.02	0.08
4	4	16	1.13	1.28	5.11
5	4	20	2.13	4.54	18.15
6	3	18	3.13	9.80	29.39
Ukupno	30	86	-	-	99.47

Najprije izračunamo prosječni broj mobitela u danu pomoću formule za ponderiranu aritmetičku sredinu:

$$\bar{x} = \frac{f_1 x_1 + \dots + f_N x_N}{N}, \text{ gdje je } N = \sum_{i=1}^N f_i.$$

$$\bar{x} = \frac{\sum_{i=1}^N f_i x_i}{N} = \frac{86}{30} = 2,87$$

Varijanca je:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N f_i (x_i - \bar{x})^2 = \frac{1}{30} \cdot 99,47 = 3,32,$$

a standardna devijacija:

$$\sigma = \sqrt{3,32} = 1,822.$$

Prosječni broj prodanih mobitela je 2.87, s prosječnim odstupanjem od prosjeka 1.82 mobitela.

Varijanca i standardna devijacija izuzetno su bitne u gospodarskoj i poslovnoj statistici jer se temeljem njih procjenjuje rizik pri donošenju poslovnih odluka.

1.2. **Relativne mjere disperzije** izražene su u relativnim brojevima. To su:

- a. koeficijent varijacije
- b. koeficijent kvartilne devijacije

- a. **Koeficijent varijacije** (V) je omjer standardne devijacije i aritmetičke sredine pomnožen sa 100.

$$V = \frac{\sigma}{\bar{x}} \cdot 100, \quad \bar{x} > 0$$

Veći koeficijent varijacije pokazuje veću raspršenost, odnosno manju reprezentativnost aritmetičke sredine. Koeficijent varijacije može i prijeći vrijednost 100% u slučajevima kad se radi o veoma heterogenom nizu.

Uočimo, *standardna devijacija* je prosječno odstupanje vrijednosti numeričke varijable od njezine aritmetičke sredine, a *koeficijent varijacije* je to odstupanje izraženo u postotku.

- b. **Koeficijent kvartilne devijacije** (V_Q) je relativna mjera disperzije i zauzima vrijednosti od 0 do 1.

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

gdje je :

V_Q – koeficijent kvartilne devijacije

Q_1 – donji kvartil

Q_3 – gornji kvartil.

Što je V_Q bliže 0, disperzija je manja a medijan reprezentativniji, i obrnuto.

Primjer 1.2.1.

Izračunajmo koeficijent kvartilne devijacije za niz iz prethodnog primjera:

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{8 - 3}{8 + 3} = \frac{5}{11} = 0.454545 .$$

U relativnom iznosu je 0.454545 ili zaokruženo i u postotku, 45%. Raspršenost je dosta velika.

Primjer 1.2.2.

Zaključne cijene dionica neke tvrtke tijekom jednog tjedna bile su 575, 500, 350, 360, 450, 400 i 375 eura.

Izračunajmo varijancu, standardnu devijaciju i koeficijent varijacije.

Napravimo tablicu.

x_i	575	500	350	360	450	400	375	3010
$x_i - \bar{x}$	145	70	-80	-70	20	-30	-55	0
$(x_i - \bar{x})^2$	21025	4900	6400	4900	400	900	3025	41550
x_i^2	330625	250000	122500	129600	202500	160000	140625	1335850

Aritmetička sredina je 430, a $N=7$.

Varijanca je prema definicijskoj formuli 5935.71428, standardna devijacija 77.04359.

Koeficijent varijacije je 17.92 %.

Primjer 1.2.3.

Anketirani su gosti hotela koji su boravili u njemu u ljetnoj sezoni 2012.

Njihova potrošnja je dana u dolarima.

Potrošnja (u američkim dolar.)	Struktura anketiranih (u %)	Razredne sredine		
	P_i	x_i	$P_i x_i$	$P_i x_i^2$
200-240	3,7	220	814	179080
240-280	16,8	260	4368	1135680
280-360	49,3	320	15776	5048320
360-440	15,2	400	6080	2432000
440-640	10	540	5400	2916000
640-840	5	740	3700	2738000
Zbroj	100		36138	14449080

Podaci o potrošnji uređeni su kao distribucija s razredima. Frekvencije su postotne i s njima se postupa kao s apsolutnima.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^k f_i (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^k f_i x_i^2 - \frac{1}{N} \left(\sum_{i=1}^k f_i x_i \right)^2}{N}}, N = \sum_{i=1}^k f_i$$

$$= \sqrt{\frac{14449080 - \frac{1}{100} \cdot 36138^2}{100}} = \sqrt{13895,2956} = 117,87831.$$

Aritmetička sredina distribucije jednaka je 361.38 am.dolara.

Koeficijent varijacije je:

$$V = \frac{\sigma}{\bar{x}} \cdot 100, \quad V = \frac{117,87831}{361,38} 100 = 32,61\%$$

Prosječna potrošnja anketiranih turista iznosila je 361,38 am. dolara, a prosječno odstupanje je bilo 117,87 dolara, ili 32.61%.

Riječ je o umjerenoj disperziji potrošnje prema prosječnoj potrošnji.

Ako izračunamo i ostale podatke :

Medijan	Donji kvartil	Gornji kvartil	Interkvartil	Koef. kvart. devijacije
327,87	287,3	387,37	100,07	0,148

interkvartil upućuje na to da je raspon varijacije potrošnje središnjih 50% turista zaokruženo na 100 am. dolara, a koeficijent varijacije da je riječ o relativno maloj disperziji 14,8%.

2. Standardizirana vrijednost

Među važnim primjenama standardne devijacije je standardizirana vrijednost. Standardizirana vrijednost z je rezultat koji se odnosi na usporedbu relativnog položaja podataka u nizovima.

Definira se kao:

$$z_i = \frac{x_i - \bar{x}}{\sigma},$$

a svojstva su:

- aritmetička sredina je 0 ($\bar{z} = 0$)
- standardna devijacija je 1 ($\sigma_z = 1$)

Standardizirana vrijednost koristi se za usporedbu numeričkih nizova s različitim mjernim jedinicama jer ne ovisi o mjernim jedinicama varijabli.

Primjer 2.1.

Zadane su plaće osam zaposlenika u jednom poduzeću:

1500, 2700, 2900, 4000, 4200, 4900, 6100, 15000.

Izračunajmo standardizirane vrijednosti ovog obilježja i prokomentirajmo rezultate.

Plaća	Odstupanje od prosjeka	Kvadrati odstupanja	Standardizirana vrijednost
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	z_i
1500	-3662.5	13413906	-0.928
2700	-2462.5	6063906	-0.624
2900	-2262.5	5118906	-0.574
4000	-1162.5	1351406	-0.295
4200	-962.5	926406.3	-0.244
4900	-262.5	68906.25	-0.067
6100	937.5	878906.3	0.238
15000	9837.5	96776406	2.493
Ukupno	-	124598750	-

Prosječna plaća iznosi 5162.5 kn.

Varijanca je $\sigma^2 = \frac{1}{8} \cdot 124598750 = 15574843,75$.

Standardna devijacija $\sigma = 3946,49$ kn

Iz izračunatih standardiziranih vrijednost zaključujemo da sve plaće osim zadnje dvije odstupaju od prosjeka u negativnim standardnim devijacijama.

Zadnja plaća najviše odstupa od prosjeka i to za 2.493 standardnih devijacija.

Primjer 2.2.

Učenik je na testu iz matematike ostvario 65 bodova, a na testu iz kemije 30 bodova. Test iz matematike imao je aritmetičku sredinu 50 bodova, standardnu devijaciju 10, a test iz kemije aritmetičku sredinu od 25 bodova, standardnu devijaciju 5. Usporedite učenikove rezultate.

Mjere raspršenosti i mogućnosti njihova izbora s obzirom na vrstu statističkog niza

Mjera raspršenosti	Vrste niza		
	nominalni	redoslijedni	numerički
raspon varijacije	ne	da	da
interkvartil i koeficijent kvartilne devijacije	ne	da	da
varijanca, standardna devijacija i koeficijent varijacije	ne	ne	da

3. Pravilo normalne distribucije (Empirijsko pravilo)

Prosuđujemo li o intervalima u kojima se nalaze proporcije broja vrijednosti numeričkog niza koristimo više pravila koja se oslanjaju na aritmetičku sredinu i standardnu devijaciju.

Općenito vrijedi da se u intervalu od $(\bar{x} - 2\sigma)$ do $(\bar{x} + 2\sigma)$ nalazi najmanje 75 % članova bilo kako raspoređenog numeričkog niza.

U intervalu od $(\bar{x} - 3\sigma)$ do $(\bar{x} + 3\sigma)$ najmanje 90 % članova,

a u intervalu od $(\bar{x} - 4\sigma)$ do $(\bar{x} + 4\sigma)$ najmanje 94 % njegovih članova.

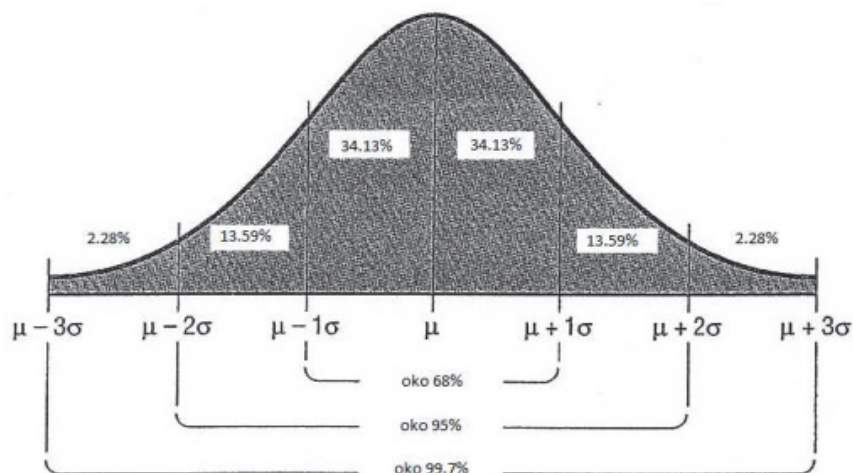
Raspored podataka ponekad se približno ravna prema **normalnoj distribuciji** (distribucija numeričkih podataka zvonolikog oblika). Ako se podaci raspoređuju po normalnoj distribuciji, tada je:

u intervalu $(\bar{x} - \sigma)$ do $(\bar{x} + \sigma)$ obuhvaćeno 68.26 % podataka,

u intervalu $(\bar{x} - 2\sigma)$ do $(\bar{x} + 2\sigma)$ 95.44% podataka,

u intervalu od $(\bar{x} - 3\sigma)$ do $(\bar{x} + 3\sigma)$ 99.73 % podataka.

Grafički prikaz normalne raspodjele je krivulja koju nazivamo *Gaussova krivulja* ili *zvonolika krivulja*.



Iz normalne distribucije slijedi da vrlo malo podataka odstupa od aritmetičke sredine više od tri standardne devijacije.

Primjer 3.1.

Koristeći se podacima o cijeni jagoda tijekom deset radnih dana istražimo ima li netipičnih iznosa cijena među njima. Cijene su bile:

18 kn, 24 kn, 25 kn, 25 kn, 26 kn, 27 kn, 28 kn, 29 kn, 30 kn i 36 kn.

Prosječna cijena jagoda iznosi 26.8 kn.

Cijena jagoda	Odstupanje od prosjeka	Kvadrati odstupanja
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
18	-8.8	77.44
24	-2.8	7.84
25	-1.8	3.24
25	-1.8	3.24
26	-0.8	0.64
27	0.2	0.04
28	1.2	1.44
29	2.2	4.84
30	3.2	10.24
36	9.2	84.64
Ukupno	-	193.6

Varijanca je $\sigma^2 = \frac{1}{10} \cdot 193.6 = 19.36$, a standardna devijacija $\sigma = 4.4$

Cijene koje bi mogle biti atipične su 18 kn i 36 kn. Provjerimo po pravilu normalne razdiobe.

$$\langle \bar{x} - \sigma, \bar{x} + \sigma \rangle = \langle 26.8 - 4.4, 26.8 + 4.4 \rangle = \langle 22.4, 31.2 \rangle$$

$$\langle \bar{x} - 2\sigma, \bar{x} + 2\sigma \rangle = \langle 28.8 - 8.8, 26.8 + 8.8 \rangle = \langle 18, 35.6 \rangle$$

$$\langle \bar{x} - 3\sigma, \bar{x} + 3\sigma \rangle = \langle 26.8 - 13.2, 26.8 + 13.2 \rangle = \langle 13.6, 40 \rangle$$

Iz navedenog vidimo da se prva vrijednost nalazi u prvom intervalu, a druga vrijednost u trećem intervalu. Zaključujemo da nema netipičnih vrijednosti cijena jer niti jedna vrijednost ne odstupa više od tri standardne devijacije od prosječne cijene.

Primjer 3.2.

Dobavljač kupcu isporučuje paketiće kave prosječne neto težine od 100 grama. Dopušteno je prosječno odstupanje od prosječne težine od 2 grama. Pretpostavimo li da je distribucija paketića kave oblika normalne distribucije sa sredinom 100 i standardnom devijacijom 2, odredimo broj paketića težine od 96 grama do 104 grama u pošiljci od 25 000 isporučenih paketića.

Kako je standardna devijacija 2 grama, iz toga slijedi da paketići od 96 grama i 104 grama odstupaju dvije standardne devijacije od prosječne težine. Iz toga proizlazi da interval od 96 grama do 104 grama odgovara intervalu $\langle \bar{x} - 2\sigma, \bar{x} + 2\sigma \rangle$ normalne distribucije. U tom je intervalu zaokruženo 95 % podataka, što u našem slučaju znači da u toj pošiljci možemo očekivati 23 750 paketića kave teških od 96 grama do 104 grama.

Zadaci 4.

1. Odredite kvartile u statističkom nizu koji se odnosi na predviđene stope zarade po dioničkim udjelima od 15 financijskih analitičara:

16, 6, 4, 8, 12, 11, 4, 3, 5, 7, 12, 14, 13, 14, 17.

2. Podaci 5, 6, 7, 8, 9, 9, 9, 10, 12, 14, 17, 17, 18, 19, 19 predstavljaju broj sati koje učenici jedne srednje škole provedu na internetu tijekom jednog tjedna. Odredite kvartile, interkvartil, raspon varijacije i relativnu mjeru disperzije.

3. Cijene deset udžbenika za četvrti razred gimnazije su:

84.60, 80.10, 101.70, 124.00, 106.20, 97.20, 97.20, 89.10, 88.00, 76.50.

Izračunajte standardnu devijaciju.

4. Izračunajte vrijednost varijance, standardne devijacije i koeficijenta varijacije za niz vrijednosti: 8, 5, 3, 9, 4, 0, 3, 2.
5. Na upitnicima za prijam u bolnicu jedan od traženih podataka je koliko cigareta pacijent popuši dnevno, ukoliko spada u kategoriju pušača. Analizom upitnika dobivene su ove vrijednosti:

20, 30, 15, 20, 25, 20, 25, 25, 40, 35, 10, 35, 10, 5, 10, 25, 30, 25, 5, 20, 25, 35.

Razvrstajte ovaj negrupirani niz i odredite kvartile, interkvartil, raspon, relativnu mjeru disperzije, aritmetičku sredinu, standardnu devijaciju i ispitaj ima li među pacijentima nekoga tko dnevno popuši atipičan broj cigareta.

6. Distribucija električnih žarulja određenog tipa prema trajanju u satima približno je simetrična, ima normalnu razdiobu, s aritmetičkom sredinom od 1 000 sati i standardnom devijacijom od 150 sati. Proizvodna partija sadrži 100 000 žarulja.
 - a) Za koliko se žarulja očekuje da će trajati od 850 do 1150 sati?
 - b) Za koliko se žarulja očekuje da će trajati od 700 do 1300 sati ?
 - c) Za koliko se žarulja očekuje da će trajati kraće od 550 sati ili dulje od 1 450 sati?

Rješenja:

1. $Q_1=5$, $Q_2=11$, $Q_3=14$
2. $Q_1=8$, $Q_2=10$, $Q_3=17$, $I_Q=9$, $R_x= 14$, $V_q=0.36$
3. $\sigma=13.32$
4. $\sigma=2.81736$, $V= 66.29075$
5. $Q_1=17.5$, $Q_2=25$, $Q_3=27.5$, $I_Q=10$, $R_x=35$, $V_q=0.22$, $\bar{x}=20.75$
 $\sigma=9.3091$, nema
6. a) 68 000 b) 95 000 c) 150

Sažetak

- mjerama disperzije brojčano se opisuje stupanj varijabilnosti statističkih podataka
- raspon varijacije – najjednostavnija približna mjera disperzije; izražen je u mjernim jedinicama obilježja; $R_x = x_{\max} - x_{\min}$
- kvantili – vrijednosti numeričke varijable koji uređen numerički ili redosljedni niz dijele na jednakobrojne dijelove; dijele li kvantili na četiri jednakobrojna dijela riječ je o kvartili, na 10 dijelova decili, na 100 dijelova percentili
- medijan se ubraja u kvantile
- najvažnija mjera disperzije je varijanca te iz nje izvedena standardna devijacija i koeficijent varijacije
- interkvartil –raspon varijacije središnjih 50 % članova niza uređenih parova

$$I_Q = Q_3 - Q_1$$

- varijanca – aritmetička sredina kvadrata odstupanja vrijednosti numeričke varijable od njezine aritmetičke sredine: $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$
- standardna devijacija – pozitivni drugi korijen iz varijance

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2}{N}}$$

- varijanca i standardna devijacija distribucije frekvencija:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^k f_i (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^k f_i x_i^2 - \frac{1}{N} \left(\sum_{i=1}^k f_i x_i \right)^2}{N}}, N = \sum_{i=1}^k f_i$$

- koeficijent devijacije – omjer standardne devijacije i aritmetičke sredine pomnožen sa 100

$$V = \frac{\sigma}{\bar{x}} \cdot 100,$$

- standardizirana varijabla z – linearna transformacija numeričke varijable x ; provodi se tako da se odstupanja vrijednosti numeričke varijable podijele sa standardnom devijacijom

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

V UTJECAJ DODAVANJA ILI UKLANJANJA PODATAKA NA SREDNJE VRIJEDNOSTI NIZA

1. Netipične vrijednosti

Izračun srednjih vrijednosti niza svodi se na jedan broj koji karakterizira cijeli skup podataka. Koliko je reprezentativan taj broj ovisi o više čimbenika. Jedan od njih su *netipične vrijednosti* koje mogu značajno utjecati na vrijednost neku od razmatranih mjera. To su izuzetno male ili velike vrijednosti koje se značajno razlikuju od većine vrijednosti niza.

Razlozi postojanja netipičnih vrijednosti: krivo mjerenje ili unos podataka u računalo, radi se o rijetkom podatku itd.

One utječu na izračun aritmetičke sredine, dok manji utjecaj imaju na izračun G i H sredine.

Na položajne srednje vrijednosti (med i mod) nemaju utjecaja.

Dobar način prepoznavanja netipičnih vrijednosti je promatranje grafičkih prikaza podataka.

Primjer 1.1.

Analiziramo prosječnu plaću za skupinu deset nasumično odabranih osoba, čije su plaće dane u kunama:

4.400 3.700 6.000 4.280 5.750 5.300 4.400 3.900 5.000 20.000

- aritmetičkom sredinom za prvih 9 plaća
- aritmetičkom sredinom za svih 10 plaća
- geometrijskom sredinom za prvih 9 plaća
- geometrijskom sredinom za svih 10 plaća
- harmonijskom sredinom za prvih 9 plaća
- harmonijskom sredinom za svih 10 plaća
- izračunajmo mod za prvih 9 plaća
- izračunajmo mod za svih 10 plaća
- izračunajmo medijan za prvih 9 plaća
- izračunajmo medijan za svih 10 plaća

Rješenje:

$$\text{a. } \bar{x} = \frac{4.400 + 3.700 + \dots + 5.000}{9} = 4.753,33 \text{ kune}$$

$$\text{b. } \bar{x} = \frac{4.400 + \dots + 20.000}{10} = 6.278 \text{ kn}$$

Razlika je 1.525 kn.

$$\text{c. i d. } \bar{x} = \sqrt[9]{4.400 \cdot \dots \cdot 5.000} = 4.692,64 \text{ kn}$$

$$\bar{x} = \sqrt[10]{4.400 \cdot \dots \cdot 20.000} = 5.424,74 \text{ kn}$$

Razlika je 732.10 kn.

$$\text{e. i f. } \bar{x} = \frac{9}{\frac{1}{4.400} + \dots + \frac{1}{5.000}} = 4.633,52 \text{ kn}$$

$$\bar{x} = \frac{10}{\frac{1}{4.400} + \dots + \frac{1}{20.000}} = 5.019,16 \text{ kn}$$

Razlika je 385.64 kn.

Zaključujemo da je aritmetička sredina najviše pod utjecajem netipičnih vrijednosti.

g. najčešće se pojavljuje plaća u iznosu 4.400 kn

h. najčešće se pojavljuje plaća u iznosu 4.400 kn

Nema promjene za mod tj. najčešća plaća u nizu nije bila pod utjecajem dodavanja netipične vrijednosti. Međutim, dodavanje određenih vrijednosti može utjecati da niz sa jednim modom postane multimodalan npr. dodamo iznos plaće od 5.000 kn.

3.700 3.900 4.280 4.400 4.400 5.000 5.000 5.300 5.750 6.000 20.000

Sada postoje dva moda: 4.400 i 5.000 kn, što znači da su najčešće plaće u promatranom uzorku osoba 4.400 i 5.000 kn.

i. Poredajmo plaće po veličini :

3.700 3.900 4.280 4.400 4.400 5.000 5.300 5.750 6.000 20.000

Medijalna točka je:

$$r = \frac{1}{2}(N+1) = \frac{1}{2}(9+1) = 5$$

što znači da peti po redu iznos plaće je medijan, u iznosu 4.400kn. To znači da je 50% od 9 osoba imalo plaću u iznosu 4.400kn ili manje, a preostala polovica 4.400 kn ili više.

$$\begin{aligned} h. \quad r_1 &= \frac{1}{2}N = 5 \\ r_2 &= \frac{1}{2}N + 1 = 6 \end{aligned} \quad \text{5. i 6. plaća po redu su središnje}$$

$$M_e = \frac{4.400 + 5.000}{2} = 4.700, \text{ što znači da u uzorku od deset osoba njih 50\% je imalo}$$

plaću u iznosu 4.700 kn ili manje, a preostalih 50% 4.700 kn ili više. Medijalna plaća porasla je za 300kn što je malo u odnosu na izračun aritmetičke sredine.

2. Ublažavanje problema utjecaja dodavanja ili uklanjanja podataka na srednje vrijednosti

Kako postupiti u slučajevima kad aritmetička sredina postane manje reprezentativna?

- i. Jedan od načina je **korištenje medijana** umjesto aritmetičke sredine.
- ii. Drugi pristup se sastoji u izračunu **modificirane aritmetičke sredine**.

Najprije poredamo statistički niz po veličini od najmanje do najveće vrijednosti. Ako se uzorak sastoji od n opažanja nekog obilježja, uz k netipičnih vrijednosti u nizu, tada možemo računati aritmetičku sredinu tako da ispustimo $\frac{k}{2}$ ekstremnih vrijednosti na svakom kraju niza i izračunamo aritmetičku sredinu za preostalih $n - k$ vrijednosti.

Možemo govoriti i o postocima, tako da računamo $x\%$ modificiranu aritmetičku sredinu, pri čemu je x postotak koliko smo netipičnih vrijednosti ispustili iz niza.

Primjer 2.1.

Izračunajmo modificiranu aritmetičku sredinu za primjer prosječnih plaća za skupinu deset nasumično odabranih osoba, čije su plaće dane u kunama:

3.700 3.900 4.280 4.400 4.400 5.000 5.300 5.750 6.000 20.000

tako da ispuštimo 10% najvećih i 10% najmanjih vrijednosti (od 10 podataka ćemo oduzeti dvije vrijednosti, jednu najmanju i jednu najveću). Novi niz je tada:

3.900 4.280 4.400 4.400 5.000 5.300 5.750 6.000

Modificirana aritmetička sredina tada iznosi

$$\bar{x} = \frac{3.900 + \dots + 6.000}{8} = 4.879kn$$

i puno je bliža originalnoj aritmetičkoj sredini za 9 osoba u početnom primjeru, koja je iznosila 4.753,33 kn, u odnosu na aritmetičku sredinu računanu za 10 osoba.

iii. Treći pristup koji možemo primijeniti **je izračun geometrijske sredine.**

Prilikom pisanja izvješća dobro je uključiti sve spomenute mogućnosti ublažavanja problema izračuna aritmetičke sredine i ne oslanjati se isključivo na jednu.

3. Uspoređivanje srodnih skupova podataka

Često je prilikom donošenja poslovnih ili gospodarskih odluka potrebno međusobno usporediti karakteristike različitih ishoda kako bi se temeljem njih mogle donijeti kvalitetne i ispravne odluke.

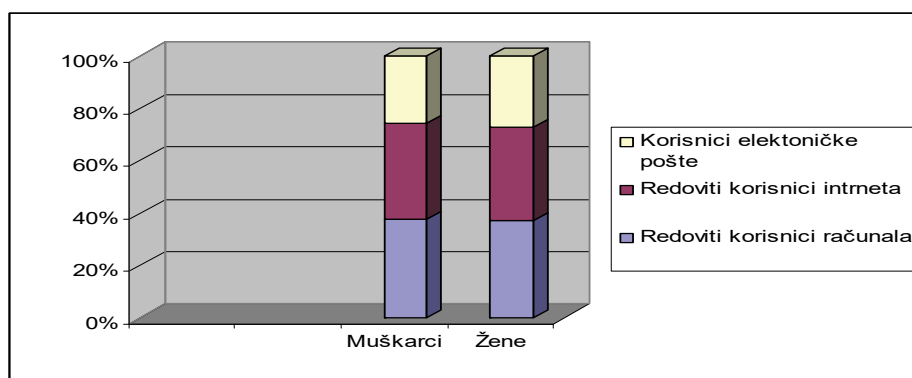
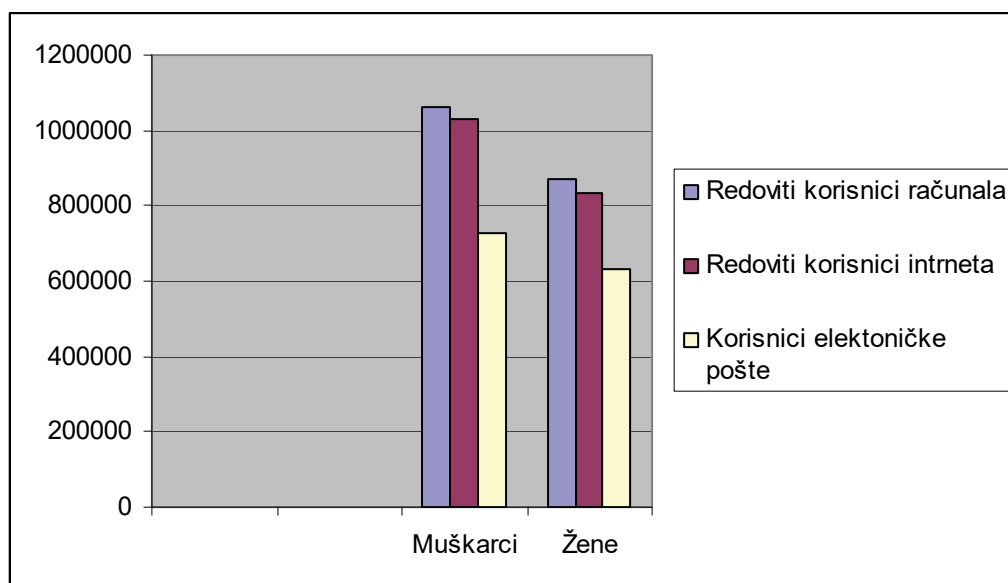
3.1. Grafičko uspoređivanje srodnih skupova podataka

Promotrimo uspoređivanje dva ili više nizova podataka na primjeru dijagrama višestrukih stupaca, razdijeljenih stupaca i strukturnih krugova.

Primjer 3.1.1.

Na temelju podataka iz tablice o korisnicima interneta i računala u Hrvatskoj izrađeni su dijagrami:

Korisnici	Muškarci	Žene
Redoviti korisnici računala	1060218	869785
Redoviti korisnici interneta	1028250	834722
Korisnici elektroničke pošte	728138	630568

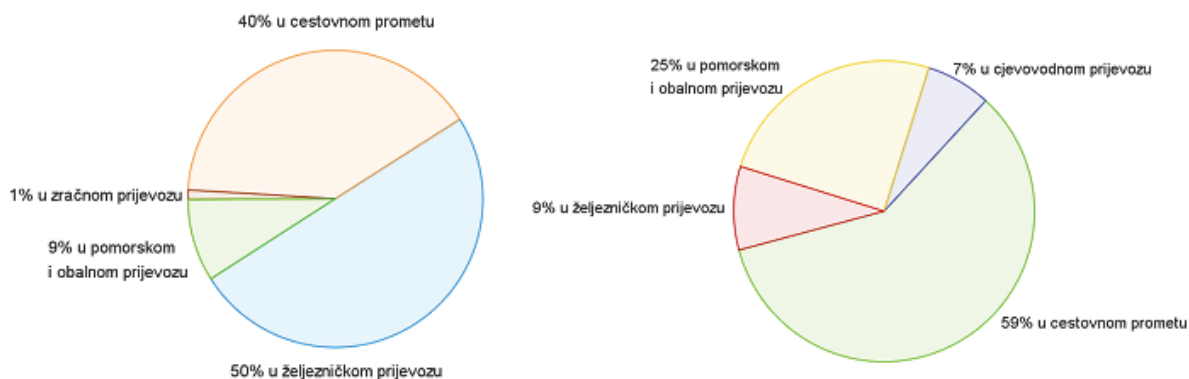


Promatranjem slika možemo zaključiti kako je veći broj korisnika računala, interneta i elektroničke pošte muškog spola. To se može objasniti, primjerice, većom stopom zaposlenosti muškaraca koji na radnom mjestu koriste računala.

Usporedimo srodne skupove podataka na primjeru strukturnih krugova.

Primjer 3.1.2.

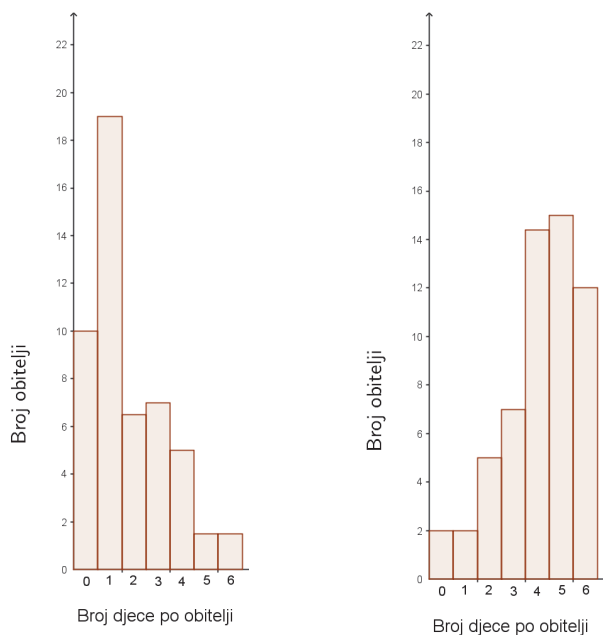
Na slici prikazana su dva strukturna kruga koji se odnose na strukturu prijevoza putnika (gore) i robe (dolje) u 2010. godini



Dok je najveći dio putnika koristio željeznički prijevoz (50% putnika), samo je 9% ukupne robe prevezeno putem istog tipa prijevoza. Možemo zaključiti kako u prijevozu putnika i robe postoje bitne strukturne razlike, koje mogu proizlaziti iz cijene pojedinog oblika prijevoza, vremena trajanja i drugih čimbenika.

Primjer 3.1.3.

Analizirajmo distribuciju frekvencija na temelju dva histograma distribucije broja djece po obitelji u dva naselja, slika A (lijevo) i slika B (desno).



Uspoređujući broj djece po obitelji u oba naselja, može se uočiti kako u naselju A ima najveći broj obitelji s jednim djetetom po obitelji, pri čemu je vrlo mali broj obitelji imao veći broj djece (5,6). S druge strane, naselje B ima poprilično drugačiju situaciju. Razlozi mogu biti

različiti, u naselju B možda ima više vrtića i škola koje su privukle obitelji sa većim brojem djece, itd. Potrebno je ispitati uzroke te pokušati riješiti ovaj problem kako bi se smanjile razlike među naseljima.

Zaključno, promatranje grafičkih prikaza srodnih skupova podataka omogućava brzo donošenje osnovnih zaključaka o pojavama koje se analiziraju. Istraživač na taj način dobije prvi uvid u rasporede podataka ali to nije dovoljno za izvođenje formalnih zaključaka.

3.2. Uspoređivanje srednjih vrijednosti i mjera raspršenosti srodnih skupova podataka

Uspoređivanje srednjih vrijednosti i mjera raspršenosti srodnih skupova podataka olakšava donošenje odluke o problemima koji se razmatraju. Ponovimo:

Mjera raspršenosti	Vrsta niza		
	nominalni	redoslijedni	numerički
raspon varijacije	ne	da	da
inerkvartil i koeficijent kvartilne devijacije	ne	da	da
varijanca, standardna devijacija, koeficijent varijacije	ne	ne	da

Napomena: koeficijent varijacije ne računa se za nizove nastale mjerenjem koje polazi od dogovorno određene nule (mjerenje temperature).

Primjer 3.2.1.

Investitor razmatra ulaganje u tri dionice: X, Y i Z. Temeljem podataka s burze za 2012. godinu izračunao je nekoliko osnovnih pokazatelja o zaključnim cijenama dionica. Rezultati su dani u tablici

	Dionica X	Dionica Y	Dionica Z
Aritmetička sredina	300 kn	500 kn	20 kn
Mod	280 kn	440 kn	18 kn
Medijan	280 kn	470 kn	21 kn
Varijanca	2500	400	1
Standardna devijacija	50 kn	30 kn	1 kn
Koeficijent varijacije	16,67%	6%	5%
Interkvartilni raspon	44 kn	20 kn	3 kn
Koeficijent kvartilne devijacije	20%	16%	8%

Ako razmatramo prosječnu cijenu dionice, uočavamo kako postoje velike razlike promatrajući samo aritmetičku sredinu. Kako bismo komentirali reprezentativnost aritmetičke sredine, potrebno je analizirati i standardne devijacije te koeficijente devijacije.

Prosječna cijena dionice X je 300 kn, uz prosječno odstupanje od prosječne cijene 50 kn (standardna devijacija), odnosno relativno 16,67% (koef.var.). Analogno tumačimo i za dionicu Y i Z.

Možemo zaključiti kako je raspršenost cijena oko prosječne cijene najveća za dionicu X s obzirom na najveći koeficijent varijacije, dok je najmanja raspršenost za dionicu Z.

To govori da je *sigurnije uložiti u dionicu Z s obzirom da se cijene kreću oko izračunate aritmetičke sredine.*

U prethodnom poglavlju spominjane su netipične vrijednosti koje utječu na izračun aritmetičke sredine, stoga je dobro izračunati interkvartil i temeljem njega koeficijent kvartilne devijacije. Na njih netipične vrijednosti neće imati utjecaja. Usporedbom koeficijenta kvartilne devijacije može se uočiti kako je raspršenost središnjih 50% vrijednosti cijena dionica najmanja za dionicu Z.

Zaključak: Iako bi investitora mogla privući najveća cijena dionice na burzi, mjere raspršenosti nam pokazuju da je najbolje odabrati dionicu Z koja ima najmanji koeficijent varijacije a i koeficijent kvart. devijacije.

Primjer 3.2.2.

Poduzetnik je odlučio analizirati strukturu svojih kupaca-dužnika kako bi mogao donijeti odluku o daljnjoj poslovnoj suradnji s njima. Promatra prosječno vrijeme naplate dugovanja

	Dužnik 1	Dužnik 2	Dužnik 3
Aritmetička sredina	90	30	55
Standardna devijacija	8	30	16
Koeficijent varijacije	8,89%	100%	29%

Poduzetnik je izračunao prosječno vrijeme naplate te je ono za dužnika 1 iznosilo 90 dana, 30 dana za dužnika 2 i 55 dana za dužnika 3. Čini se po tome da je najbolje poslovati sa dužnikom 2.

Međutim, izračunato je prosječno odstupanje od prosječne naplate dugovanja i ono iznosi redom 8, 30 i 16 dana. Relativno to iznosi 8,89%, 100% i 29%. Sada ispada da je prosječno vrijeme naplate za dužnika 1 najbolje jer je relativno odstupanje od prosjeka najmanje.

Dužnik 2 ima značajno odstupanje od prosjeka, što znači da je znao podmiriti dug u vrlo kratkom razdoblju, a drugi put u jako dugom.

Primjer 3.2.3.

U kontroli kvalitete trajnosti žarulje prikupljeni su uzorci dva tipa žarulja. Želi se ispitati prosječna trajnost žarulja u satima.

	Žarulja tipa A	Žarulja tipa B
Aritmetička sredina	1502	1799
Mod	1500	1489
Koeficijent varijacije	3%	7%

Prosječno trajanje žarulje A je 1502 sata a žarulje B 1799 sati. Čini se da je žarulja B kvalitetnija. Međutim, usporedbom modova može se zaključiti da je najčešće vrijeme trajanja žarulja tipa A bilo 1500 sati, dok za B je 1489 sati.

Veća razlika između aritmetičkih sredina i modova kod žarulje tipa B upućuje kako B imaju najčešću trajnost 1489 sati jer je u uzorak odabrana jedna ili nekolicina žarulja koje su imale netipično dugo vrijeme trajanja u odnosu na ostale i time utjecale na samo prosječno vrijeme trajanja.

Usporedba koeficijenta varijacije pokazuje da je prosječno vrijeme trajanja žarulje A bolji pokazatelj.

Za proizvodni pogon to bi značilo da je potrebno izvršiti određene modifikacije u proizvodnji žarulje B (popravak strojeva ili zamjena materijala od kojeg se proizvode i slično).

Zadaci 5

1. Što su netipične vrijednosti?
2. Navedite razloge postojanja netipičnih vrijednosti u nizu.
3. Kako možemo brzo uočiti netipične vrijednosti u nizu ?

4. Kako dodavanje netipičnih vrijednosti utječe na aritmetičku a kako na geometrijsku i harmonijsku sredinu?
5. Utječe li dodavanje netipičnih vrijednosti na mod i medijan statističkog niza i zašto?
6. Kako unimodalni statistički niz može postati bimodalni ili multimodalni?
7. Nabrojite kako možemo ublažiti pojam netipičnih vrijednosti u nizu?
8. Što je modificirana aritmetička sredina?
9. Analiziramo prosječno vrijeme čekanja u redu u jednoj banci, te su dani podaci o čekanju pojedinih stranaka u minutama: 4,8,10,11,12,20,22,25,34,37,42,55
Izračunajte modificiranu aritmetičku sredinu tako da iz niza ispustite najmanju i najveću vrijednost.
10. Kako bi donio opravdanu odluku o ulaganju u dionice, investitor je razmotrio prosječne cijene triju dionica: 1,2,3. Temeljem podataka o kretanju cijena dionica u ožujku 2013. godine izračunao je nekoliko osnovnih pokazatelja o zaključnim cijenama dionica.

	Dionica 1	Dionica 2	Dionica 3
Aritmetička sredina	25 kn	29 kn	31 kn
Medijan	22 kn	29 kn	30 kn
Standardna devijacija	5 kn	2 kn	15 kn
Koeficijent varijacije	20%	6,90%	48,39%

Interpretirajte sve pokazatelje za sve tri dionice, prokomentirajte ih i donesite odluku u koju dionicu od navedenih tri ulagati i zašto.

11. Proizvođač mlijeka i sireva izbacuje novu vrstu namaza od sira koju planira najprije plasirati u tri veća trgovačka lanca. pri tom želi razmotriti koliko je prosječno vrijeme naplate plasiranog proizvoda koje nude analizirani trgovački lanci. Rezultati su u danima.

	Trgovački lanac		
	A	B	C
Aritmetička sredina	40	70	65
Standardna devijacija	36	12	7
Koeficijent varijacije	80%	17,14%	10,77%

Interpretirajte sve pokazatelje i komentirajte ih.

U koji lanac biste najprije plasirali proizvod i zašto?

12. Poduzeće proizvodi kožne jakne u 4 pogona. Uprava je odlučila analizirati efikasnost svakog pogona kako bi utvrdili postoje li kakva veća odstupanja od proizvodne norme te otkloniti ako postoje kakvi problemi. Prosječna norma je oko 100 jakni na dan.

	Pogon 1	Pogon 2	Pogon 3	Pogon 4
Prosječna dnevna proizvodnja	98	102	95	93
Standardna devijacija	7	8	6	15
Koeficijent varijacije	7%	7,84%	6,32%	16,13%

Interpretirajte sve pokazatelje i komentirajte ih. U kojem pogonu je potrebno podrobnije analizirati proces proizvodnje i otkloniti uzroke problema?

Rješenja 5

$$9. \quad \bar{x} = \frac{8 + 10 + \dots + 42}{10} = 22,1 \text{ min}$$

10. Investitor će najprije uložiti novac u dionicu 2 s obzirom da je prosječna cijena te dionice najreprezentativnija s obzirom na najmanji koeficijent varijacije.
11. U trgovački lanac C s obzirom da je njegov koeficijent varijacije najmanji. Iako je najkraće vrijeme naplate za lanac A najmanje, pripadajući koef.varij.upućuje da postoji velika varijabilnost u vremenu naplate za lanac A
12. Potrebno je analizirati pogon 4, s obzirom da je prosječna proizvodnja jakni u tom pogonu najmanja od svih, kao i što je varijabilnost dnevne proizvodnje u tom pogonu najveća (16,13%).

VI PRAVAC REGRESIJE

Između dvaju skupova statističkih podataka može se pojaviti određena veza: linearna, eksponencijalna, logaritamska, logistička, potencijnska.... Ako je ta veza linearna, pronalažanjem jednadžbe takozvanog regresijskog pravca, možemo procjenjivati i prognozirati koje će vrijednosti poprimiti varijabla za one podatke koji nisu u početnom skupu.

Primjer 1.1.

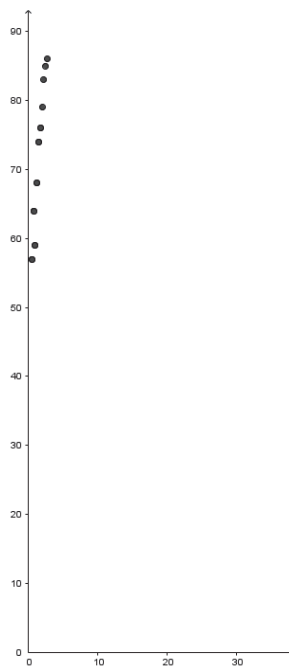
Svakome od šest pacijenata zabilježena je starost i visina sistoličkog tlaka izmjerena pri posjetu ordinaciji. Podaci su dani tablično:

pacijent	starost (x_i)	sistolički tlak (y_i)
1	36	120
2	40	123
3	43	127
4	45	128
5	46	130
6	51	135

Grafički prikazimo ove podatke u koordinatnom sustavu u kojem na x -os nanese starost, a na y -os tlak. Procijenite koliki će tlak imati 48-godišnji pacijent.

Točke $(36,120)$, $(40,123)$,..., $(51,135)$ unesemo u koordinatni sustav.

Ova vrsta grafa naziva se dijagram raspipanja (engl. scatter diagram).



Iz grafa vidimo da između starosti i sistoličkog tlaka postoji veza i to skoro linearna. Kako pronaći jednadžbu pravca koji se najbolje prilagođava nacrtanim točkama? Drugim riječima, želimo pronaći pravac koji bi najbolje reprezentirao ovisnost tih dviju varijabli.

Takav pravac naziva se **regresijski pravac**. Njegova jednadžba ima oblik: $\hat{y} = ax + b$. Taj pravac ima svojstvo da je među svim mogućim pravcima zbroj kvadrata odstupanja varijable y od $\hat{y}$ najmanji. Konkretno, brojeve a i b računamo pomoću formula:

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum y_i}{n} - a \cdot \frac{\sum x_i}{n}$$

pri čemu oznake $\sum x_i$, $\sum y_i$, $\sum x_i y_i$, $\sum x_i^2$ imaju ovo značenje:

$$\sum x_i = x_1 + x_2 + \dots + x_n$$

$$\sum y_i = y_1 + y_2 + \dots + y_n$$

$$\sum x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

$$\sum x_i^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

Dopunimo našu tablicu s dva stupca u kojima su brojevi x_i^2 i $x_i y_i$, a u zadnjem retku napišimo zbrojeve po stupcima.

pacijent	starost (x_i)	tlak (y_i)	x_i^2	$x_i y_i$
1	36	120	1296	4320
2	40	123	1600	4920
3	43	127	1849	5461
4	45	128	2025	5760
5	46	130	2116	5980
6	51	135	2601	6885
zbroj Σ	261	763	11 487	33 326

Uvrstimo li dobivene brojeve u formulu za a i b dobivamo:

$$a = \frac{n \sum x_i y_i - (\sum x_i) (\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} = \frac{6 \cdot 33\,326 - 261 \cdot 763}{6 \cdot 11\,487 - 261^2} = 1.015$$

$$b = \frac{\sum y_i}{n} - a \cdot \frac{\sum x_i}{n} = \frac{763}{6} - 1.015 \cdot \frac{261}{6} = 83.014$$

Jednadžba regresijskog pravca glasi: $\hat{y} = 1.015x + 83.014$.

Ucrtajmo pravac u dijagram rasipanja.

Sad možemo procijeniti i koliki bi bio tlak osobe stare 48 godina.

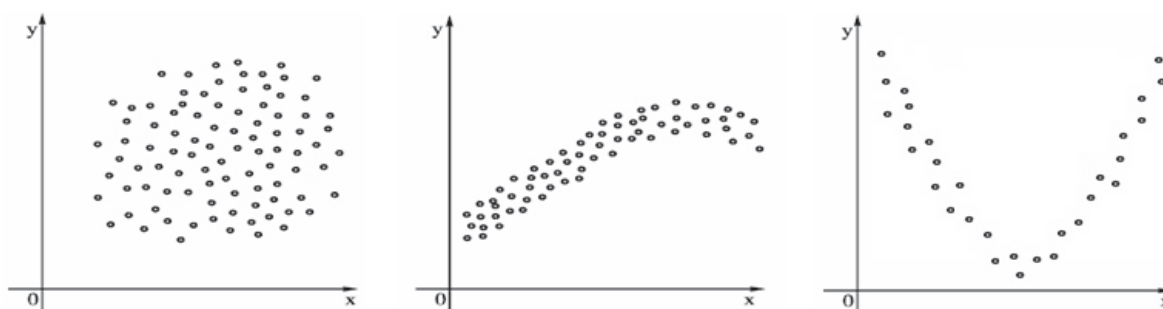
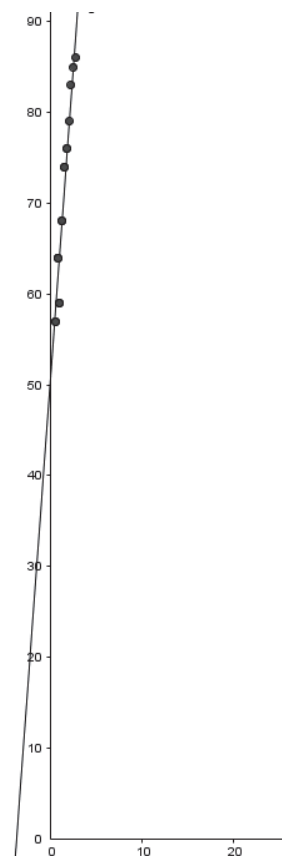
Uvrstimo li $x = 48$ u jednadžbu regresijskog pravca, dobivamo:

$$\hat{y} = 1.015 \cdot 48 + 83.014 = 131.7 \approx 132.$$

Dakle, procijenjujemo da bi 48-godišnjak imao sistolički tlak **132** mmHg.

Ako nam diagram rasipanja sugerira da veza nije linearna, onda opisani račun nema smisla. Evo nekoliko takvih diagrama rasipanja.

Na prvom grafu nema neke očite veze između podataka, dok na druga dva grafa vidimo da veza postoji ali nije linearna.



Ako je veza linearna, pronalaženjem formule koja opisuje tu vezu, tj. pronalažanjem jednadžbe regresijskog pravca, možemo procjenjivati koje će vrijednosti poprimiti varijabla y za one podatke x koji nisu u početnom skupu podataka.

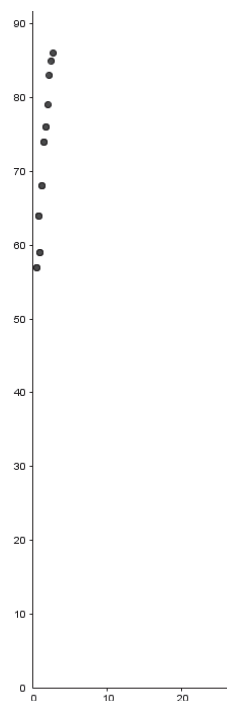
Primjer 1.2.

Deset učenika prvog razreda bilo je upitano koliko su se sati pripremali za ispit iz matematike. Njihovi odgovori na to pitanje uspoređeni su s bodovima što su ih dobili na ispitu. Maksimalni broj bodova na testu iznosio je 100.

učenik	vrijeme u satima (x_i)	broj ostvarenih bodova na testu (y_i)
1	0.50	57
2	0.75	64
3	1	59
4	1.25	68
5	1.50	74
6	1.75	76
7	2	79
8	2.25	83
9	2.5	85
10	2.75	86

- nacrtajte dijagram rasipanja
- napišite jednadžbu regresijskog pravca
- procijenite, ako se neki učenik pripremao 0.25 h koji je njegov najvjerojatniji rezultat na ispitu
- koliko sati se učenik trebao pripremati da bi ostvario maksimalni broj bodova na testu?

Iz dijagrama rasipanja vidimo da je veza linearna.



Dopunimo našu tablicu:

učenik	vrijeme u satima (x_i)	broj ostvarenih bodova na testu (y_i)	x_i^2	$x_i y_i$
1	0.50	57	0.25	28.5
2	0.75	64	0.5625	48
3	1	59	1	59
4	1.25	68	1.5625	85
5	1.50	74	2.25	111
6	1.75	76	3.0625	133
7	2	79	4	158
8	2.25	83	5.0625	186.75
9	2.5	85	6.25	212.5
10	2.75	86	7.5625	236.5
Zbroj Σ	16.25	731	31.5625	1258.25

Izračunajmo koeficijente a i b .

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} = \frac{10 \cdot 1258.25 - 16.25 \cdot 731}{10 \cdot 31.5625 - 16.25^2} = 13.648$$

$$b = \frac{\sum y_i}{n} - a \cdot \frac{\sum x_i}{n} = \frac{731}{10} - 13.648 \cdot \frac{16.25}{10} = 50.922$$

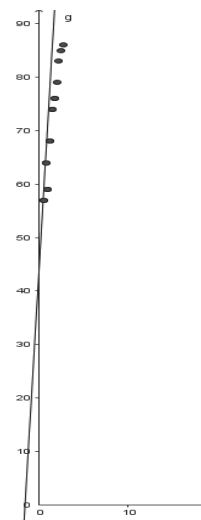
Jednadžba regresijskog pravca glasi: $\hat{y} = 13.648x + 50.922$.

Ako se neki učenik pripremao 0.25 h koji je njegov najvjerojatniji rezultat na ispitu:

$$\hat{y} = 13.648 \cdot 0.25 + 50.922 \approx 54$$

a da bi ostvario maksimalni broj bodova na testu, učenik se trebao pripremati:

$$100 = 13.648 \cdot x + 50.922 \Rightarrow x \approx 3.6 \text{ sati.}$$



Zadaci 6

1. Dane su sljedeće vrijednosti varijabli:

x_i	13.8	3.1	38.5	15.1	28.2	53.6
y_i	6	2.7	13.6	3.4	10.1	19.7

Nacrtajte dijagram rasipanja. Kako glasi jednačba regresijskog pravca?

Ucrtajte regresijski pravac.

2. Četiri bolnička centra za mjesec dana hospitalizirali su x_i pacijenata koji su imali troškove liječenja y_i . Podaci su iskazani u tablici. Nacrtajte dijagram rasipanja i izračunajte jednačbu regresijskog pravca. Procijenite koliki bi trošak liječenja bio u bolnici u kojoj je mjesečno hospitalizirano 1300 pacijenata.

bolnica	broj pacijenata (x_i)	troškovi liječenja u tisućama kunama (y_i)
A	1200	7200
B	950	5415
C	830	5040
D	1100	6490

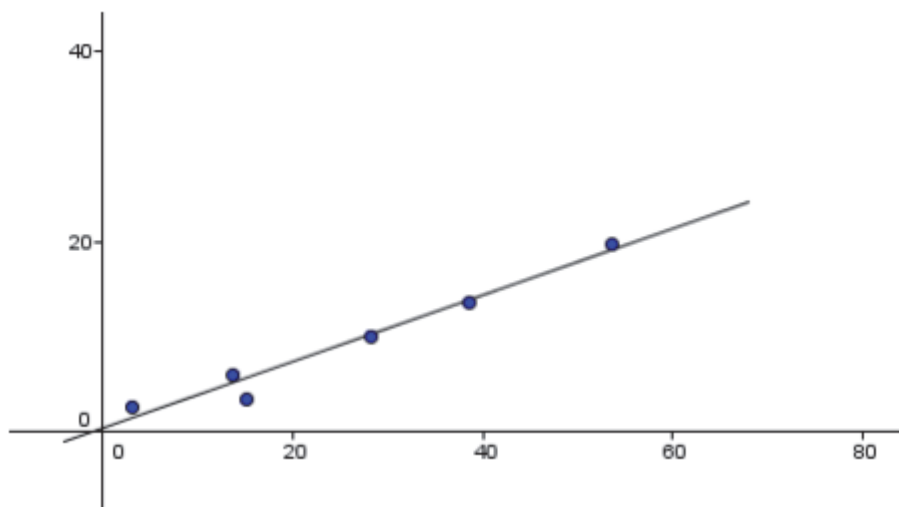
3. Osiguravajuće društvo želi odrediti vezu između broja sati provedenih na poslu tijekom tjedna i broja ozljeda na radu u tom razdoblju. Podaci kojima raspolaže su:

sati rada (x_i)	32	40	36	46	41	33
broj ozljeda (y_i)	0	2	2	5	3	1

Nacrtajte dijagram rasipanja i izračunajte jednačbu regresijskog pravca. Koliki je očekivani broj ozljeda ako osoba radi 44 sata tjedno?

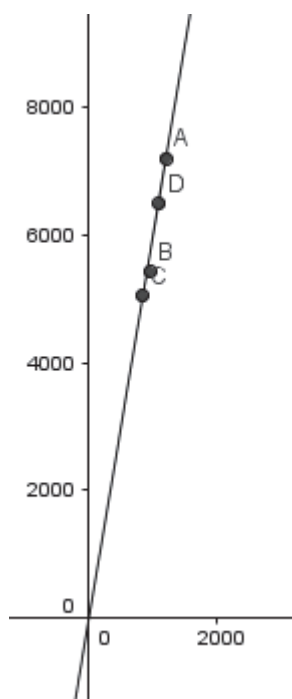
Rješenja 6

1.



$$\hat{y} = 0.35x + 0.39$$

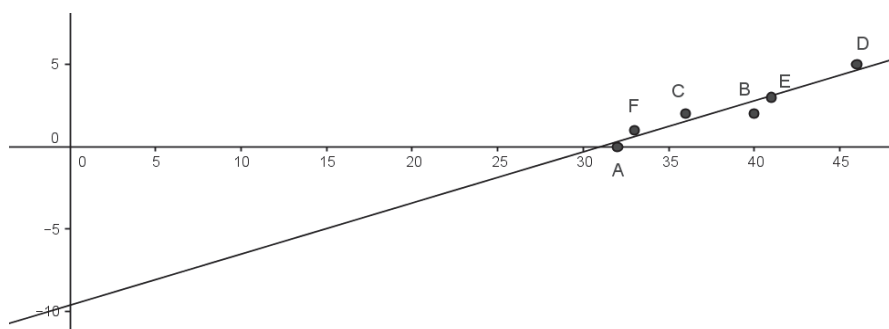
2.



$$\hat{y} = 5.997x - 80.69$$

Trošak liječenja za 1300 pacijenata: $\hat{y} = 7715.41$

3.



$$\hat{y} = 0.31x - 9.61$$

Očekivani broj ozljeda: $\hat{y} = 4$

VII STATISTIKA UZ POMOĆ PROGRAMA GEOGEBRA

Nakon upoznavanja klasičnog pristupa obradi i analizi podataka došlo je vrijeme za korištenje računala. Računalo je korisno za brzu i pouzdanu obradu velikog broja podataka, kvalitetnu statističku analizu i prezentaciju rezultata.

Za statističku obradu podataka postoji jako puno programa. Na školskim računalima instaliran je *Microsoft Office* s programom za tablični proračun *Excel*. U ovom priručniku upoznat ćemo se s mogućnostima GeoGebre, koja je postala moćan program za statističku analizu. GeoGebra je jednostavna za uporabu. Način rada s proračunskom tablicom sličan je onom kao i u drugim programima. Između GeoGebrinih prozora komunikacija je dvosmjerna. Osim što se promjena podataka u tablici istovremeno odražava na grafički prikaz, moguće je i obrnuto što s ostalim programima najčešće nije slučaj.

Možemo koristiti kao alat za obradu, analizu i prezentaciju podataka, ali možemo ju rabiti i za upoznavanje s temeljnim pojmovima statistike, njihovo bolje razumijevanje i zorno predočavanje.

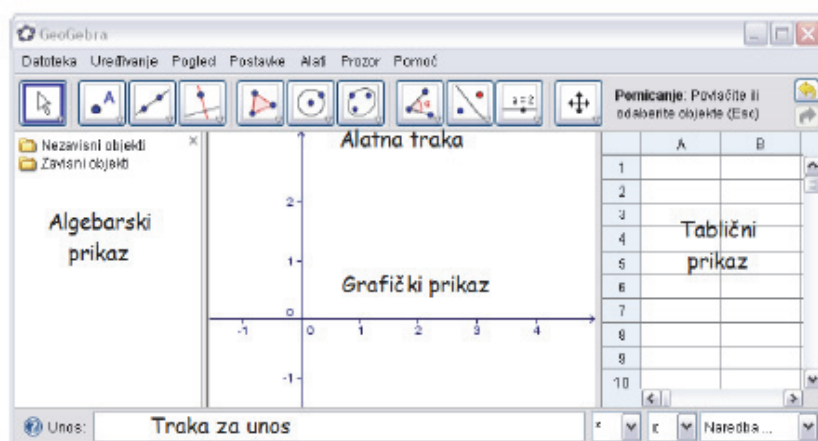
1. Instaliranje i pokretanje GeoGebre

Otvorite Internet preglednik i idite na www.geogebra.org. Kliknite na gumb *Downloads*.

i odaberite instaliranje za računalo, mobitel ili tablet.

Program se automatski instalira. Vi samo trebate potvrditi sve poruke koje bi se mogle pojaviti sa *OK* ili *YES*. Jednom kad instalirate GeoGebru možete raditi i bez uključene internetske veze.

Nakon pokretanja GeoGebre pojavljuje se sljedeći prozor.



GeoGebra ima tri različita načina promatranja matematičkih objekata: *grafički prikaz*, *algebarski prikaz* i *tablični prikaz*. Oni omogućavaju predstavljanje matematičkih objekata na tri različita načina: grafički (npr. točka, graf funkcije), algebarski (npr. koordinate točke, jednažba) ili u ćelijama tabličnog prikaza.

Svi prikazi jednog objekta su dinamički povezani i automatski će se promijeniti ako promijenite bilo koji prikaz, bez obzira na koji je način objekt izvorno kreiran.

2. Konstrukcije i naredbe u GeoGebri

Konstrukcije u GeoGebri možemo provesti korištenjem:

- a. miša i alata na *alatnoj traci*

Ako promatramo algebarski prikaz, stoje nam na raspolaganje sljedeći alati:

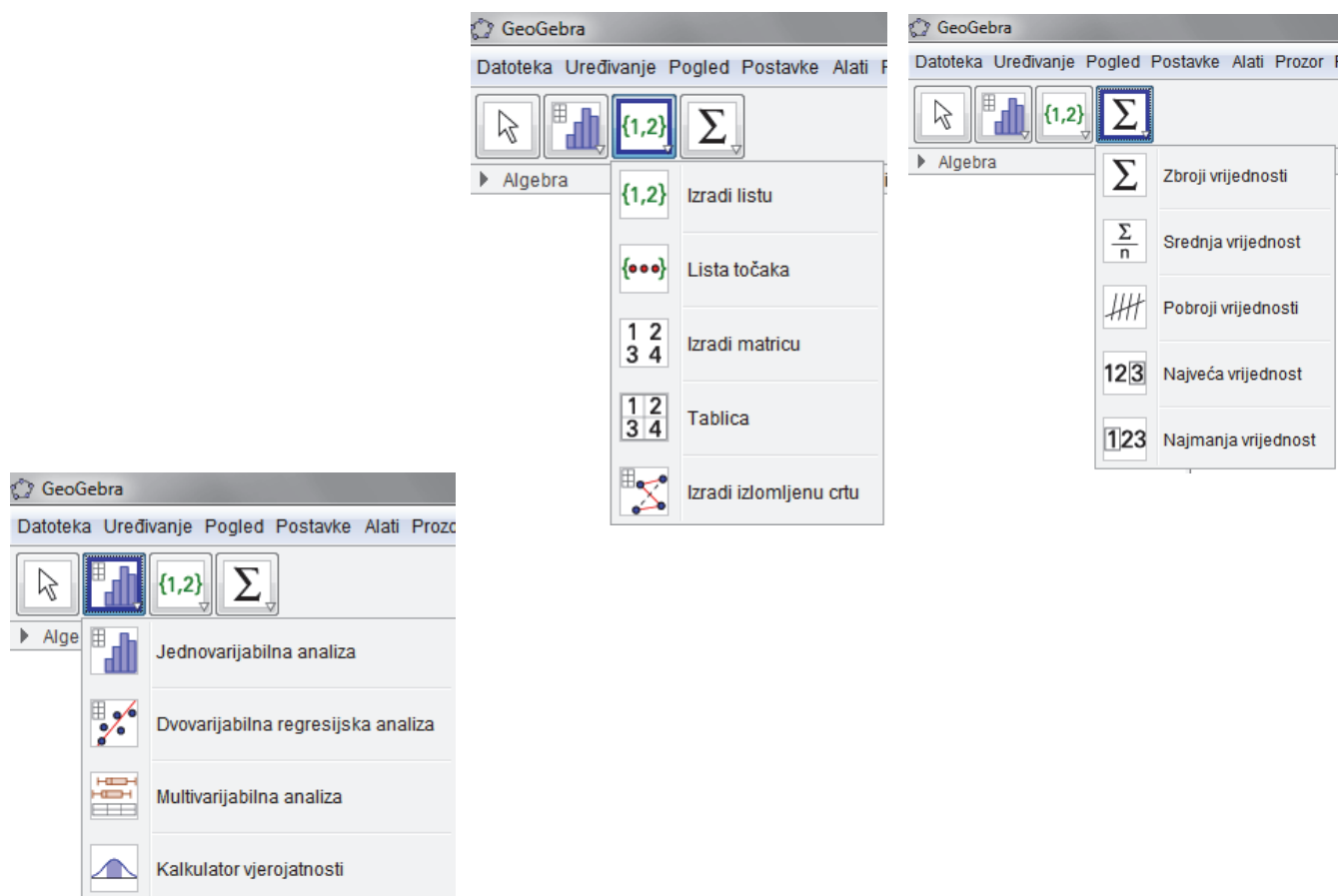


a ako smo u tabličnom prikazu:



Aktivirajte željeni alat klikom na gumb koji prikazuje odgovarajuću ikonu. Otvorite padajući izbornik alata klikom na strelicu u donjem desnom kutu odgovarajućeg gumba i odaberite drugi alat iz padajućeg izbornika alata. Prelaskom miša preko alata javlja se poruka s uputom kako taj alat radi.

Evo detaljnijeg pregleda statističkih alata:

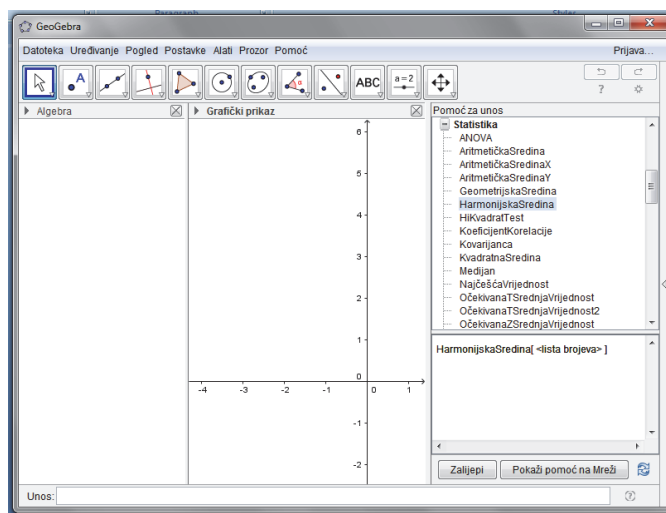


Za korištenje alata potrebno je odabrati stupac ili redak u *tablici*, a zatim kliknuti na alat.

b. korištenjem *trake za unos*

Obratite pažnju na malu ikonu (uokvireni upitnik) na desnom rubu *trake za unos*. Klikom na tu ikonu otvara se prozor sa sortiranim naredbama po područjima i ispod njega prozor u kojem se prikazuje sintaksa odabrane naredbe. Obratite i pažnju na gumb *Zalijepi* koji odabranu naredbu sa sintaksom izravno prenosi u *traku za unos*.

Naredbe se mogu unositi i upisom nekoliko prvih slova pa iz ponuđenih odabranu potvrdimo s tipkom *Enter*.



Raspoložive statističke naredbe koje unosimo u *traku za unos* odnose se na cijeli raspon podataka unesen u tablicu. Evo nekih naredbi koje možemo koristiti pri obradi podataka.

OPĆE NAREDBE ZA OBRADU PODATAKA

Jedinstveni[<lista neobrađenih podataka>]

Daje skup svih različitih vrijednosti koje se javljaju u skupu podataka poredanih po veličini.

Frekvencija[<lista neobrađenih podataka>]

Daje listu broja pojavljivanja svakog pojedinog elementa iz ulazne liste. U ulaznoj listi trebaju biti brojevi ili tekstovi. Ulazna lista se prethodno sortira prema rastućem poretку.

Razredi[<lista podataka L>, <broj razreda>]

Daje listu granica razreda. Prva je granica jednaka najmanjem elementu liste L , posljednja je granica jednaka najvećem elementu L , dok su udaljenosti između sljedećih granica razreda jednake.

DIJAGRAM NAREDBE

a. Dijagram pravokutnika

DijagramPravokutnika[pomak po y osi, skala na y osi, lista neobrađenih podataka]
Daje dijagram pravokutnika koristeći dane neobrađene podatke. Vertikalnu poziciju u koordinatnom sustavu kontrolira varijabla <i>pomak po y osi</i> dok na visinu utječe koeficijent <i>skala na y osi</i> .
<u>Primjer:</u>
DijagramPravokutnika[0, 1, {2,2,3,4,5,5,6,7,7,8,8,8,9}]
DijagramPravokutnika[pomak po y osi, skala na y osi, početna vrijednost, Q1, medijan, Q3, završna vrijednost]
Daje dijagram pravokutnika za dane statističke podatke na intervalu <i>[početna vrijednost, završna vrijednost]</i> .

b. Dijagram StabloList

DijagramStabloList[<lista>]
Daje dijagram stablo-list zadane liste brojeva. Brojevi koji značajno odstupaju ne prikazuju se u dijagramu nego izdvojeno. Značajno odstupajući brojevi su oni koji su izvan intervala $[Q1 - 1.5 (Q3 - Q1), Q3 + 1.5 (Q3 - Q1)]$, gdje su Q kvartili.

c. Dijagram točaka

DijagramTočaka[<lista neobrađenih podataka>]
Crta dijagram točaka za zadanu listu neobrađenih podataka (kao listu točaka).
Za svaki broj "a" koji se pojavljuje u listi "k" puta, dijagram daje točke $(a, 1), (a, 2), \dots, (a, k)$.

d. Histogram**Histogram[lista granica razreda, lista visina]**

Proizvodi histogram s pravokutnicima zadane visine. Granice razreda određuju širinu i poziciju svakog pravokutnika histograma.

Primjer: Histogram[{0, 1, 2, 3, 4, 5}, {2, 6, 8, 3, 1}]

izradit će histogram s 5 pravokutnika zadane visine. Prvi je pravokutnik pozicioniran nad intervalom $[0, 1]$, drugi nad intervalom $[1, 2]$, i tako dalje.

Histogram[lista granica razreda, lista neobrađenih podataka]

Proizvodi histogram koristeći neobrađene podatke. Granice razreda određuju širinu i poziciju svakog pravokutnika histograma i koriste se za određivanje koliko elemenata pripada svakom razredu.

Primjer: Histogram[{1, 2, 3, 4}, {1.0, 1.1, 1.1, 1.2, 1.7, 2.2, 2.5, 4.0}]

izraditi će histogram s 3 pravokutnika, visine 5 (prvi pravokutnik), 2 (drugi pravokutnik), i 1 (treći pravokutnik).

e. Poligon frekvencija**PoligonFrekvencija[<lista granica razreda>, <lista visina>]**

Crta poligon frekvencija s vrhovima u danim visinama. Granice razreda određuju x koordinatu svakog vrha.

f. Stupčasti dijagram

StupčastiDijagram[lista neobrađenih podataka, širina stupaca]
Proizvodi stupčasti dijagram koristeći zadane neobrađene podatke, stupci su zadane širine.
Primjer: StupčastiDijagram[{1,1,1,2,2,2,2,3,3,3,5,5,5}, 1]
StupčastiDijagram[lista podataka, lista frekvencija]
Proizvodi stupčasti dijagram koristeći listu podataka s odgovarajućim frekvencijama.
Napomena: Lista podataka treba sadržavati brojeve koji rastu za stalni iznos.
Primjeri: StupčastiDijagram[{10,11,12,13,14}, {5,8,12,0,1}] StupčastiDijagram[{5, 6, 7, 8, 9}, {1, 0, 12, 43, 3}] StupčastiDijagram[{0.3, 0.4, 0.5, 0.6}, {12, 33, 13, 4}]
StupčastiDijagram[lista podataka, lista frekvencija, širina stupaca w]
Proizvodi stupčasti dijagram koristeći listu podataka s odgovarajućim frekvencijama, sa stupcima širine w.
Napomena: Lista podataka treba sadržavati brojeve koji rastu za stalni iznos.
Primjeri:
StupčastiDijagram[{10,11,12,13,14}, {5,8,12,0,1}, 0.5] (ostavlja razmak između stupaca)
StupčastiDijagram[{10,11,12,13,14}, {5,8,12,0,1}, 0] (izraditi će linijski dijagram)

SREDNJE VRIJEDNOSTI

a. Aritmetička sredina

AritmetičkaSredina[lista brojeva]

b. Geometrijska sredina

GeometrijskaSredina[<lista brojeva>]
--

c. Harmonijska sredina

HarmonijskaSredina[<lista brojeva>]

d. Mod

NajčešćaVrijednost[<lista brojeva>]

Primjeri:

NajčešćaVrijednost[{1,2,3,4}] daje praznu listu {}

NajčešćaVrijednost[{1,1,1,2,3,4}] daje praznu listu {1}

NajčešćaVrijednost[{1,1,2,2,3,3,4}] daje praznu listu {1,2,3}

e. Medijan

Medijan[<lista brojeva>]

Primjeri:

Medijan[{1,2,3}] daje 2

Medijan[{1,1,3,4,8,8}] daje 3.5.

MJERE RASPRŠENOSTI

a. Kvartili

Q1[lista brojeva]
određuje donji kvartil zadanih brojeva
Q3[lista brojeva]
određuje gornji kvartil zadanih brojeva

b. Varijanca

Varijanca[lista brojeva]

c. Standardna devijacija

SD[lista brojeva]

LINEARNA PRILAGODBA (REGRESIJA)

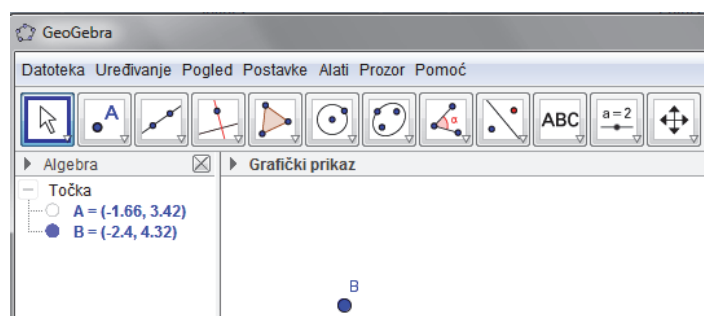
PrilagodbLinearne[lista točaka]
Izračunava (y ovisi o x) pravac regresije zadanih točaka.
PrilagodbLinearneX[lista točaka]
Izračunava (x ovisi o y) pravac regresije zadanih točaka.

Najjednostavnije da skupove podataka do kojih smo sami došli unesemo u ćelije tabličnog prikaza. Svaka ćelija tabličnog prikaza ima specifični naziv čime je omogućeno izravno adresiranje svake ćelije. Na primjer, ćelija u stupcu A i redu 1 se zove ćelija A1. Podatke možemo prenijeti i direktno. Označimo ih npr. u *Excelu*, koristimo naredbu *Copy* i u *Zalijepimo* ih u *GeoGebri*.

Naziv ćelije se može upotrijebiti u naredbama kako bi se adresirao sadržaj pripadajuće ćelije. U ćelije možete unijeti ne samo brojeve, već sve tipove matematičkih objekata (na primjer, koordinate točaka, funkcije, naredbe).

Za velike količine unesenih podataka praktično je kreirati *Liste*. Označimo raspon ćelija u tabličnom prikazu, kliknemo desnim klikom na ćeliju i odaberemo *Izradi>Listu*. Kasnije se u pisanju naredbe pozovemo na tu *Listu*.

GeoGebra će odmah po unosu objekta prikazati njegov grafički prikaz. Ako ne želimo prikaz tog objekta možemo ga sakriti, klikom u *algebarskom prikazu*. Ukoliko je kružić objekta ispunjen bojom one će biti vidljiv u *grafičkom prikazu*, u suprotnom bit će skriven.



3. Vježba

Zadatak 1 (stupčasti dijagram i tablica frekvencija)

Analizirajte učestalost izvučenih brojeva na igri Hrvatske lutrije 7 od 39 tijekom 2010.godine. Podatke prikažite stupčastim dijagramom i tablicom frekvencija. Jesu li izvučeni svi brojevi? Koji su najmanje puta, a koji najviše? Koliki je raspon frekvencija?

Pripreme

U izborniku *Pregled* odaberite *Tablica*, *Algebra* i *Grafički prikaz*.

Na stranicama Hrvatske lutrije mogu se pronaći traženi podaci u *Excel* tablici.

Koraci konstrukcije

1			Upišite pronađene podatke u <i>GeoGebra</i> tablicu tako da ih označite u <i>Excelu</i> , koristite naredbu <i>Copy</i> i u <i>Zalijepite</i> ih u <i>GeoGebri</i> .
2			Odaberite pripadajuće ćelije i kliknite desnim klik miša i u padajućem izborniku odaberite <i>Izradi>Lista</i> .
3			U <i>traku za unos</i> teksta, na dnu ekrana, unesite <i>Jedinstveni[lista1]</i> i potvrdite <i>Enterom</i> . Time smo dobili <i>listu2</i> u kojoj su sve vrijednosti koje se javljaju u skupu podataka poredane po veličini.
4			Ponovo, U <i>traku za unos</i> teksta, unesite <i>Frekvencija[lista1]</i> i potvrdite <i>Enterom</i> . Time smo dobili <i>listu3</i> u kojoj su vidi učestalost pojavljivanja svake vrijednosti.
5			U <i>traku za unos</i> teksta, na dnu ekrana, unesite <i>StupčastiDijagram[lista2, lista3, 0.8]</i> i potvrdite <i>Enterom</i> . (koristili smo naredbu <i>StupčastiDijagram</i> [lista podataka, lista frekvencija, širina stupaca w] gdje 0.8 predstavlja širinu stupaca)
6			U <i>traku za unos</i> teksta, na dnu ekrana, unesite <i>TablicaFrekvencija[lista1]</i> i potvrdite <i>Enterom</i> . U grafičkom prikazu vidimo tablicu frekvencija kao tekst.

Neki savjeti



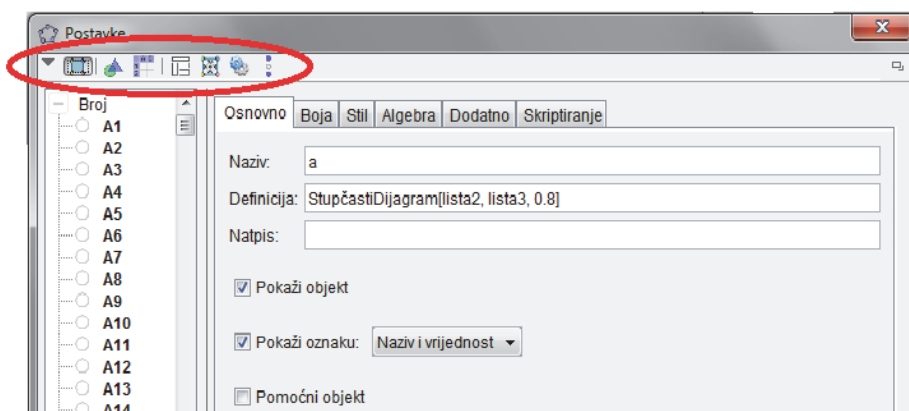
Odaberite na *Alatnoj traci* alat za *Tekst*.

Umetnite naslovni tekst *Loto 7/39*.

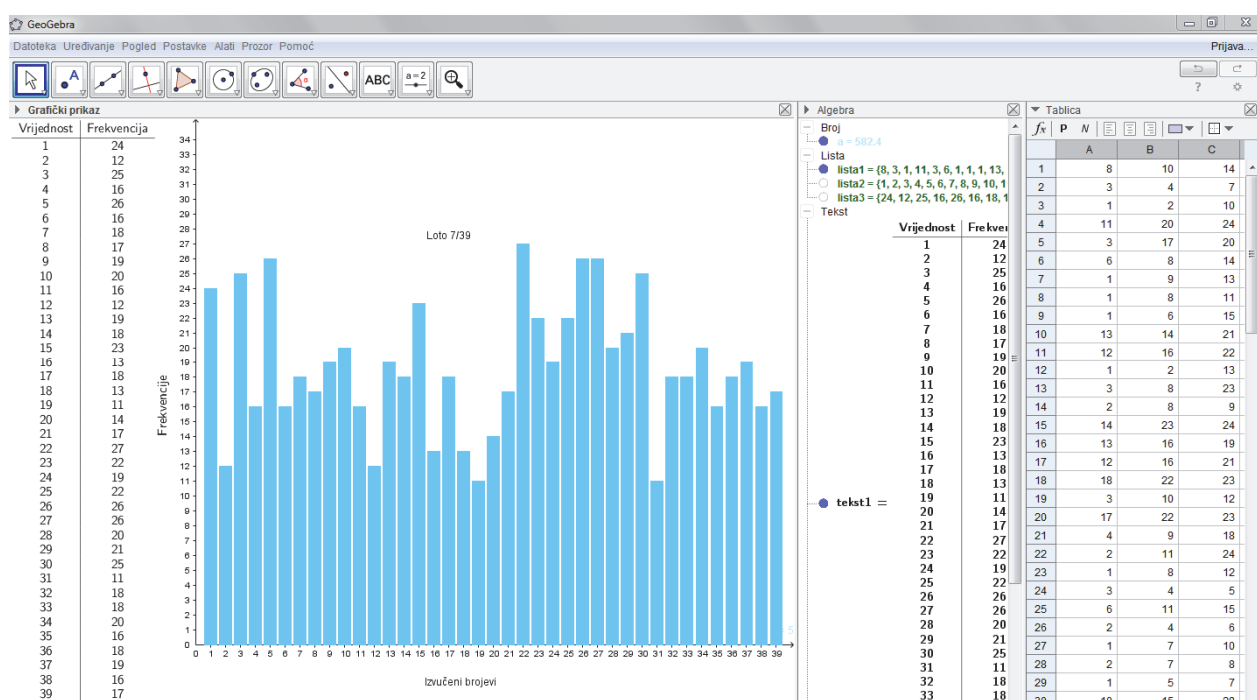
Umetnite i nazive za podatke na *x* osi *Izvučeni brojevi* te *Frekvencije* na *y* osi.

Želimo li tekst *Frekvencije* zarotirati za 90° moramo u *alatnu traku* unijeti naredbu *RotirajTekst*["frekvencije", 90°].

Lako možete mijenjati izgled objekta (boje, tip, debljinu crte...). Upotrijebite desni klik na mišu dijagrama i odaberite *Svojstva*. Na vrhu izbornika nalaze se kartice s postavkama za *Svojstva objekta*, *Grafički prikaz*, *Tablicu*...Istražite koje sve mogućnosti postoje.



U kartici *Svojstva>Boja* možete promijeniti boju stupaca. U kartici *Postavke-Grafički prikaz>xOs>yOs* možete sakriti negativne dijelove koordinatnih osi.



Zadatak 2. (histogram)

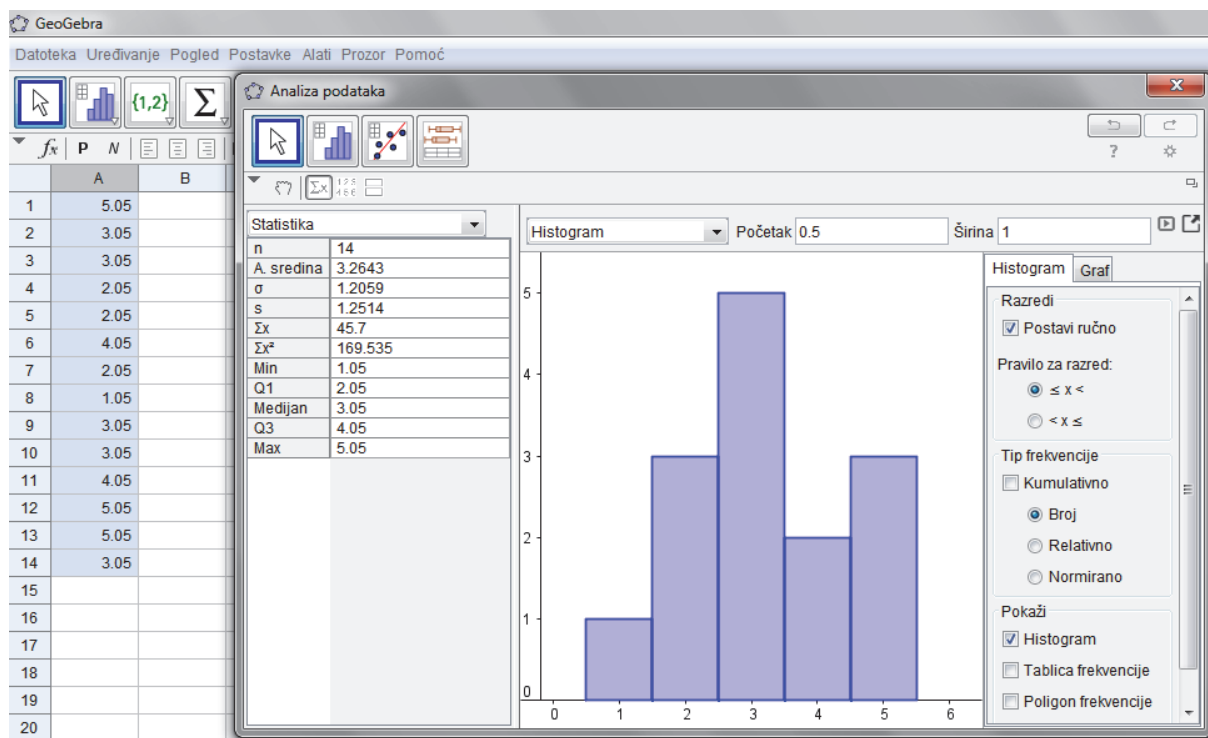
Napravite histogram i očitajte aritmetičku sredinu, medijan, minimum i maksimum unesenih vrijednosti.

Pripreme

U izborniku *Pregled* odaberite *Tablica* i *Grafički prikaz*.

Koraci konstrukcije

1		Upišite podatke u stupac A u tablici: 5,3,3,2,2,4,2,1,3,3,4,5,5,3
2		Odaberite pripadajuće ćelije od A1 do A14 i odaberite alat <i>Jednovarijabilna analiza</i> .
3		Odaberite odgovarajući broj <i>Razreda</i> na vrhu prozora. U ovom primjeru odaberite <i>5 Razreda</i> , zato što imamo pet različitih veličina.
4		Klikom na strelicu u gornjem desnom kutu u izborniku <i>Histogram</i> označite <i>Razredi>Postavi ručno</i> . Za <i>Početak</i> upišite 0.5 i pritisnite tipku <i>Enter</i> , a za <i>Širinu</i> upišite 1 i pritisnite <i>Enter</i> .
5		Pronađite aritmetičku sredinu, medijan, maksimum i minimum na lijevoj strani u statističkom dijelu prozora.



Zadatak 3. (brkata kutija)

Dijagram pravokutnika ili *Brkatu kutiju* nismo posebno obrađivali unutar priručnika. On je koristan prikaz iz kojeg možemo lako očitati medijan, donji i gornji kvartil, raspon podataka i ekstremne vrijednosti. Podsjetimo se, kvartili su mjere raspršenosti skupine podataka. Donji kvartil dijeli podatke tako da je ispod njega 25%, a iznad njega 75% svih podataka. Gornji kvartil obrnuto. Brkata kutija zorno predočava te granice.

Nacrtajmo brkatu kutiju za sljedeći niz podataka:

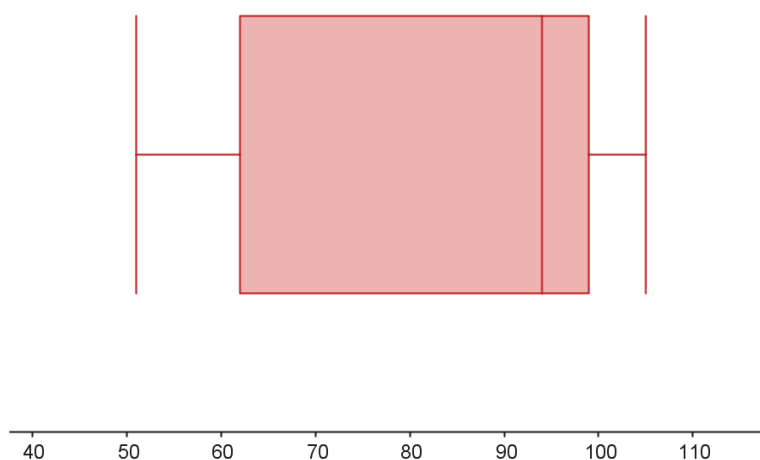
51, 62, 78, 94, 96, 99, 105

Pripreme

U izborniku *Pregled* odaberite *Tablica* i *Grafički prikaz*.

Koraci konstrukcije

1		Upišite podatke u stupac A u tablici: 51, 62, 78, 94, 96, 99, 105
2		Odaberite pripadajuće ćelije od A1 do A7 i odaberite alat <i>Jednovarijabilna analiza>Pravokutni dijagram</i> .
5		Odaberite gumb za Statistiku i provjerite odgovara li nacrtani dijagram očitanim podacima.



Iz dijagrama možemo očitati donji kvartil $Q1=62$, gornji kvartil $Q3=99$, srednji kvartil ili medijan $M=94$, a interkvartilni raspon $Q3-Q1=37$.

Zadatak 4 (kružni dijagram)

U GeoGebri ne postoji alat kojim se crta kružni dijagram. Potrebno je crtati kružne isječke.

Izradimo kružni dijagram koji prikazuje broj rođenih po mjesecima u Krapinsko-zagorskoj županiji u 2009.. godini. Te podatke pronaći ćemo na stranici Državnog zavoda za statistiku (www.dzs.hr).

3.2. ŽIVOROĐENI PREMA MJESECU ROĐENJA U 2009. PO ŽUPANIJAMA
LIVE BIRTHS, BY THE MONTH OF BIRTH, BY COUNTIES, 2009

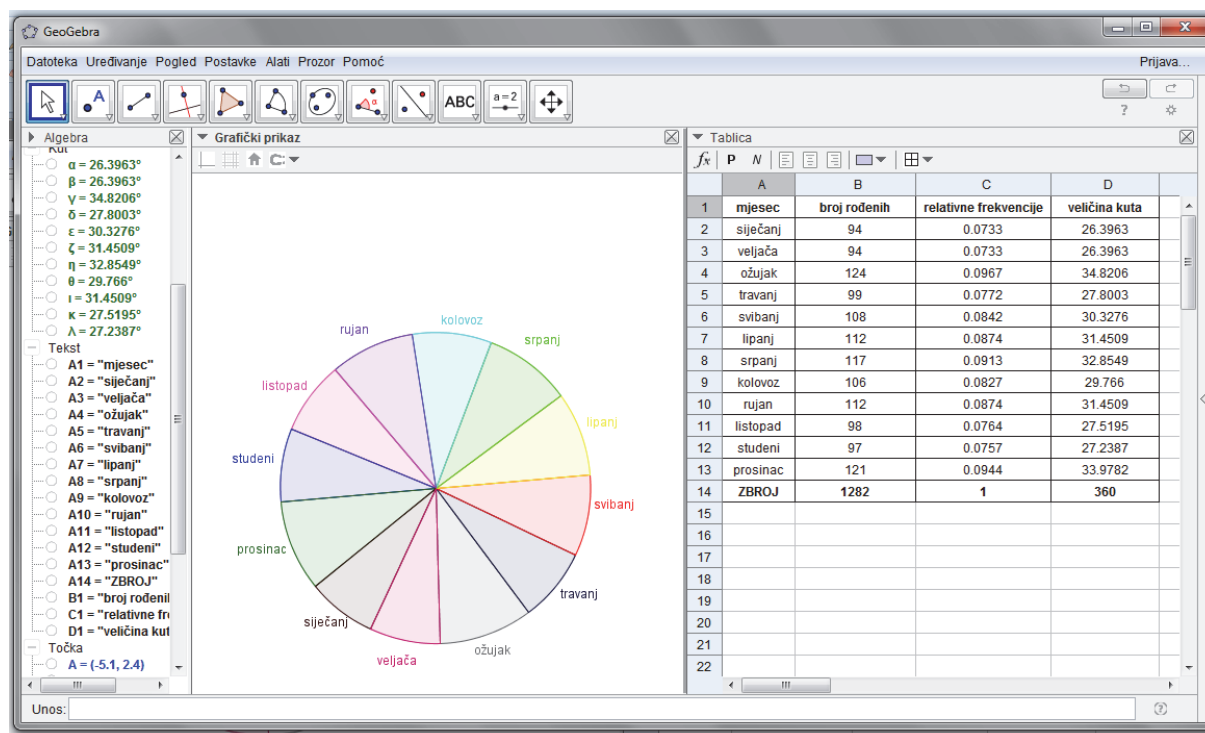
	Ukupno Total	Mjesec rođenja Month of birth											
		I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.
Republika Hrvatska Republic of Croatia	44 577	3 767	3 426	3 563	3 292	3 461	3 554	4 180	3 935	4 118	3 823	3 600	3 858
Zagrebačka županija County of Zagreb	3 407	297	266	267	251	250	274	328	300	309	316	262	287
Krapinsko-zagorska županija County of Krapina-Zagorje	1 282	94	94	124	99	108	112	117	106	112	98	97	121

Pripreme

U izborniku *Pregled* odaberite *Tablica*, *Algebra* i *Grafički prikaz*.

Koraci konstrukcije

1		Upišite podatke u <i>tablični prikaz</i> kao što je zadano u zadatku. Izračunajte ukupan broj rođenih u Krapinsko-zagorskoj županiji za 2009.g.
2		Dodajte stupac i izračunajte relativne frekvencije za svaki mjesec (npr. broj rođenih u siječnju dijelimo s ukupnim brojem rođenih u 2009.) Unosimo u tablicu nazive ćelija: =B2 / B14 i potvrdimo <i>Enterom</i> .
3		U zadnjem stupcu izračunajmo veličine kutova pripadnih kružnih isječaka: npr. za siječanj računamo: =C2*360 i potvrdimo <i>Enterom</i> . Napomena: zbroj svih relativnih frekvencija mora biti 1, a zbroj svih veličina kutova 360°.
4		U <i>grafičkom prikazu</i> , desnim klikom na crtaću plohu, sakrijte koordinatne osi.
5		Odaberite alat <i>Kut zadane veličine</i> na <i>alatnoj traci</i> . Kliknite na crtaću plohu dva puta da bi nacrtali točku kraka A i vrh B. Zatim će se pojaviti prozor u koji moramo unijeti veličinu kuta. Unesemo za siječanj veličinu kuta 26.3963°. Dobili smo točku A', odnosno kut ABA' veličine 26.3°. Ponovimo postupak za veljaču. Odaberemo za prvu točku kraka točku A', zatim vrh B i unesemo veličinu kuta za ožujak. Ponavljamo postupak za sve mjesece.
		Klikom na kružice u <i>algebarskom prikazu</i> sakrijmo oznake veličina kutova, i desnim klikom na svaku točku isključimo <i>pokaži oznaku</i> .
		Da bismo završili konstrukciju moramo izabrati alat <i>Kružni isječak</i> , kliknuti prvo na središte a potom na dvije točke. Nacrtajte sve kružne isječke. Da biste promijenili njihove oznake, kliknite na kružni isječak i odaberite <i>Preimenuj s prikazanog izbornika</i> , a da biste promijenili njihove boje, kliknite na kružni isječak i odaberite <i>Svojstva objekta>Boja s prikazanog izbornika</i> .



Zadatak 5. (stablo-list)

Dani su rezultati testa iz hrvatskog jezika u dva razreda,

3.a	39, 50, 23, 25, 27, 23, 22, 34, 30, 46, 37, 46, 16, 14, 27, 29, 34, 36, 37, 37, 48, 48, 33
3.b	34, 35, 5, 38, 45, 44, 12, 7, 19, 19, 23, 22, 46, 26, 27, 27, 8, 19, 22, 34, 37, 18, 20

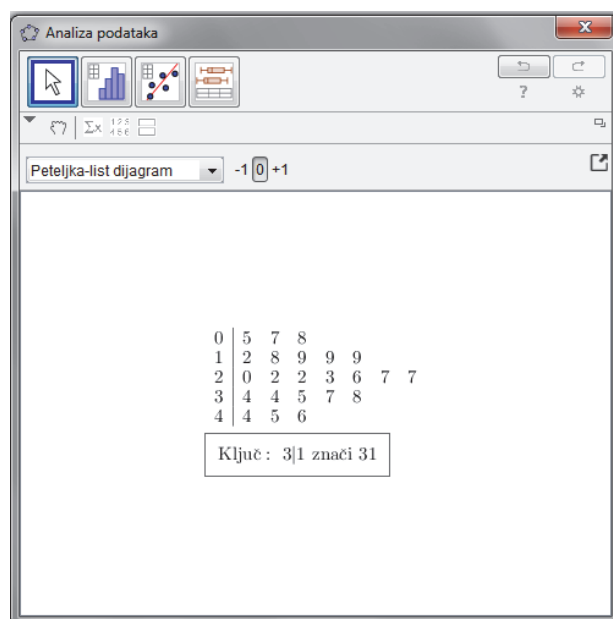
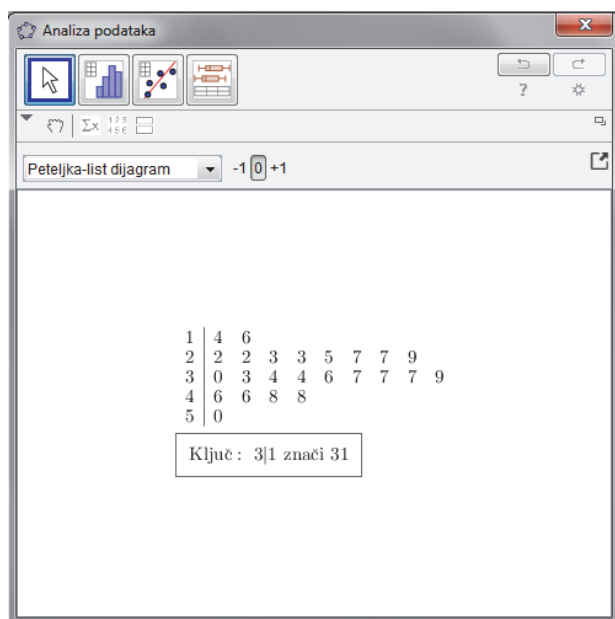
Napravite stablo-list dijagram i protumačite rezultate.

Pripreme

U izborniku *Pregled* odaberite *Tablica* i *Grafički prikaz*.

Koraci konstrukcije

1		Upišite podatke u stupce A i B <i>tabličnog prikaza</i> .
2		Odaberite pripadajuće ćelije od A1 do A24 i odaberite alat <i>Jednovarijabilna analiza>Peteljka-list dijagram</i> .
3		Ponovite postupak za stupac B.



Kad su podaci ovako zorno prikazani, odmah možemo zaključiti o rezultatima. U 3.a razredu najlošiji rezultat je 14 bodova, dok su u 3.b razredu čak tri rezultata ispod 10 bodova. U 3.a postignut je maksimalni broj bodova, što se nije dogodilo u 3.b. Međutim, rezultati u 3.b su manje raspršeni i imaju tendenciju grupiranja u nižim vrijednostima. Kod rezultata 3.a razreda ističe se grupa od 30 do 39 u kojoj se nalazi najveći broj podataka.

Zadatak 6. (pravac regresije i dijagram rasipanja)

Tablica prikazuje podatke o broju umrlih u dobi od 35 do 64 godina od bolesti srca, te godišnje količine popušanih cigareta po osobi za 21 zemlju svijeta.

Zemlja	Godišnja konzumacija cigareta po osobi	Stopa smrtnosti na 100 000 ljudi
Australia	3220	238
Austrija	1770	182
Belgija	1700	118
Canada	3350	212
Danska	1500	145
Finska	2160	233
Francuska	1410	60
Grčka	1800	41
Island	2770	111
Irska	2770	187
Italija	1510	114
Meksiko	1680	32
Nizozemska	1810	125
Novi Zeland	3220	212
Norveška	1090	136
Španjolska	1200	44
Švedska	1270	127
Švicarska	2780	125
Velika Britanija	2790	194
SAD	3900	257
Njemačka	1890	150

Napravite dijagram rasipanja, odredite jednadžbu pravca regresije i izračunajte:

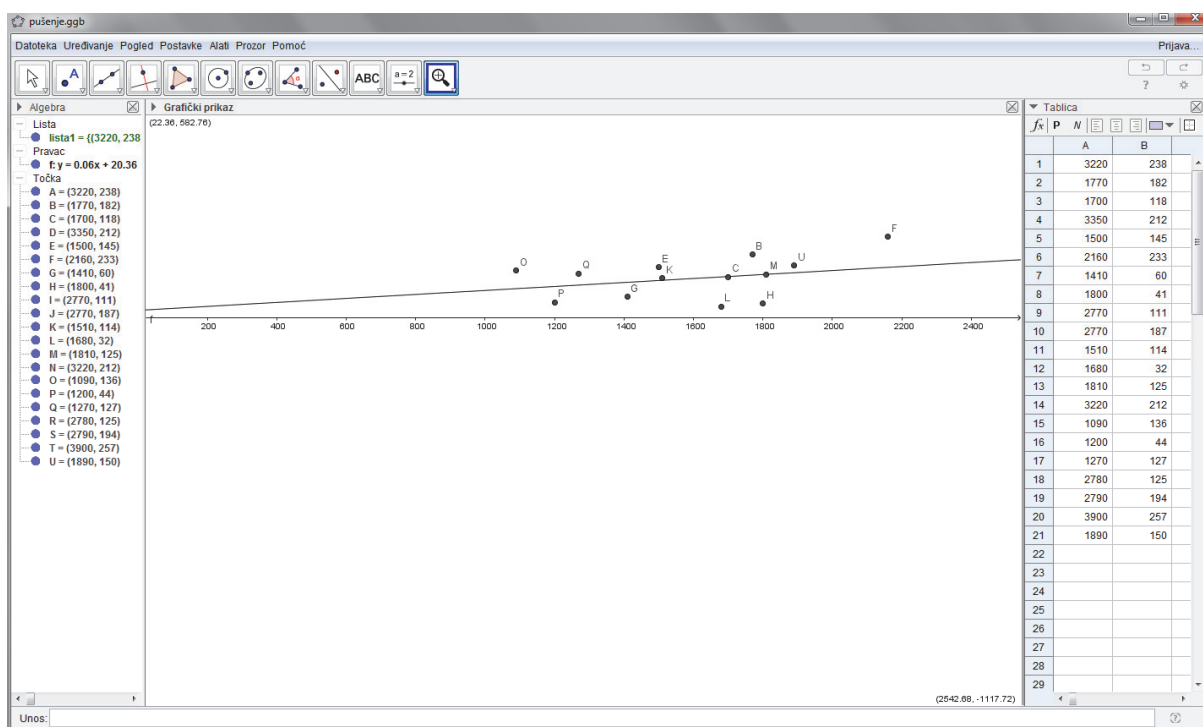
- prosječnu stopu smrtnosti za zemlju u kojoj je godišnja konzumacija 2 000 cigareta po osobi
- koliko prosječno cigareta konzumiraju državljani zemlje u kojoj je stopa smrtnosti 240 na 100 000 ljudi.

Pripreme

U izborniku *Pregled* odaberite *Tablica* i *Grafički prikaz*.

Koraci konstrukcije

1		Upišite podatke u stupce A i B <i>tabličnog prikaza</i> .
2		Označite sve ćelije, kliknite desnom tipkom miša i odaberite <i>Izradi>Lista točaka</i> . GeoGebra će odmah prikazati te točke u koordinatnom sustavu.
3		U traku za unos unesemo naredbu: <i>PrilagodbaLinearna[lista1]</i> GeoGebra će nacrtati pravac regresije i u <i>algebarskom prikazu</i> prikazati njegovu jednadžbu: $y=0.06x+20.36$



Riješimo zadatak do kraja:

- Za $x=2000$, izračunamo $y=140.36$.
- Za $y=240$, x je približno 3 360.

VIII PROJEKTA NASTAVA

1.1. Što je projektna nastava?

Projekt (lat. plan, namjera, nacrt, skica) je svaki zaokružen, cjelovit i složen pothvat čija se obilježja i cilj mogu definirati, a mora se ostvariti u određenom vremenu te zahtijeva koordinirane napore nekoliko ili većeg broja ljudi, službi, poduzeća i sl.

U okviru projektne nastave, podjela projekata može se provesti na nekoliko načina:

- prema vremenu trajanja istraživanja (dnevni, tjedni, mjesečni, tijekom polugodišta ili tijekom cijele školske godine)
- prema broju uključenih subjekata (individualni, u paru, u grupni, u razredni ili školi)
- prema metodama rada (teorijski ili eksperimentalni) (Cindrić, 2006).

Za svaki (dobar) projekt je važno da ima cilj i rezultira proizvodom (produktom), podrazumijeva složeniji zadatak (koji se razlaže na jednostavnije, rutinske), ima određeno trajanje, tj. rok za završetak, u pravilu uključuje rad više (grupa) ljudi i podrazumijeva suradnju i koordinaciju svih (grupa) sudionika.

Učenički **matematički projekt** podrazumijeva rad tima učenika na složenijem matematičkom problemu, često bliskom realnom svijetu.

1.2. Koji su ciljevi projektne nastave?

- razvijanje samostalnosti i odgovornosti u radu
- razvijanje kompetencije planiranja
- osamostaljivanje učenika u radu
- osposobljavanje učenika za korištenje raznih sredstava za rad i pravilnu procjenu raspoloživih resursa
- prakticiranje samostalnog pronalaženja informacija i kritički odnos prema njima
- osposobljavanje učenika za integriranje zajednički dobivenih radnih rezultata
- razvijanje osobne odgovornosti za realizaciju projekta
- upoznavanje s procesom učenja i rada u jednoj socijalnoj grupi
- stvaranje proizvoda za vlastito korištenje ili za korištenje od strane drugih
- savladavanje slučajeva razilaženja teorije i prakse, kao i strogo postavljenih granica nastavnih predmeta, uzimajući u obzir aspekte koji nadilaze struku
- učenje izvorne stvarnosti

1.3. Realizacija jednog matematičkog projekta

0. FAZA – Priprema projekta

- odabir projektne teme
- priprema učenika za temu (potrebno predznanje...)
- organizacija projekta (vremenska, prostorna organizacija, potrebni resursi prema broju učenika i grupa)

1. FAZA – Postavljanje problema

- postavljanje problema - problem mora biti postavljen dovoljno otvoreno i složeno (po mogućnosti, da izlazi iz okvira matematike)
- kratko odgovoriti na slijedeća pitanja:
 - opis problema - zašto?
 - ciljevi - što?
 - ciljane grupe - za koga?
 - aktivnosti
- brainstorming (navala ideja)
- sređivanje i grupiranje ideja – razlaganje problema u parcijalne probleme
- izbor parcijalnih problema koji će se obraditi
- preciziranje zadataka
- određivanje strukture tima
- podjela u grupe

2. FAZA – Rad u grupama

- prikupljanje radnih materijala (izvori informacija)
- međurezime rada:
 - izvještaji o radu grupa,
 - prezentacija međurezultata, kritika
 - nove smjernice od tima ili nastavnika

3. FAZA – Integriranje radnih rezultata

- završni izvještaji grupa i prezentacije rezultata
- podnošenje dokumentacije, dopune
- integriranje rezultata u cjeloviti proizvod
- vrednovanje ostvarenog, kritika
- finalno sređivanje

4. FAZA – Matematički doprinos

- sređivanje matematičkih rezultata
- uspostavljanje veze s nastavnim programom
- prezentacija proizvoda korisnicima

Učenički matematički projekti trebaju biti orijentirani na proizvode:

- matematički panoi, plakati
- matematički članci, časopisi, web stranice
- predavanja za druge učenike
- radni materijali za druge učenike.

1.4. Uloga učenika i nastavnika u projektu

Ključna kvalifikacija učenika uključenih u projekt je sposobnost timskog rada.

Od učenika se očekuje:

- suradnički rad i savladavanje konflikta u grupi
- sposobnost artikuliranog iskazivanja vlastitog mišljenja
- sposobnost prihvatanja tuđih mišljenja
- razvijanje sposobnosti za uzajamnu kritiku vlastitog i tuđeg rada
- davanje inicijative za rad sukladno njihovom interesu i sposobnostima
- traženje rješenja problema
- aktivno planiranje svih etapa i tijeka rada
- sudjelovanje u samoocjenjivanju i vrednovanju rezultata rada
- uviđanje nužnosti i smisla podjele rada

Nastavnik ima ulogu menadžera projekta. On preuzima sljedeće zadatke:

- izbor projektnih tema (u većini slučajeva)
- podjela u grupe, ponekad i zajednički rad s učenicima
- vođenje pri preciziranju postavljanja zadataka

- organiziranje podnošenja međuizvještaja
- priprema pomoći u radu
- odgovoran je za organizaciju i način rada projektnog tima
- mora biti u stanju prepoznati probleme i pokušati ih konstruktivno i sustavno riješiti

1.5. Primjer projektne teme

Primjer Izbjeglička kriza u Europi (na tlu Zagorja)

Polazište za brainstorming: film *Izbjeglička kriza u 6 minuta*. <https://youtu.be/RvOnXh3NN9w>

- izbjeglička kriza i statistika (opis problema, ciljevi, ciljane grupe, aktivnosti)
- svaki učenik ideje bilježi na kartice
 - koje vrste podataka možemo statistički obrađivati
 - koji učenici tj. razredi će sudjelovati u istraživanju
 - izrada ankete- osmišljavanje pitanja
 - analiza ankete- po kojim kriterijima
- preciziranje zadataka
- podjela u grupe
- podjela zadataka
- prikupljanje radnih materijala (izvori informacija)
- tabeliranje
- grupiranje podataka
- grafičko prikazivanje
- analiza podataka korištenjem statističkih metoda
- izrada plakata, prezentacija obrađenih sadržaja

PRILOG 1_ Anketa o izbjegličkoj krizi

PRILOG 2_ Neobrađena tablica podataka

PRILOG 3_ Prezentacija rezultata na njemačkom jeziku.

LITERATURA

1. Šošić, I. (2014). *Statistika*. Zagreb: Školska knjiga.
2. Brleković, K., Noskov, J. (2014). *Matematika 1, Matematika 2, udžbenik sa zbirkom zadataka za 1., 2. razred medicinskih škola*. Zagreb: Školska knjiga.
3. Salomon Padjen, Đ., Šego, B., Škrinjarčić, T. (2014). *Matematika 1, Matematika 2, Matematika 3, Matematika 4, udžbenik sa zbirkom zadataka za strukovne škole*. Zagreb: Alka script.
4. Mrkonjić, I., Šišić, S. (2014). *Matematika 2, udžbenik sa zbirkom zadataka za 2. razred medicinskih i zdravstvenih škola s dvogodišnjim programom matematike*. Zagreb: Neodidacta d. o. o.
5. Bakula, B., Varošaneć, S. (2014). *Matematika 1, udžbenik i zbirka zadataka za 1.razred medicinskih škola*. Zagreb: Element.
6. Bakula, B., Varošaneć, S. (2014). *Matematika 2, udžbenik i zbirka zadataka za 2.razred medicinskih škola*. Zagreb: Element.
7. Kovač Striko, E., Fratović, T., Ivanković, B.(2008). *Vjerojatnost i statistika*. Zagreb: Fakultet prometnih znanosti.
8. Bennett, A., Nelson, L. (2001). *Mathematics for elementary teachers*. New York: McGraw Hill.
9. Šuljić, Š. (2011.). Statistika uz pomoć programa GeoGebra 4.0. *Matematika i škola*, 61 (10), 39-43.

PRILOG 1

Ova je anketa anonimna. Na svako pitanje odaberi odgovor koji je najbliži tvom mišljenju, a ako nemaš mišljenje o određenom pitanju odgovor ostavi praznim. Hvala puno unaprijed! ☺

Razred:

Spol: M Ž

Pohađam: a) vjeronauk b) etiku

U sljedećih nekoliko rečenica zaokruži broj koji najbolje opisuje koliko se slažeš s određenom tvrdnjom pri čemu vrijedi:

1 – u potpunosti se ne slažem 2- djelomično se ne slažem 3 - niti se slažem niti se ne slažem 4 – djelomično se slažem 5 – u potpunosti se slažem

Migranti koji trenutno dolaze u Europu su izbjeglice i trebamo ih pustiti u Hrvatsku te im pružiti pomoć.

1 2 3 4 5

Izbjeglička je kriza samo paravan za ulazak terorista u Europu.

1 2 3 4 5

Hrvatska bi kao i Mađarska i Slovenija trebala uvesti strožu kontrolu granica.

1 2 3 4 5

Ako se trenutna situacija s migracijama nastavi, Europa će postati muslimanski kontinent.

1 2 3 4 5

Migrante iz Sirije je potrebno što prije deportirati natrag.

1 2 3 4 5

Ne osjećam nikakav strah u vezi dolaska izbjeglica u Hrvatsku.

1 2 3 4 5

Da sam ja izbjeglica, bio/bila bih zadovoljan/a postupanjem drugih država i ljudi prema meni.

1 2 3 4 5

Europska Unija čini dovoljno da zaustavi priljev migranata.

1 2 3 4 5

Moj se stav o ovoj temi podudara sa stavom barem jednog od mojih roditelja.

1 2 3 4 5

Biste li primili izbjeglicu u svoj dom?

a) Da, svakako. b) Da, ali samo žene i djecu. c) Da, ali samo kršćanske izbjeglice.

d) Da, ali samo kada bi mi to bilo plaćeno. e) Ne, nikako. f) Neki drugi odgovor.

Tko je **najveći** krivac za situaciju u Siriji?

a) SAD b) zemlje članice EU-a c) Rusija d) Islamska Država (ISIL) e) netko drugi

PRILOG 2

	Razred	Spol (M/Ž)	Predmet (V/E)	Pitanja 1									Pitanja 2	
				1	2	3	4	5	6	7	8	9	1	2
1	2at	2	2	4	3	5	4	4	2	3	2	3	e	d
2	2at	1	2	4	3	4	3	3	5	5	2	3	e	a
3	2at	1	2	3	4	5	4	5	3	3	2	5	e	d
4	2at	1	2	4	3	4	2	2	4	5	3	4	a	d
5	2at	1	1	4	3	3	4	2	2	4	3	3	a	d
6	2at	1	1	4	3	5	4	4	4	5	4	5	a	f
7	2at	1	2	4	3	4	1	3	3	5	3	1	a	d
8	2at	1	2	3	3	5	3	3	1	3	2	3	f	d
9	2at	1	2	4	3	4	2	4	4	3	2	2	0	d
10	2at	1	1	2	5	5	3	5	2	4	2	5	e	e
11	2at	1	2	3	3	4	4	3	2	2	2	4	e	a
12	2at	1	0	2	4	5	4	5	2	4	1	2	e	a
13	2at	1	1	1	5	5	5	5	5	5	1	1	e	d
14	2at	1	2	3	3	1	2	3	3	2	1	4	e	d
15	2at	1	1	1	5	5	5	5	2	2	1	5	e	d
16	2at	1	2	1	4	5	5	5	2	1	4	5	e	e
17	2at	1	1	1	5	5	5	5	1	2	1	2	e	d
18	2at	1	2	2	4	5	4	5	2	2	2	3	e	d
19	2at	2	1	1	5	5	5	3	4	3	3	2	f	0
20	2bt	1	1	4	5	5	4	5	4	5	2	4	e	d
21	2bt	1	1	4	5	5	3	2	4	3	4	4	e	d
22	2bt	1	2	1	2	5	3	5	0	0	0	0	0	0
23	2bt	1	2	1	5	5	5	5	0	0	0	0	0	0
24	2bt	1	2	1	5	5	5	5	1	0	3	5	e	d
25	2bt	1	1	1	5	5	5	5	1	3	3	5	e	0
26	2bt	1	1	1	5	5	5	5	2	1	1	5	e	d
27	2bt	1	2	3	3	4	5	3	3	4	4	3	e	d
28	2bt	1	2	5	2	2	1	2	5	2	1	1	b	d
29	2bt	1	1	3	5	4	2	4	2	3	1	2	b	a
30	2bt	1	1	3	4	5	3	2	4	5	3	3	f	d
31	2bt	1	1	3	3	3	1	4	5	5	5	1	e	d
32	2bt	1	1	3	5	5	3	3	3	1	3	4	a	d
33	2bt	1	2	3	4	4	2	3	4	5	4	4	a	d
34	2bt	1	1	2	3	2	4	4	4	3	3	4	e	d
35	2bt	1	1	3	3	4	4	4	3	3	2	4	e	a
36	2bt	1	1	1	5	5	5	5	1	0	5	5	f	o
37	2bt	1		3	3	5	4	2	2	4	4	5	e	d
38	2bt	1	1	3	5	5	3	5	4	4	2	5	f	c
39	2bt	1	1	2	3	3	2	2	5	3	2	3	e	d
40	2bt	1	1	2	4	5	3	4	4	3	2	4	c	d
41	2bt	1	1	4	2	3	2	3	5	4	3	3	b	d
42	2j	2	1	5	5	5	1	5	1	5	1	1	e	d
43	2j	2	2	3	3	5	3	3	3	5	3	3	b	d
44	2j	2	1	4	3	5	4	5	4	5	2	4	e	d
45	2j	2	2	1	1	5	1	1	5	2	1	1	e	d
46	2j	1	1	1	1	5	5	5	5	3	1	5	b	d
47	2j	1	2	2	1	5	5	5	5	4	1	5	b	d
48	2j	2	1	3	3	3	4	5	4	5	2	4	e	d
49	2j	2	1	3	4	4	3	3	2	4	4	5	e	d

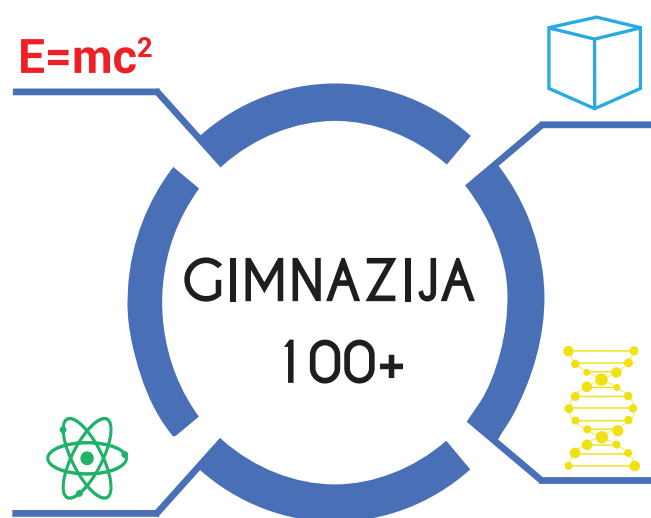
50	2j	2	1	1	1	5	1	1	5	1	1	1	e	d
51	3h	1	1	1	1	1	1	1	5	1	5	1	c	d
52	3h	1	1	1	5	5	5	5	1	5	3	5	e	b
53	3h	1	1	1	5	5	5	5	3	5	1	5	e	a
54	3h	1	1	1	4	5	1	5	3	5	1	5	e	a
55	3h	1	1	1	5	5	5	5	1	5	3	5	e	b
56	3h	1	1	1	1	5	5	5	3	4	5	2	e	a
57	3h	1	1	1	1	1	1	1	5	1	5	4	e	d
58	3h	1	1	1	1	1	1	1	1	1	1	5	e	b
59	2bt	1	1	4	5	5	1	3	2	1	1	5	e	a
60	3dt	2	1	4	3	3	3	1					a	d
61	3dt	2	1	4	3	5	3	3	3	3	3	5	f	d
62	3dt	2	1	4	5	5	5	3	2	5	3	4	c	a
63	3dt	2	2	3	4	5	2	3	4	2	3	3	d	a
64	3dt	2	1	1	5	5	5	5	1	2	1	2	e	a
65	3dt	2	1	1	5	5	5	5	2	2	1	3	e	a
66	3dt	2	1	3	3	4	4	3	3	3	2	3	e	d
67	3dt	2	1	2	5	5	5	5	1	3	2	1	f	a
68	3dt	2	1	1	5	5	5	5	3	5	1	5	e	d
69	3dt	2	1	4	1	5	3	1	5	3	3	3	e	d
70	3dt	2	2	1	5	5	5	5		1	1	5	e	d
71	3dt	2	1	3	3	4	2	2	2	2	1	5	e	d
72	3dt	2	2	3	4	5	5	3	3	4	1	4	e	d
73	3dt	2	2	3	3	5	3	5	3	4	4	4	e	a
74	3dt	2	2	5	3	5	4	4	5	4	3	3	a	a
75	3dt	2	2	3	4	5	4	3	2	3	5	3	f	a
76	3dt	2	2	5	3	4	2	1	5	3	1	2	a	a
77	3dt	1	1	4	3	5	2	3	2	5	3	3	e	d
78	3dt	1	1	2	4	5	4	5	2	5	3	2	e	a
79	3dt	1	1	3	5	4	3	5	4	3	2	4	e	a
80	3dt	1	1	3	4	5	4	3	4	2	1	5	f	d
81	3dt	1	1	1	5	5	5	5	1	1	5	5	e	d
82	3dt	1	1	2	5	5	3	4	3	2	1	4	4	d
83	3dt	1	2	1	5	5	5	5	4	5	2	5	e	d
84	3dt	1	2	2	3	5	1	4	2	4	2	5	e	a
85	3dt	1	1	1	2	5	1	4	5	5	1	3	e	a
86	3dt	1	1	2	5	5	3	4	3	2	1	4	e	a

Srednja škola Krapina

Šetalište hrvatskog narodnog
preporoda 6
49 000 Krapina
Tel: +385 49 382 111
Fax: +385 49 382 113
E-mail: ss-krapina@kr.t-com.hr
www-ss-krapina.skole.hr

Srednja škola Pregrada

Stjepana Škreblina bb
49 218 Pregrada
Tel: +385 49 382 150
Fax: +385 49 382 159
E-mail: ss-pregrada@kr.t-com.hr
web: ss-pregrada.skole.hr



www.gimnazija-100-plus.eu

Ministarstvo znanosti, obrazovanja i sporta

www.mzos.hr
esf@mzos.hr

Agencija za strukovno obrazovanje i obrazovanje odraslih

Organizacijska jedinica za upravljanje strukturnim instrumentima (DEFCO)
www.asoo.hr/defco/
defco@asoo.hr