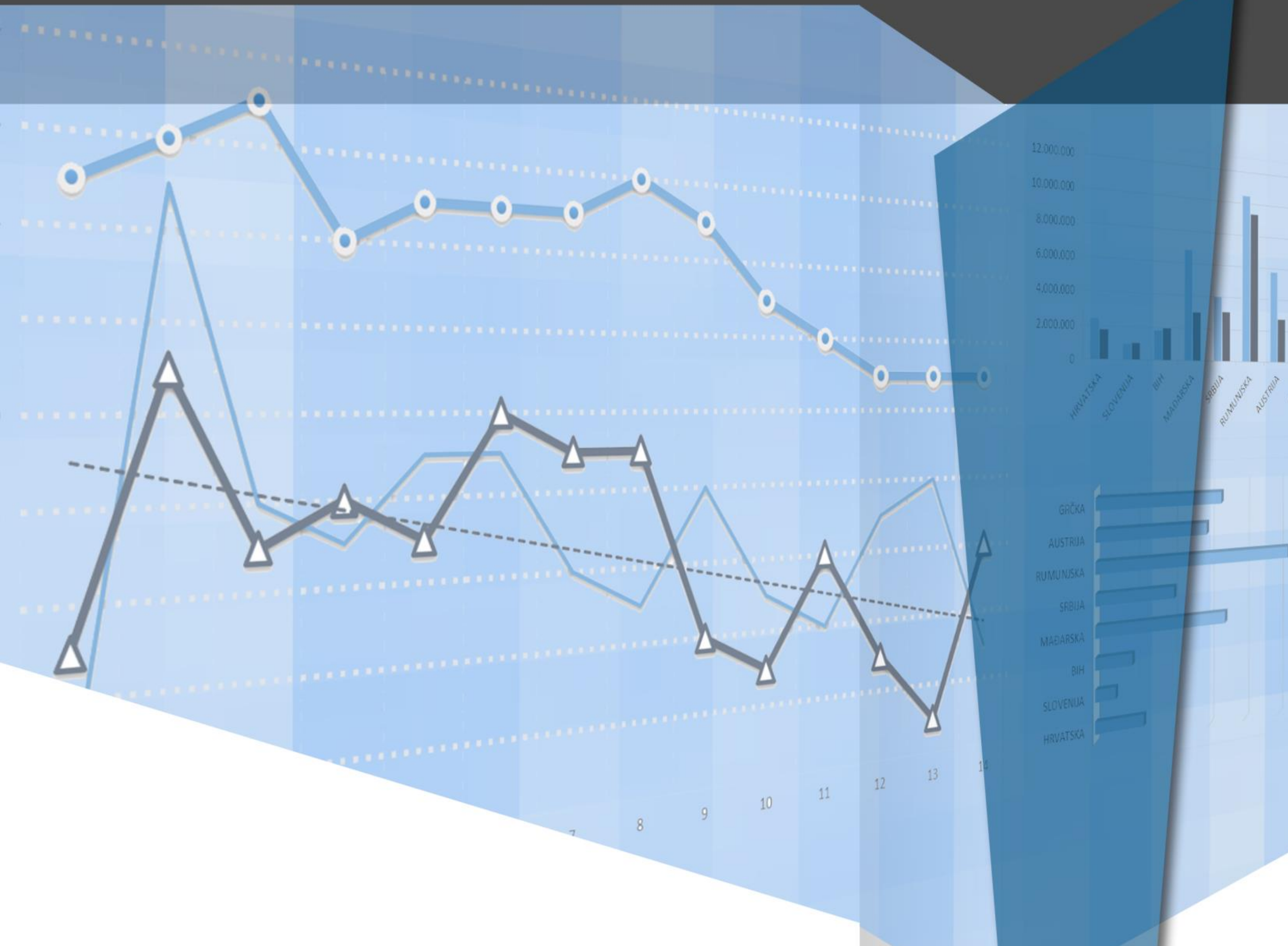
EUROPSKA UNIJA

ULAGANJE U BUDUĆNOST



STATISTIKA

izabrane teme

Josip Matejaš

Mladen Matejaš

Priručnik je izrađen u sklopu projekta „HEUREKA – spoznajom do uspjeha“ kojeg je financirala Europska unija.

Autori:
prof. dr. sc. Josip Matejaš
Mladen Matejaš dipl. ing.

„Projekt je financirala Europska unija u 100%-om iznosu iz Europskog socijalnog fonda kroz Operativni program „Razvoj ljudskih potencija 2007.-2013., poziv na dostavu projektnih prijedloga HR.3.1.20 Promocija kvalitete i unaprjeđenje sustava odgoja i obrazovanja na srednjoškolskoj razini.“

„Sadržaj ove publikacije /emitiranog materijala isključiva je odgovornost Srednje škole Ivanec“

SREDNJA ŠKOLA IVANEC – nositelj projekta
Ravnateljica: mr.sc. Lidija Kozina dipl.oec
Eugena Kumičića 7, 42 240 Ivanec
Telefon: 042 782 344
Faks: 042 781 512
E-mail: info@ss-ivanec.hr
Web: http://www.ss-ivanec.hr/

SREDNJA ŠKOLA MATE BLAŽINE LABIN – partner na projektu
Ravnatelj: Čedomir Ružić, prof.
Rudarska 4, 52 220 Labin
Telefon: 052 856 277
Faks: O52 855 329
E-mail: ssmb@ss-mblazine-labin.skole.hr
Web: http://www.ssmb.hr

Posredničko tijelo razine 1
Ministarstvo znanosti, obrazovanja i sporta
Ulica Donje Svetice 38, 10000 Zagreb
E-mail: esf@mzos.hr
Web: http://public.mzos.hr

Posredničko tijelo razine 2
Agencija za odgoj i strukovno obrazovanje i obrazovanje odraslih, Organizacijska jedinica za upravljanje strukturnim instrumentima
Radnička cesta 37b, 10000 Zagreb
E-mail: defco@asoo.hr
Web: http://www.asoo.hr/defco

Za više informacija o EU fondovima u RH:
www.mrrfeu.hr
www.strukturnifondovi.hr

Svim nastavnicima i učenicima želimo uspješan rad.

Autori

PREDGOVOR

Ovaj priručnih prvenstveno je namijenjen učenicima gimnazijskog usmjerenja koji slušaju fakultativni predmet Statistika u sklopu projekta HEUREKA spoznajom do uspjeha ali i svima ostalima koji žele ili trebaju upoznati statistiku i njenu primjenu u svakodnevnom životu.

Gradivo je prezentirano na jednostavan i slikovit način sa cjelovitim objašnjenjima utemeljenim na odgovarajućim primjerima. U priručniku se obrađuju osnovni statistički pojmovi i metode koji se uvode formalnim definicijama i objašnjavaju preko motivacijskih primjera. Svi ostali pojmovi i veličine koje proizlaze iz osnovnih detaljno su izvedeni i prezentirani pomoću elementarnih matematičkih postupaka.

Priručnik sadrži više od 200 riješenih zadataka sa svim potrebnim uputama i postupcima za rješavanje. Zadaci su praktične prirode i vrlo raznovrsni pa pokrivaju gotovo sva područja ljudskog djelovanja i istraživanja (ekonomija, fizika, tehnika, astronomija, biologija, zemljopis itd.). Na taj se način pokazuje sveobuhvatna uloga i prisutnost statistike u svakodnevnom životu, različitim strukama i znanstvenim disciplinama. U zadacima se sugeriraju mnogobrojne teme za samostalni ili grupni rad učenika u obliku domaće zadaće, prezentacije, seminarskog rada ili istraživačkog rada uz mentorstvo nastavnika. Sve navedene komponente čine priručnik prikladnim za korištenje u suvremenom projektnom tipu nastave u srednjim školama.

U nadi da će materijali iz ovog priručnika omogućiti bolji uvid u značaj i univerzalnost primjene statističkih metoda u modernoj svakodnevnici svim čitateljima, a posebno učenicima i nastavnicima koji će ga koristiti u nastavi, želimo uspješan rad.

Najljepše se zahvaljujemo prvom recenzentu prof. Marici Dolenec-Jurinić na konstruktivnim primjedbama i komentarima. Također se zahvaljujemo na savjetima i drugom recenzentu prof. Đaniju Žufić.

Autori

SADRŽAJ

OSNOVNI POJMOVI

PREDMET I METODE ISTRAŽIVANJA

Svakodnevni tijek života u prirodi i društvu uvjetuju i oblikuju različite pojave. Česte promjene i prilagodbe novim situacijama postale su nužnost u suvremenoj svakodnevnici. Da bi promjene bile uspješne i kvalitetne potrebno je pravilno analizirati postojeće stanje a što je jedna od glavnih zadaća statistike. *Statistika* je znanost koja se bavi prikupljanjem podataka za promatranu pojavu, koristeći definirane načine prikupljanja, te njihovim uređivanjem, prikazivanjem i analizom u svrhu donošenja određenih zaključaka i njihovom primjenom u sličnim i novim situacijama. Drugim riječima primjenom statističkih metoda sažimaju se obavijesti koje nam podaci pružaju. Na taj način uočavamo slabosti i nedostatke postojećeg stanja te njihove uzroke i posljedice. Nakon provedene analize postojećeg stanja, korelacijska i regresijska analiza omogućava nam praćenje i predviđanje (simulaciju) prostornog i vremenskog trenda promatranih pojava uz nove stvarne ili hipotetske uvjete. Sve to doprinosi boljem razumijevanju strukture, principa i zakonitosti promatranih pojava. Time stječemo mogućnost svjesnog mijenjanja prirode i društva u željenom smjeru radi unapređenja i poboljšanja kvalitete života pojedinca i društva u cjelini a što statistici daje puni smisao i značaj.

Razlikujemo deskriptivne i inferencijalne statističke metode. *Deskriptivne metode* se sastoje od uređivanja (grupiranja) i prikazivanja prikupljenih podataka (tablično ili grafički) te izračunavanja njihovih numeričkih obilježja. Dobiveni zaključci se odnose samo na te podatke bez mogućnosti poopćavanja.

Inferencijalne metode polaze od uzorka podataka (podskupa) koji se analiziraju a zaključci se poopćavaju na cijeli skup (cjelinu) sa određenom vjerojatnošću.

Statistički skup je skup elemenata (jedinica) nad kojima se analiza provodi. On može biti konačan ili beskonačan, stvaran ili hipotetski. Elementi skupa imaju neka svojstva (obilježja ili varijable) koje se mogu statistički analizirati. Skup podataka za neko obilježje elemenata statističkog skupa čini populaciju ili osnovni skup namijenjen statističkoj analizi. Podaci se prikupljaju metodama koje osiguravaju njihovu što veću točnost i vjerodostojnost.

Primjer 1. Skup učenika nekog razreda može biti statistički skup. Učenici su elementi (jedinice) skupa. Možemo analizirati neka njihova obilježja (varijable): visinu, težinu, ocjene, dob, spol, navike, udaljenost od kuće do škole, itd. Za svako od tih obilježja formira se skup podataka (populacija) koji onda možemo grupirati, prikazivati i numerički analizirati.

Obilježja (varijable) koja služe kao predmet analize mogu biti *kvalitativne* (kategorijalne) i *kvantitativne* (numeričke). Tako su na primjer boja, oblik, izgled, položaj, agregatno stanje i sl. kvalitativna a duljina, težina, volumen, cijena, temperatura, brzina i sl. kvantitativna obilježja. Ova podjela nije isključiva jer se neka obilježja mogu tretirati kao kvalitativna i kao kvantitativna (npr. ocjena izražava kvalitetu ali i kvantitetu nečijeg znanja ili aktivnosti). Pojedino obilježje (varijabla) se pojavljuje u različitim oblicima (modalitetima, stupnjevima ili vrijednostima). Tako na primjer boja može biti plava, zelena, crvena (oblici obilježja), temperatura može biti 0^oC, 25^oC, 100^oC (vrijednosti varijable). Vidimo da se pojmovi obilježja i oblici prirodno češće vežu uz kvalitativna a varijable i vrijednosti uz kvantitativna svojstva promatrane pojave. Kvalitativna obilježja mogu biti nominalna (atributivna, geografska) te uređajna (redna, obilježja ranga). Kvantitativna obilježja mogu biti diskretna (skokovita) ili kontinuirana (neprekidna). Općenito diskretna varijabla može poprimati svoje susjedne vrijednosti u određenim razmacima. Unutar tih razmaka vrijednost ne postoji. Tako na primjer broj učenika u razredu može biti samo cijeli broj, recimo 23 ili 24 ali ne može biti između 23 i 24. Slično tome u novčaniku možemo imati 15.15 kuna ili 15.16 kuna ali ne možemo imati neki iznos između njih jer je 1 lipa najmanja novčana jedinica. Kontinuirana varijabla može poprimati neprekinute vrijednosti (bez razmaka) pa nema susjednih. Tako na primjer temperatura može biti $20^o C$ ili $21^o C$ ali može imati i bilo koju vrijednost između njih, u boci možemo imati 3 ili 4 dl soka ali i bilo koju količinu između njih.

Podaci o vrijednostima obilježja, s obzirom na izvor, mogu biti *primarni* i *sekundarni*. Primarni podaci su rezultat primarnih istraživanja provedenih u skladu sa postavljenim ciljevima a prikupljaju se anketama, mjerenjima, promatranjima, eksperimentima i sl. Sekundarni podaci su rezultat istraživanja drugih istraživača. To su podaci čija vrsta i opseg ne ovise o ciljevima i potrebama danog istraživanja a dostupni su iz različitih baza podataka ili se mogu dobiti na zahtjev (besplatno ili kupnjom) od institucija (banke, zavodi, ustanove i sl.) ili pojedinaca. Svi prikupljeni podaci se pripremaju (uređuju) za statističku obradu.

ZADACI

1. Navedite neka obilježja sljedećih statističkih skupova: skup kuća (zgrada) u ulici, skup automobila na parkiralištu, skup putnika u zrakoplovu, skup mobitela u vašem razredu, skup država članica EU.
2. Koje od sljedećih obilježja (varijabli) su diskretne a koje kontinuirane: datum, visina, težina, broj stolica, planirani broj sati nastave, održani broj sati nastave, površina, cijena električne energije, potrošnja električne energije, broj boca na polici, količina tekućine u tim bocama, broj kamenčića na plaži, količina pijeska na plaži, dubina, slanost vode, broj zvijezda na nebeskom svodu, vrijeme ustajanja?
3. Koju varijablu (diskretnu ili kontinuiranu) čine podaci očitani sa digitalnih odnosno analognih mjernih uređaja (sat, termometar, brzinomjer, itd.)?
4. Ako za analizu kretanja temperature u vašem mjestu koristite: (a) podatke meteorološke službe objavljene u medijima, (b) podatke koje ste dobili vlastitim mjerenjem, o kojim vrstama podataka se radi? Koji su vjerodostojniji?
5. Ako na temelju dobro riješenog ispita znanja u jednom trećem razredu pohvalimo: (a) samo učenike tog razreda, (b) sve učenike trećih razreda, na temelju koje statističke metode smo tako postupili?

RJEŠENJA

1. Skup kuća: kućni broj, stambena površina, broj stanara, vrijednost, starost, dimenzije, itd. Skup automobila: boja, marka, starost, broj vrata, snaga, brzina, potrošnja goriva, datum registracije, itd. Skup putnika: ime, prezime, dob, zaposlenje, obrazovanje, svrha puta, destinacija, itd. Skup mobitela: cijena, kapacitet, brzina, oblik, boja, itd. Skup država: površina, broj stanovnika, jezik, geografski položaj, glavni grad, klima, itd.
2. Diskretne (d) a kontinuirane (k): datum (d), visina (k), težina (k), broj stolica (d), planirani broj sati nastave (d), održani broj sati nastave (k), površina (k), cijena električne energije (d), potrošnja električne energije (k), broj boca na polici (d), količina tekućine u tim bocama (k), broj kamenčića na plaži (d), količina pijeska na plaži (k), dubina (k), slanost vode (k), broj zvijezda na nebeskom svodu (d), vrijeme ustajanja (k).
3. Diskretne sa digitalnih a kontinuirane sa analognih.
4. (a) Sekundarni, (b) primarni. U ovom slučaju primarni su vjerodostojniji.
5. Na temelju: (a) deskriptivne, (b) inferencijalne metode (metode uzorka).

ETAPE U PROVOĐENJU ISTRAŽIVANJA

Statističke analize pojava u prirodi i društvu na temelju prikupljenih podataka provode se kroz sljedeće etape.

1. Određuje se predmet i cilj istraživanja, izbor i način prikupljanja podataka te se definiraju statistički skupovi i obilježja koja će se analizirati.
2. Prikupljaju se podaci pri čemu treba osigurati njihovu vjerodostojnost i točnost.
3. Vrši se obrada (uređivanje) i prikaz podataka (grupiranje te tabelarni i/ili grafički prikaz).
4. Primjenjuju se numeričke metode za analizu tako uređenih podataka (deskriptivne ili inferencijalne ako podaci čine uzorak pojave), Izbor metode ovisi o postavljenim ciljevima analize, vrsti statističkog skupa (niza) i obilježja koje se analizira.
5. Tumače se dobiveni rezultati iz tabela, grafičkih prikaza i numeričkih pokazatelja te se donose odgovarajući zaključci u skladu sa tim rezultatima.
6. Primjenjuju se dobiveni zaključci u sličnim i novim situacijama u teoriji i praksi u svrhu organiziranja, provođenja i predviđanja postupaka, aktivnosti i ishoda.

Navedene etape mogu se okvirno sažeti u tri faze: *prikupljanje podataka, obrada i analiza podataka* te *tumačenje dobivenih rezultata*. Svrha statističke analize u praksi nije samo utvrđivanje i opisivanje činjeničnog stanja za promatranu pojavu već i pronalaženje mjera i postupaka za njeno unapređivanje i poboljšanje. Upravo se na temelju statističke analize utvrđuju razlozi trenutnog stanja iz kojih onda proizlaze načini za promjenu tog stanja ako za to postoji želja ili potreba. Zbog toga je statistička analiza nezaobilazni dio planiranja, organiziranja, poslovnog odlučivanja, investiranja i sl.

ZADACI

1. Definirajte anketni upitnik i provedite anketu u razredu na temelju koje ćete simulirati etape u provođenju statističke analize.
2. Pretpostavite da trebate organizirati proslavu (rođendana, godišnjice, nekog važnog događaja i sl.). Navedite i razradite pojedine etape od ideje do realizacije.
3. Ako niste zadovoljni sa svojim uspjehom iz nekog predmeta, utvrdite razloge za takvo stanje te pronađite načine za poboljšanje.
4. Osmislite metodu kojom bi što točnije odredili broj stabala u nekom većem šumskom kompleksu.
5. Kako bi što točnije odredili broj riba u jezeru?
6. Izaberite proizvoljno pojavu (događaj), zadajte cilj analize te definirajte etape za realizaciju postavljenog cilja.

RJEŠENJA

1. Prijedlozi pitanja:
   Koliko prosječno vremena dnevno provodite za računalom, televizorom i mobitelom zajedno?
   Koliko prosječno vremena dnevno provodite u učenju?
   Koju ocjenu iz matematike ste imali prošle godine?
   Koju ocjenu očekujete iz ovog predmeta?
   Koliko ste imali ljubavnih veza koje su trajale duže od mjesec dana?
   U kojem zanimanju (poslu) vidite sebe za 10 godina?
   Itd.
   U provođenju pojedinih etapa dati inicijativu učenicima. U svakoj etapi navesti različite mogućnosti pristupa (ne treba provoditi etape u cijelosti jer to na ovoj početnoj razini znanja nije ni moguće). Cilj je ukazati učenicima na teme koje će se obrađivati kako bi stekli dojam o sadržaju predmeta.
2. Sakupite prijedloge diskusijom. Uzmite u razmatranje mjesto održavanje proslave (vlastiti dom, priroda, učionica, restoran, itd.) zatim broj uzvanika, pripremu hrane i pića (samostalna priprema, kupnja ili kombinirano), protokol i sadržaj same proslave (govori, glazba, zaduženja). Nakon prikupljanja podataka usporedite pojedine opcije s obzirom na cilj same proslave i realne mogućnosti (troškove, potrebni prostor, vrijeme, angažman, pomoć drugih osoba itd.) te odaberite najpovoljniju.
3. Analizirajte postojeću situaciju (vrijeme provedeno u učenju, predznanje, pažnja na satu, pisanje zadaća, ponavljanje, redovito ili povremeno učenje i sl.). Na temelju ovih pokazatelja uočit ćete razloge slabijeg uspjeha kao i načine poboljšanja (trebate više vremena posvetiti učenju ili potražiti pomoć ili više ponavljati i vježbati itd.).
4. Navedite vlastite ideje. Jedan od načina je da se na nekoliko (npr. 10) različitih mjesta u šumi prebroje stabla na određenoj površini (npr. 20 stabala na 35 m^2, 15 stabala na 50 m^2, itd.), izračuna prosječna gustoća (broj stabala na zadanu jedinicu površine) te pomnoži sa površinom šume.
5. Navedite vlastite ideje. Jedan od načina je da se ulovi i markira (označi) određeni broj riba (M) koje se puste natrag na različitim mjestima u jezero. Nakon nekog vremena na više mjesta ponovimo ulov. Od ukupnog broja ulovljenih riba (u) određeni broj je markiram (m). Ukupan broj riba u jezeru (x) dobijemo iz razmjera $x : M = u : m$.
6. Razradite vlastite ideje.

UREĐIVANJE I PRIKAZIVANJE PODATAKA

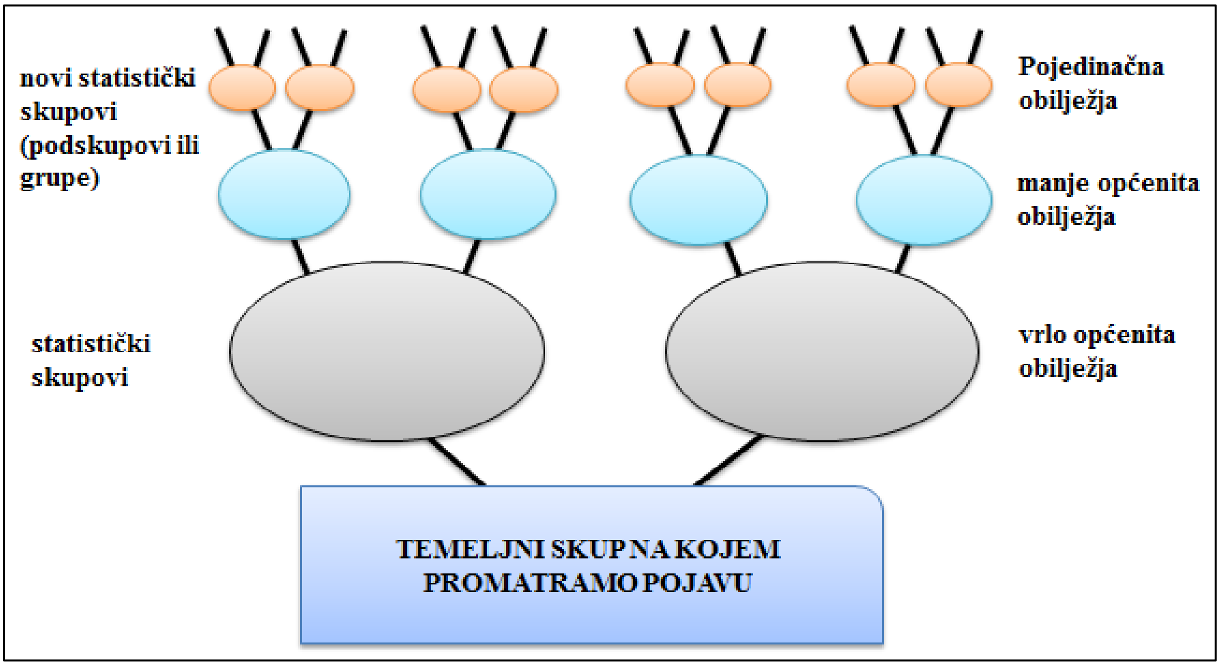
Nakon što se prikupe podaci u skladu sa definiranim predmetom i ciljem istraživanja prelazi se na njihovu obradu. Prvi korak u obradi podataka je uređivanje i prikaz podataka. U tu svrhu se najviše koriste metode grupiranje podataka prema vrstama i oblicima obilježja i ciljevima analize te tabelarni i grafički prikaz. Navedeni prikazi omogućuju početnu (vizualnu ali i kvantitativnu) analizu podataka koja je u mnogim slučajevima i dovoljna.

STATISTIČKI SKUPOVI I NIZOVI

Ako želimo analizirati neku pojavu prvo treba izdvojiti jedinice (elemente) koji reprezentiraju promatranu pojavu i time formirati statistički skup. Svi elementi statističkog skupa međusobno su jednaki po općim svojstvima ali se mogu razlikovati po pojedinačnim svojstvima (obilježjima). Iz tako dobivenog skupa možemo formirati nove skupove (podskupove ili grupe) po određenim pravilima (zahtjevima) ovisno o obilježju koje želimo analizirati. Grupe mogu biti međusobno disjunktne što je i najčešći slučaj a mogu sadržavati sve elemente polaznog skupa (deskriptivni pristup) ili samo neke – uzorak (inferencijalni pristup. Iz tih grupa možemo nastaviti dalje izdvajanje. Na taj način statističke skupove možemo formirati na različitim razinama ovisno o potrebama i ciljevima analize (slika 1). Osnovna metoda koja se pri tome koristi je grupiranje podataka prema definiranim pravilima te vrstama i oblicima obilježja. Dobiveni skupovi mogu se po potrebi urediti po nekom pravilu (redoslijedu) u *statistički niz* (rastući, padajući, abecedni, kronološki itd.). Dakle, razlika između skupa i niza je u tome što u skupu nije bitan redoslijed podataka a u nizu jest. Tako je na primjer skup $\{1,2,x,y\}$ jednak skupu $\{1,2,y,x\}$ i također skupu $\{y,2,x,1\}$ itd., dok je niz $1,2,x,y$ različit od niza $1,2,y,x$ te su ova oba niza različita od $2,x,1,y$ itd. Prema tome, niz je skup na kome je definiran uređaj među elementima.

Postupcima izdvajanja i grupiranja velik broj pojedinačnih podataka sažima se u manji broj grupa pri čemu su svi podaci u istoj grupi jednakog ili sličnog obilježja.

Primjer 2. Ako želimo analizirati strukturu stanovništva neke županije tada temeljni skup čine svi stanovnici koji žive (prijavljeni su) na teritoriju te županije. To je ujedno statistički skup iz kojeg izdvajamo grupe sa istim obilježjima (npr. predškolska djeca, školarci – osnovna i srednja škola, studenti, zaposleni, nezaposleni, umirovljenici itd.). Ako nadalje želimo analizirati strukturu skupa zaposlenih osoba možemo izdvojiti zaposlene u javnom i privatnom sektoru, iz tih skupova možemo izdvojiti zaposlene u proizvodnim i uslužnim djelatnostima koje dalje možemo dijeliti na zaposlene u poljoprivredi i industriji zatim trgovini, ugostiteljstvu, obrazovanju itd.



Slika 1. Razine formiranja statističkih skupova

ZADACI

1. Jedna osnovna škola koja ima 16 razrednih odjeljenja (I-VIII razred u A i B smjeni) organizira posjet muzeju za 4 odjeljenja. Definirajte nekoliko pravila po kojima ćete izdvojiti ta 4 odjeljenja.
2. Definirajte nekoliko pravila po kojima ćete u vašem razredu formirati dvije ekipe za sportsko natjecanje. Uredite svaku tako formiranu ekipu u niz.
3. Navedite nekoliko razina za formiranje statističkih skupova ako je polazni skup javni putnički prijevoz. Koja obilježja ovdje možemo analizirati?
4. Iz skupa riječi {pada, ide, trava, sat, vrijeme, stoji, kiša, raste} formirajte jedan ili više nizova.
5. Koliko se različitih skupova a koliko nizova može formirati od ova četiri elementa: $4,\ 0,\ -1,\ 2/3$? Među njima izdvojite rastući i padajući niz.

6. U nekoj trgovini prodaju se mliječne čokolade, čokolade sa lješnjacima i rižom (oznake M, L i R) od 100, 200 i 300 grama (oznake 1,2 i 3). Uvidom u ispis blagajne tokom jednog dana prodaja je tekla ovako (kronološki statistički niz):
M3, L1, L2, M1, M1, R3, L2, R1, R2, R1, M2, L3, M3, L1, R1. R1, R1, M2, L1, L3, R1, M3, R2, L2, L1, R3, R1, M1, M2, M3, R2, M2, R1, L1, R1, L1, M3, R1, L2, L3, M1, M3, R2, R2, L3, M3, M1, R2, M2, M1, R2, R2, L3, R3, M1, R3, M1, R1, R1, L1.
Pregledno prikažite navedeni ukupni promet tako da su jasno vidljivi pojedini podskupovi (iste vrste i jednake količine).

RJEŠENJA

1. Najmlađi i najstariji (oba prva i oba osma razreda), završni razredi (oba sedma i oba osma razreda), početni razredi (oba prva i oba druga razreda), niži razredi iste smjene (I-IV smjene A ili smjene B), viši razredi iste smjene (V-VIII smjene A ili B) itd.
2. Ekipe mogu obuhvatiti sve učenike u razredu ili samo neke (uzorak). Muška i ženska ekipa, ekipe formirane prema mjestu stanovanja, prema uspjehu, prema abecedi, prema željama i dogovoru, prema zahtjevu nastavnika itd. Grupe možemo formirati u niz prema abecedi, visini, snazi, ulogama u ekipi, dogovoru itd.
3. Prva razina: zračni, vodeni i kopneni prijevoz. Za npr. kopneni druga razina je željeznički i cestovni prijevoz. Za cestovni treća razina je autobusni, tramvajski i taxi prijevoz. Možemo nastaviti i dalje: četvrta razina noćni i dnevni prijevoz, peta redovite i izvanredne linije itd. Obilježja: kapaciteti, broj prevezenih putnika, broj linija, vremenski raspored, redovitost (kašnjenje), cijena itd.
4. Možemo formirati jedan niz npr. po abecedi, po broju slova u riječi i sl. ili ih grupirati u dva niza npr. niz glagola i niz imenica po abecedi. Možemo također grupirati imenice i glagole u jedan niz npr. kiša pada, trava raste, vrijeme ide, sat stoji.
5. Samo jedan skup a $4 \cdot 3 \cdot 2 \cdot 1 = 24$ niza pri čemu je $-1, 0, 2/3, 4$ rastući (uzlazni), dok je $4, 2/3, 0, -1$ padajući (silazni) niz.
6. Saslušati ideje učenika. Jedan od načina je tabelarni prikaz, npr.

Čokolade	**100 g**	**200 g**	**300 g**	**Ukupno**
mliječna	8	5	7	20
sa lješnjacima	7	4	5	16
sa rižom	12	8	4	24
Ukupno	27	17	16	60

ODNOS DVAJU PODATAKA

Promatrajući svijet oko sebe uočavamo stalne promjene. Točnije rečeno stvari, pojave i događaje spoznajemo upravo po njihovim promjenama. Promatramo li kako se neka pojava mijenja tokom vremena ili je uspoređujemo sa drugom pojavom dolazimo do dubljih spoznaja o njezinoj prirodi i svojstvima. Kao što smo već spomenuli neke promjene su kvalitativne naravi dok ih je većina kvantitativne naravi. Kvantitativna svojstva i njihove promjene možemo mjeriti kvantitativnim mjernim jedinicama i izražavati realnim brojevima. To nam omogućuje njihovu usporedbu, analizu i izvođenje određenih zaključaka. Za bilo kakvu usporedbu ili analizu ključno je imati bar dva različita stanja (dvije različite vrijednosti) pojave koju promatramo. Na primjer kad bi vrijeme stalno bilo vedro a temperatura stalno ista ne bi postojala potreba za analizom vremena i davanjem vremenskih prognoza. Pojava bi bila stalno ista (konstantna) i ne bi imali što analizirati i uspoređivati. Uz postojanje bar dviju različitih vrijednosti analiza može početi.

Temelj svake analize je međusobna usporedba dvaju podataka. Ako se radi o kvalitativnom obilježju usporedba (ukoliko je provediva) uglavnom proizlazi iz praktičnih razloga i subjektivne je naravi (crveno ili plavo, desno ili lijevo, plastično ili metalno, itd.). Važnija je i najčešća usporedba obilježja sa pridruženim numeričkim vrijednostima a koja je objektivne naravi, Odnos dvaju istovrsnih numeričkih podataka x i y najčešće se izražava *apsolutnim iznosom* (razlikom $x-y$ ili $y-x$ koja pokazuje za koliko je jedan podatak veći ili manji od drugog) ili *relativnim iznosom* (kvocijentom x/y ili y/x koji pokazuje koliko puta je jedan podatak veći ili manji od drugog). Oba načina izražavanja imaju svoje prednosti ovisno o situaciji koju promatramo.

Primjer 3. Cijena proizvoda A je 2 kune a proizvoda B 200 kuna. Ako svaki od njih poskupi 1 kunu, koji je poskupio više?

Ako želimo kupiti proizvod A ili B po novoj cijeni platit ćemo 1 kunu više nego prije poskupljenja. Sa tog stajališta apsolutno poskupljenje za A i B je jednako. Sa relativnog stajališta, u odnosu na postojeću cijenu, očito nije. Naime, relativno poskupljenje je

$$\mathrm{A}: \ \frac{1}{2}=0.5=50\%, \qquad \mathrm{B}: \ \frac{1}{200}=0.005=0.5\%.$$

Vidimo da je poskupljenje proizvoda A sa 2 na 3 kune vrlo veliko (50%) dok je poskupljenje proizvoda B sa 200 na 201 kunu samo 0.5% (100 puta manje nego za A) a što vjerojatno nećemo ni primijetiti. Relativni iznos je u ovom primjeru bolji pokazatelj.

Primjer 4. Od ukupnog broja stanovnika je 60% ženskih i 40% muških osoba. Koliko ženskih osoba ima više u odnosu na muške?

Na temelju zadanih relativnih iznosa ne možemo odgovoriti na ovo pitanje. Potrebno je poznavati i apsolutne iznose. Naime broj ženskih osoba može biti 3 od 5 ili 60 od 100 ili 12000 od 20000 itd.

U skladu sa ovim razmatranjima imamo sljedeću definiciju.

DEFINICIJA 1. NEKA JE x ZADANI (REFERENTNI) PODATAK IZRAŽEN KAO REALNI BROJ. APSOLUTNA I RELATIVNA UDALJENOST ILI ODSTUPANJE NEKOG DRUGOG PODATKA y OD PODATKA x JE

$$\Delta(x,y)=y-x \ \text{(apsolutna)}, \qquad \delta(x,y)=\frac{\Delta(x,y)}{x}=\frac{y-x}{x}=\frac{y}{x}-1 \ \text{(relativna)}.$$

PRI TOME RELATIVNA UDALJENOST NIJE DEFINIRANA ZA $x=0$.

Vidimo da je apsolutna udaljenost izražena u istim mjernim jedinicama kao i promatrani podaci dok je relativna neimenovani broj pa se obično izražava u postocima. Relativna udaljenost pokazuje udaljenost po jedinici promatranog podatka (veličine). Ovo značenje dolazi do izražaja kad promatramo promjenu neke veličine. Primijetimo da su navedene udaljenosti funkcije dviju varijabli, $\Delta(x,y)$ i $\delta(x,y)$. Ako je iz konteksta jasno na koje podatke se udaljenosti odnose varijable se mogu i izostaviti (pišemo samo Δ i δ).

Apsolutna i relativna udaljenost može biti pozitivna, negativna ili nula. Apsolutna udaljenost je pozitivna ako se usporedba vrši sa većom vrijednošću $(y > x)$, negativna za manju vrijednost $(y < x)$ a nula za jednaku $(y = x)$. Slično za relativnu udaljenost ako je referentna vrijednost pozitivna $(x > 0)$ a u protivnom obrnuto. To ne treba miješati sa pojmom apsolutne vrijednosti realnog broja koja nikad nije negativna ($|-3| = 3,\ |4| = 4,\ |0| = 0$). Ako nas ne zanima smjer (predznak) udaljenosti tada je dovoljno promatrati samo njenu veličinu, apsolutnu vrijednost apsolutne udaljenosti $|\Delta|$ ili apsolutnu vrijednost relativne udaljenosti $|\delta|$.

Iz definicije 1 slijedi

$$\Delta(x, y) = y - x = -(x - y) = -\Delta(y, x),$$
$$\delta(x, y) = \frac{y - x}{x} = -\frac{x - y}{y} \cdot \frac{y}{x} = -\frac{y}{x} \cdot \delta(y, x),$$

što znači da je apsolutna udaljenost od y do x iste veličine ali suprotnog smjera (predznaka) kao apsolutna udaljenost od x do y, dok odnos relativnih udaljenosti ovisi o omjeru promatranih podataka y / x. Ako je neka od ovih udaljenosti poznata tada iz definicije 1 dobivamo izraze za referentni podatak x kao i za podatak y sa kojim ga uspoređujemo,

$$y = x + \Delta, \qquad x = y - \Delta, \tag{1}$$

$$y = (1 + \delta)x, \qquad x = \frac{y}{1 + \delta}. \tag{2}$$

U praksi je vrlo često poznata samo gornja granica (ograda) A ili R za ove udaljenosti ($|\Delta| \le A$ ili $|\delta| \le R$). U tom slučaju imamo

$$|\Delta| \le A \Rightarrow -A \le \Delta \le A \Rightarrow -A \le y - x \le A \Rightarrow \begin{cases} x - A \le y \le x + A \\ y - A \le x \le y + A. \end{cases} \tag{3}$$

Kod relativne udaljenosti vrijedi

$$|\delta| \le R \Rightarrow -R \le \delta \le R \Rightarrow -R \le \frac{y}{x} - 1 \le R \Rightarrow \begin{cases} (1 - R)x \le y \le (1 + R)x \\ \dfrac{y}{1 + R} \le x \le \dfrac{y}{1 - R} \end{cases} \tag{4}$$

za $x > 0$. Ako je $x < 0$ u izrazima za x i y sve znakove $\le$ treba zamijeniti sa $\ge$.

Navedene udaljenosti imaju vrlo široku primjenu u praksi. Koristimo ih kod usporedbe podataka, kod praćenja promjena neke veličine, npr. kod postotnog, promilnog, kamatnog računa i sl. (apsolutna i relativna promjena ili prirast) te kod aproksimacije točnih podataka približnim, npr. zaokruživanje brojeva, približno računanje u konačnoj aritmetici računala i sl. (apsolutna i relativna pogreška).

ZADACI

1. Napišite detaljan izvod formula (1), (2), (3) i (4) sa svim koracima.
2. Odredite apsolutnu i relativnu udaljenost broja 5 od broja 50 te zatim broja 50 od broja 5. Objasnite dobivene rezultate.
3. Zaokružite razlomak $16/33$ prvo na četiri a zatim na jednu decimalu te odredite apsolutne i relativne greške zaokruživanja.
4. Ako smo kod prepisivanja broja 4.222 zabunom izostavili jednu dvojku, koliku smo relativnu pogrešku napravili? Isto pitanje za broj 4222.
5. Koji je račun točniji: $10.0002^2 \approx 100$ ili $\sqrt{3} \approx 1.732$?
6. Ako neki broj različit od nule aproksimiramo nulom, kolika je relativna greška aproksimacije?
7. Ako je apsolutna greška -0.05 a relativna -0.0002 kolika je točna vrijednost?
8. Sljedeći brojevi dobiveni su zaokruživanjem na posljednju zadržanu znamenku: 4.805, 1.7, 23. Što možemo reći o točnim vrijednostima?
9. U kojem se intervalu nalazi približna vrijednost ako je točna vrijednost 4950 a relativna greška nije veća od 10%? Isto pitanje za točnu vrijednost ako je 4950 približna vrijednost.
10. Procijenili smo da će nogometnom derbiju prisustvovati približno 24000 navijača. Ako greška naše procjene nije veća od 20%, u kojim je granicama točan broj navijača?
11. Nakon poskupljenja od 5% cijena nekog proizvoda je 252 kune. Kolika je cijena bila prije poskupljenja?
12. Ako cijenu povećamo za 7% a zatim novu cijenu nakon nekog vremena smanjimo za 7%, što se u stvari dogodilo sa polaznom cijenom?
13. Povećamo li početni iznos za neki postotak (δ), za koji postotak treba smanjiti novonastali iznos da se dobije početni? Interpretirajte dobiveni rezultat na primjeru PDV-a.
14. Na nabavnu cijenu neke robe odobrava se 5% popusta, zatim se obračunava 8% marže i 25% PDV-a. Ako je prodajna cijena te robe 256 kuna i 50 lipa, kolika je nabavna cijena?
15. Neko trgovačko poduzeće ima 4 prodajne jedinice čiji je udio u ukupnim prihodima redom 20%, 25%, 15% i 40%. Ako se prihodi prve jedinice povećaju za 3%, druge smanje za 4% a treće povećaju za 2%, kako treba poslovati četvrta prodajna jedinica pa da se ukupni prihodi poduzeća povećaju za 2%?
16. Srednja udaljenost Mjeseca od Zemlje iznosi 384385 km a najveće odstupanje od te udaljenosti je $\pm 5.494231\%$. Kolika je najveća (apogej) i najmanja (perigej) udaljenost Mjeseca od Zemlje te kolika je razlika tih dviju udaljenosti?
17. Ako 15000 EUR oročimo uz 4% godišnjih kamata koliko ćemo imati za 10 godina?
18. Za koje vrijeme se početna glavnica uz 4% godišnjih složenih kamata udvostruči?
19. Ako danas na štednji imamo 20000 EUR, koliko smo uložili prije 10 godina ako je godišnja kamatna stopa bila 6% prve dvije godine, 5% sljedeće tri i 4% zadnjih pet?
20. Za gradnju temeljne ploče nebodera oblika kvadra prema planiranim dimenzijama trebalo bi 5000 m^3 betona. Zbog naknadno utvrđenih loših svojstava podloge odlučeno je povećati dubinu ploče za 10%, širinu za 12% te smanjiti dužinu za 5%. Koliko betona treba za ploču sa ovako promijenjenim dimenzijama?

RJEŠENJA

1. Izvodi su elementarni. Prilikom prebacivanja članova preko znaka (ne)jednakosti treba paziti na promjenu predznaka a prilikom množenja sa nazivnikom x (formula (4))

treba razlikovati slučajeve $x<0$ kada znakovi nejednakosti mijenjaju smjer odnosno $x>0$ kada ostaju isti.

2. Broj 5 je udaljen od broja 50 za $\Delta=-45$ ili za $\delta=-0.9=-90\%$ a broj 50 od broja 5 za $\Delta=45$ ili za $\delta=9=900\%$. Naime, ako referentni broj 50 smanjimo na 5 smanjili smo ga za 45 ili za 90% (negativni predznak promjene). Slično, ako referentni broj 5 povećamo na 50 povećali smo ga za 45 ili za 900% (pozitivni predznak promjene).
3. $x=16/33=0.484848...\approx 0.4848=y \Rightarrow \Delta=-4.848...\cdot 10^{-5}$, $\delta=-10^{-4}=-0.01\%$;
   $x=16/33=0.484848...\approx 0.5=y \Rightarrow \Delta=0.0151515...$, $\delta=0.03125=3.125\%$.
4. Za $x=4.222,\ y=4.22$ imamo $\delta=-0.0004737...\approx -0.047\%$ a za $x=4222,\ y=422$ je $\delta=-0.90004737...\approx -90\%$.
5. Usporedimo apsolutne vrijednosti relativnih pogrešaka računanja. Za prvi račun je $|\delta|=3.99988...\cdot 10^{-5}$ a za drugi $|\delta|=2.93337...\cdot 10^{-5}$ pa je drugi točniji.
6. $|\delta|=\left|\dfrac{0-x}{x}\right|=1=100\%$.
7. $\delta=\dfrac{\Delta}{x} \Rightarrow x=\dfrac{\Delta}{\delta}=\dfrac{-0.05}{-0.0002}=250$.
8. Kod zaokruživanja (prema najbližem broju) apsolutna vrijednost apsolutne pogreške ne smije prelaziti polovicu vrijednosti posljednje zadržane znamenke u zaokruženom broju. Kako je u zaokruženom broju $y=4.805$ posljednja zadržana znamenka tisućinka imamo $|\Delta|\le 0.5\cdot 10^{-3}=0.0005=A$, pa iz formule (3) slijedi $4.8045\le x\le 4.8055$. Slično za $y=1.7$ imamo $A=0.5\cdot 10^{-1}=0.05$ pa je $1.65\le x\le 1.75$. Za $y=23$ je $A=0.5\cdot 10^{0}=0.5$ pa je $22.5\le x\le 23.5$.
9. Koristimo formulu (4). Prvo je $x=4950$ i $R=10\%=0.1$ pa je $4455\le y\le 5445$ a zatim je $y=4950$ i $R=10\%=0.1$ pa je $4500\le x\le 5500$.
10. Koristimo formulu (4): $y=24000,\ R=20\%=0.2 \Rightarrow 20000\le x\le 30000$.
11. Za postotnu promjenu neke veličine koristimo formulu (2). Imamo $y=252$, $\delta=5\%=0.05$ pa je $x=y/(1+\delta)=240$ kuna.
12. $C(1+0.07)(1-0.07)=C(1+\delta) \Rightarrow \delta=-0.0049=-0.49\%$. Nakon ove dvije promjene polazna cijena je smanjena za 0.49%.
13. $x(1+\delta)(1-\varepsilon)=x \Rightarrow 1-\varepsilon=\dfrac{1}{1+\delta} \Rightarrow \varepsilon=\dfrac{\delta}{1+\delta}$ (umjesto $1-\varepsilon$ možemo staviti $1+\varepsilon$ pa dobivamo $\varepsilon=-\delta/(1+\delta)$ a što znači smanjenje). Za $\delta=0.25$ je $\varepsilon=0.2$ (PDV je 25% nabavne cijene a 20% prodajne cijene).
14. $NC\cdot 0.95\cdot 1.08\cdot 1.25=PC \Rightarrow NC\cdot 1.2825=256.50 \Rightarrow NC=200$ kuna.
15. $20\cdot 1.03+25\cdot 0.96+15\cdot 1.02+40\cdot(1+\delta)=100\cdot 1.02 \Rightarrow 40\cdot(1+\delta)=42.1 \Rightarrow$ $\delta=0.0525=5.25\%$.
16. $384385\cdot(1\pm 0.05494231)=405504$ i 363266 km. Razlika je 42238 km.
17. Na oročenu štednju kamate se pripisuju glavnici na kraju svakog razdoblja (godine) a obračunavaju se u odnosu na glavnicu sa početka tog razdoblja (složeni kamatni račun). Ako je p kamatna stopa te C_0 početna a $C_1, C_2, ...$ vrijednost glavnice nakon prvog, drugog, … razdoblja, imamo: $C_1=C_0(1+p)$, $C_2=C_1(1+p)=C_0(1+p)^2$, … i

općenito, nakon n razdoblja, $C_n = C_0(1+p)^n$. Dakle za $C_0 = 15000$, $p = 0.04$ i $n = 10$ imamo $C_{10} = 15000 \cdot 1.04^{10} = 22203.66$ EUR.

18. $C_n = 2C_0 \Rightarrow C_0 \cdot (1+p)^n = 2C_0 \Rightarrow (1+p)^n = 2 \Rightarrow n = \dfrac{\log 2}{\log(1+p)}$.

    Za $p = 0.04$ je $n \approx 17.673$ godina.

19. $C_0 \cdot 1.06^2 \cdot 1.05^3 \cdot 1.04^5 = 20000 \Rightarrow C_0 = 12638.15$ EUR.

20. Ako je x dužina ploče, y širina a z dubina, imamo $xyz = 5000$, pa je $0.95x \cdot 1.12y \cdot 1.1z = 1.1704xyz = 1.1704 \cdot 5000 = 5852\ \mathrm{m}^3$ betona.

KVALITATIVNI PODACI

Kvalitativna (kategorijalna) obilježja mogu biti nominalna (atributna i geografska) i uređajna (obilježja ranga). Izražavaju se opisno a ako su izražena brojčano nisu moguće računske operacije sa tim brojevima (npr. razred, stupanj stručne spreme, kućni broj itd.). Pripadni statistički skup uređujemo grupiranjem podataka prema vrsti obilježja. Za svaku vrstu obilježja formiramo jednu grupu podataka. Ukupan broj podataka u skupu (nizu) nazivamo *opseg skupa* a broj podataka u pojedinoj grupi *apsolutna frekvencija* te vrste obilježja (veličina grupe). Dakle, ako se skup sastoji od k međusobno disjunknih grupa sa apsolutnim frekvencijama $f_1, f_2, \ldots, f_k$ tada je zbroj svih apsolutnih frekvencija jednak opsegu skupa,

$$f_1 + f_2 + \ldots + f_k = N \quad \left(\sum_{i=1}^{k} f_i = N\right). \tag{5}$$

Omjer pojedine apsolutne frekvencije i opsega skupa nazivamo *relativna frekvencija*. Uz navedene oznake za relativne frekvencije p_i vrijedi

$$p_i = \frac{f_i}{N},\ i = 1, 2, \ldots, k \quad \Rightarrow \quad p_1 + p_2 + \ldots + p_k = 1 \quad \left(\sum_{i=1}^{k} p_i = 1\right). \tag{6}$$

Vidimo da se relativne frekvencije uvijek nalaze u intervalu [0,1] pa se najčešće izražavaju u postocima. Općenito se u statističkoj analizi vrlo često koristi omjer (kvocijent) dviju veličina koje ima smisla međusobno dijeliti. To su *statistički omjerni brojevi* (različite vrste indikatora, koeficijenata, statističkih pokazatelja, stopa, prosječnih veličina, relativnih brojeva itd.), kao što su na primjer: gustoća naseljenosti, BDP po stanovniku, koeficijent pokrivenosti uvoza izvozom, prosječna plaća, kamatna stopa, stopa inflacije, prosječna ocjena, indeksi itd. Jedna vrsta omjernih brojeva koji se često koriste u analizi kvalitativnih podataka su *indeksi niza kvalitativnih podataka*. To su omjeri apsolutnih frekvencija sa zadanom baznom veličinom,

$$I_i = \frac{f_i}{B},\quad i = 1, 2, \ldots, k. \tag{7}$$

Ovisno o izboru bazne veličine B indeks može biti manji, jednak ili veći od 1 a najčešće se izražava u postocima (množi sa 100). Pri tome bazna veličina uvijek ima indeks 1 (100%),

Uređeni i grupirani podaci pregledno se prikazuju u *statističkim tabelama*. Tabele mogu poslužiti za prvi korak u analizi podataka, npr. praćenjem i uspoređivanjem frekvencija uočavamo strukturu promatrane pojave. Po potrebi mogu se izračunati i tabelirati i određeni omjerni brojevi. Slijedeći korak je *grafički prikaz* podataka koji omogućuje vizualni uvid u strukturu pojave. U upotrebi su raznovrsni grafički prikazi od kojih su najčešći površinski dijagrami (jednostavni, dvostruki, razdijeljeni i strukturni stupci, dijagram indeksa, proporcionalni krugovi, strukturni krugovi i polukrugovi) zatim slikovni dijagrami i statističke karte. Upoznajemo ih kroz konkretne primjere.

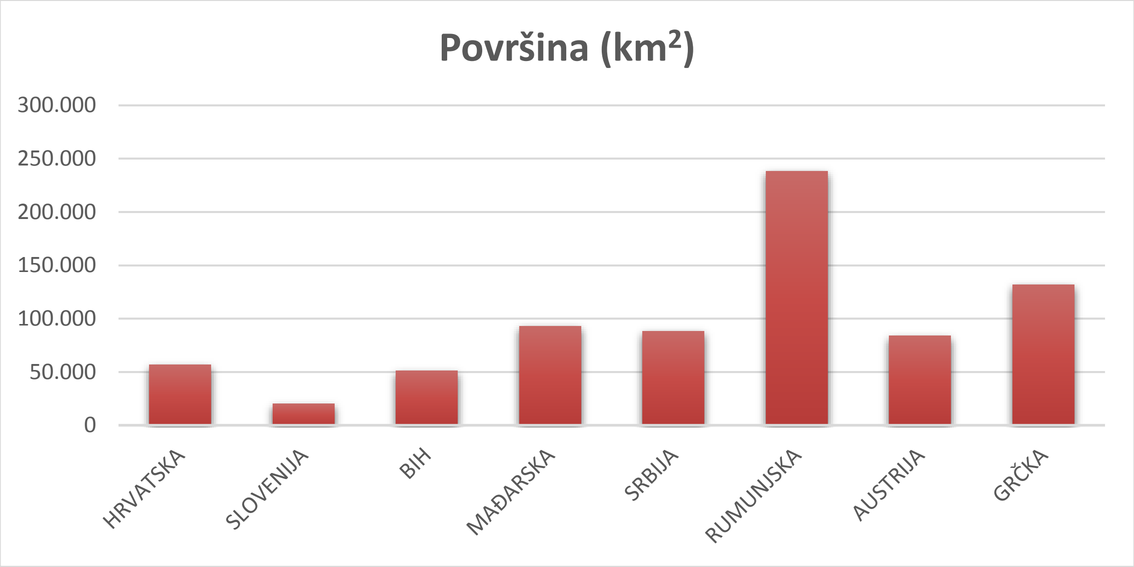
Primjer 5. Za odabrane zemlje u regiji (Hrvatska, Slovenija, BiH, Mađarska, Srbija, Rumunjska, Austrija i Grčka) u tabeli 1 dani su podaci za površinu, broj stanovnika, broj gradskog stanovništva i BDP (izvor Wikipedia, 2015). Prikažimo i analizirajmo dane podatke različitim grafičkim metodama.

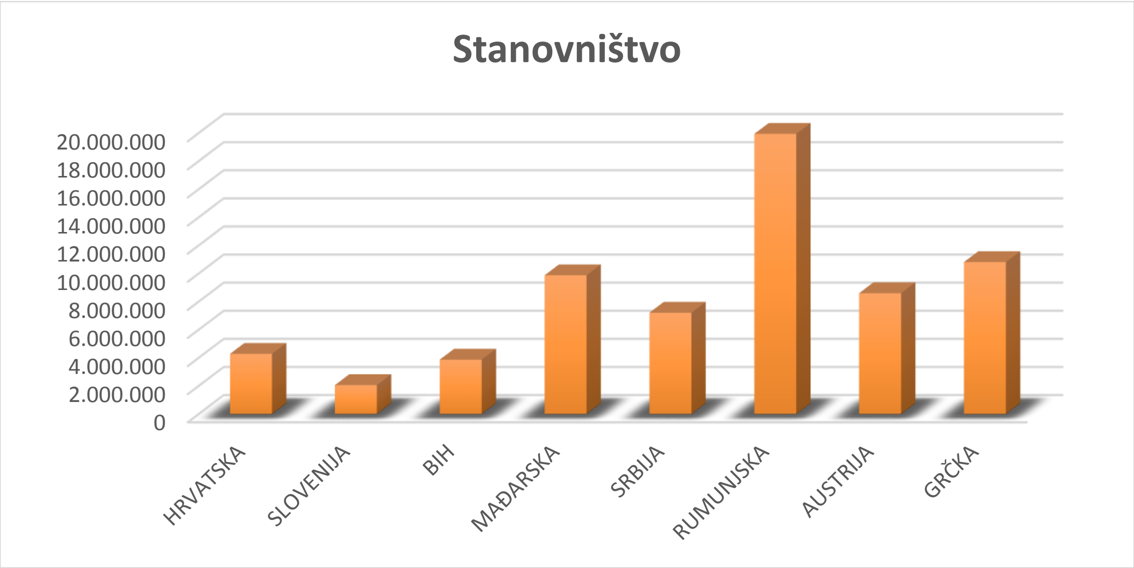
	Površina (km^2)	**Stanovništvo**	**Gradsko stanovništvo**	**BDP (milijarde USD)**
HRVATSKA	56.594	4.284.889	2.476.666	59,9
SLOVENIJA	20.273	2.061.085	1.028.481	43,0
BIH	51.197	3.871.643	1.870.004	19,1
MAĐARSKA	93.030	9.877.365	6.864.769	145,2
SRBIJA	88.361	7.209.764	4.066.307	42,6
RUMUNJSKA	238.391	19.942.642	10.529.715	215,3
AUSTRIJA	83.879	8.602.112	5.823.630	389,6
GRČKA	131.957	10.815.197	6.640.531	207,0

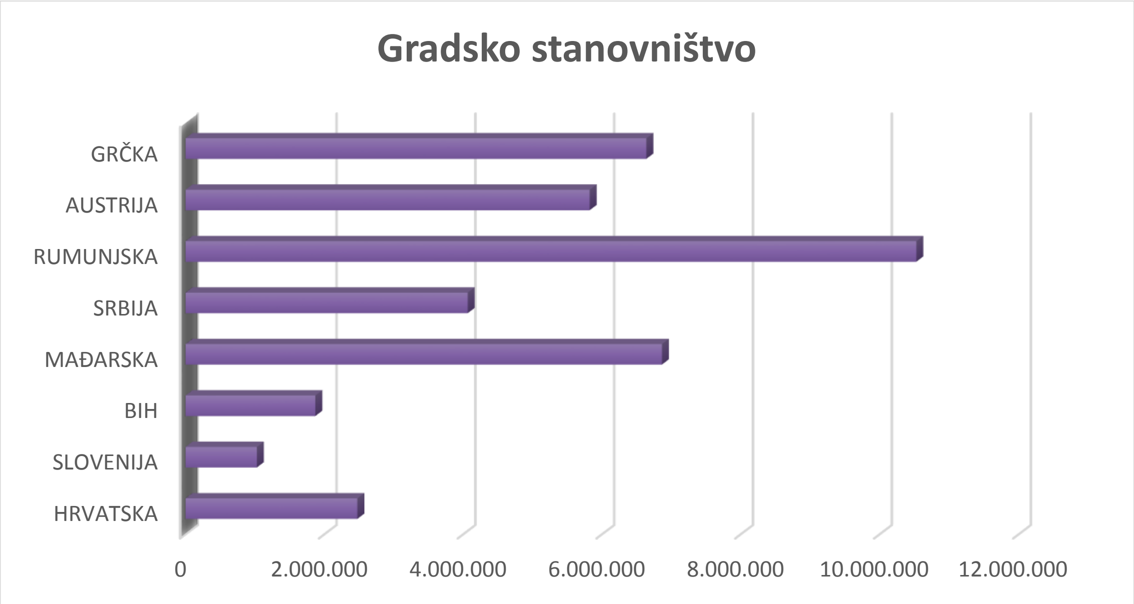
Tabela 1. Odabrane zemlje regije

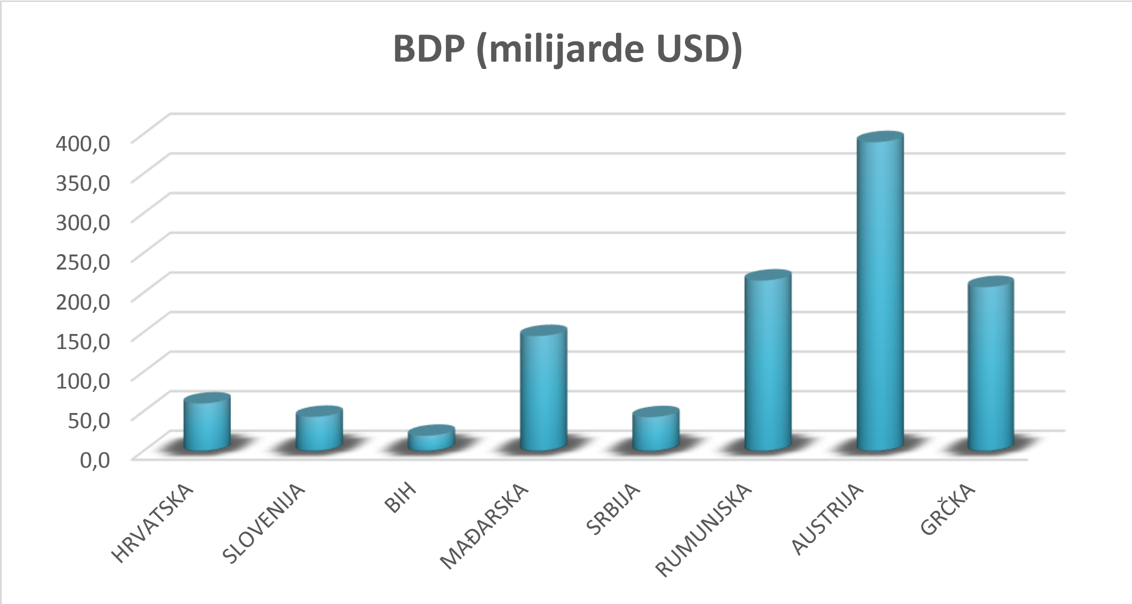
Uobičajeni način prikazivanja kvalitativnih podataka je pomoću *površinskih dijagrama*. Površine odabranih geometrijskih likova proporcionalne su frekvencijama (apsolutnim ili relativnim). Najjednostavniji prikaz je *dijagram odvojenih uspravnih ili položenih stupaca* istih baza (plošni – pravokutnici, trokuti ili trodimenzionalni – kvadri, valjci, piramide, stošci i sl.). Kako su im iste baze frekvencije su predočene njihovim visinama.

Iz prikazanih dijagrama na slici 2 možemo analizirati odnos apsolutnih iznosa (frekvencija) promatranih veličina (površine, broja stanovnika itd.) po pojedinim obilježjima (zemljama regije).


Površina (km²)
300.000
250.000
200.000
150.000
100.000
50.000
0
HRVATSKA
SLOVENIJA
BIH
MAĐARSKA
SRBIJA
RUMUNJSKA
AUSTRIJA
GRČKA


Stanovništvo
20.000.000
18.000.000
16.000.000
14.000.000
12.000.000
10.000.000
8.000.000
6.000.000
4.000.000
2.000.000
0
HRVATSKA
SLOVENIJA
BIH
MAĐARSKA
SRBIJA
RUMUNJSKA
AUSTRIJA
GRČKA


Gradsko stanovništvo
GRČKA
AUSTRIJA
RUMUNJSKA
SRBIJA
MAĐARSKA
BIH
SLOVENIJA
HRVATSKA
0
2.000.000
4.000.000
6.000.000
8.000.000
10.000.000
12.000.000

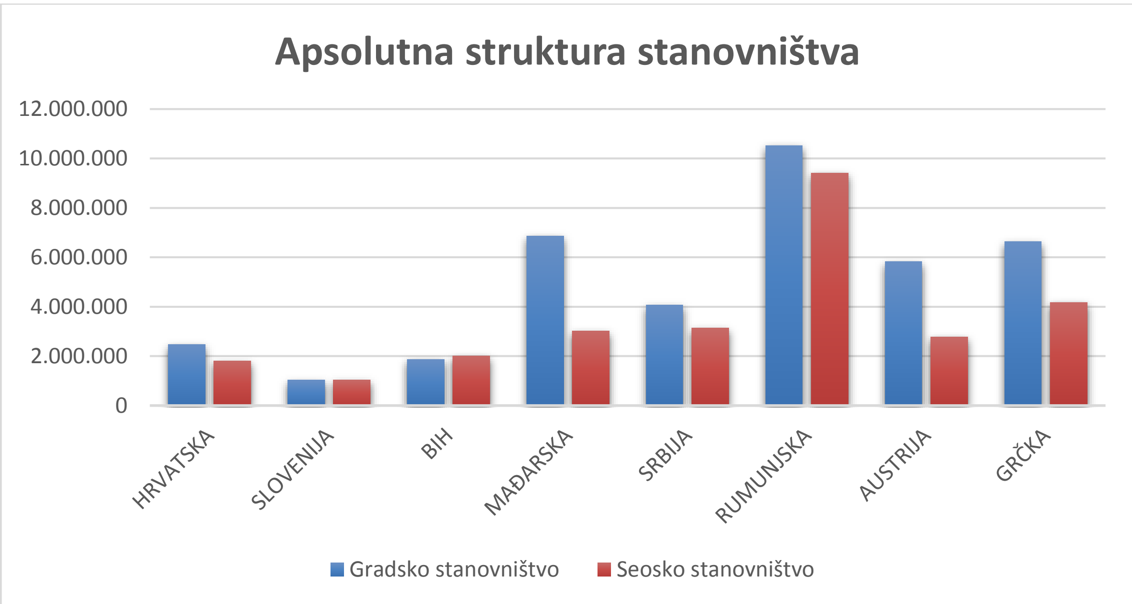


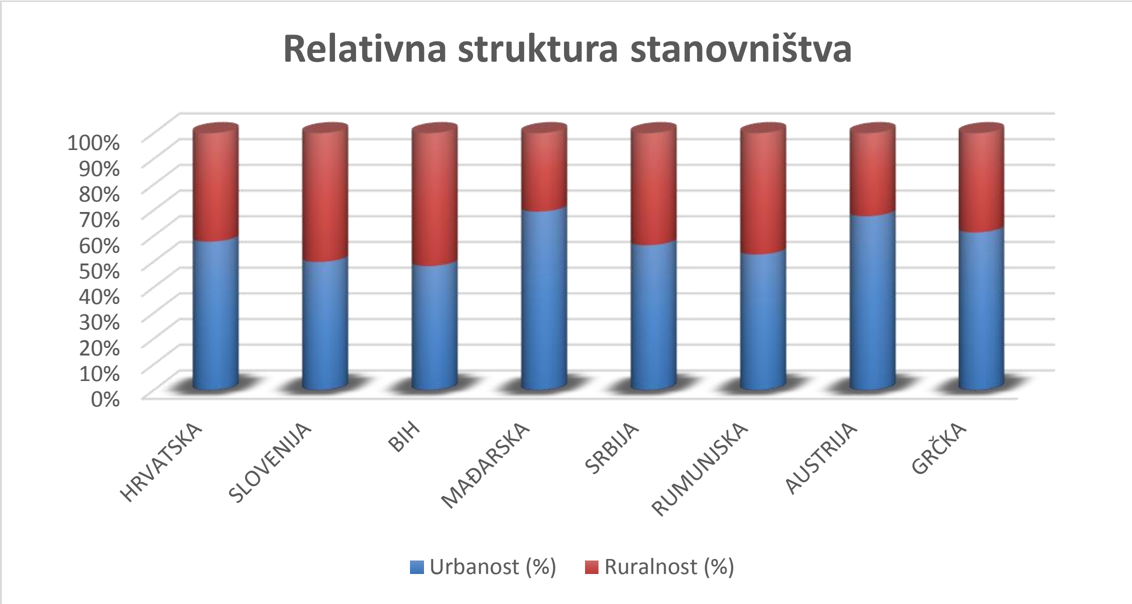
Slika 2. Prikaz jednostavnim uspravnim i položenim stupcima

Za usporedbu dvaju ili više kvalitativnih nizova (kao i za usporedbu razdijeljenih frekvencija istog niza) možemo koristiti *dijagram dvostrukih ili višestrukih stupaca*. Pri tome je prirodno uspoređivati ista obilježja izražena u istim mjernim jedinicama. U našem primjeru prikazujemo apsolutnu strukturu stanovništva (gradsko i seosko) dijagramom dvostrukih stupaca te relativnu strukturu dijagramom razdijeljenih stupaca (slika 3). U tu svrhu definirali smo relativne pokazatelje (statističke omjerne brojeve) urbanost i ruralnost,

$$\text{urbanost} = \frac{\text{broj gradskog stanovništva}}{\text{ukupan broj stanovnika}}, \qquad \text{ruralnost} = 1 - \text{urbanost},$$

koje smo izrazili u postocima (tabela 2).





Slika 3. Prikaz dvostrukim i razdijeljenim strukturnim stupcima

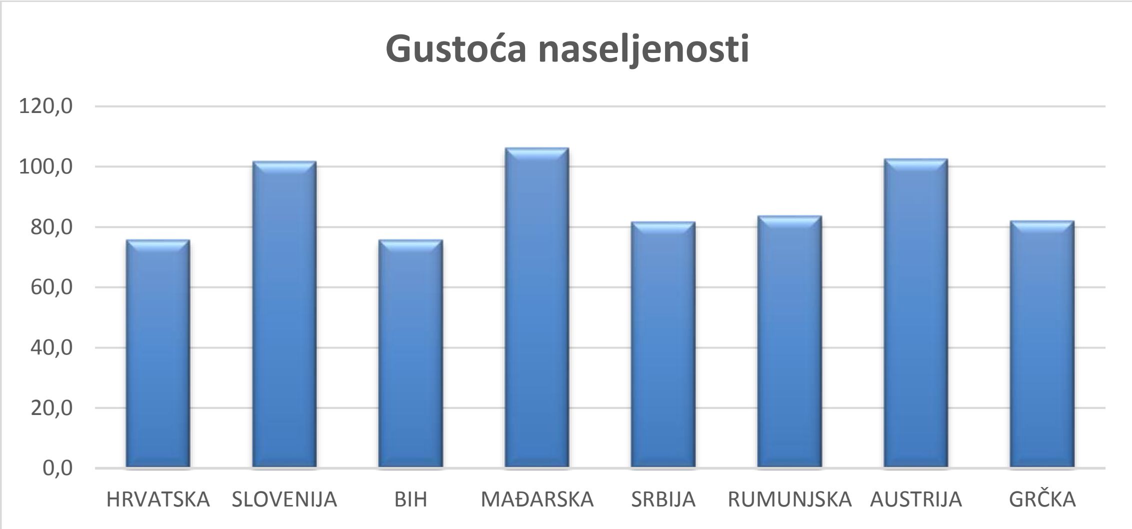
	Urbanost (%)	Ruralnost (%)	Gustoća naseljenosti	Indeksi gustoće naseljenosti	BDP po stanovniku	Indeksi BDP po stanovniku
HRVATSKA	57,80	42,20	75,7	100,0	13979	100,0
SLOVENIJA	49,90	50,10	101,7	134,3	20863	149,2
BIH	48,30	51,70	75,6	99,9	4933	35,3
MAĐARSKA	69,50	30,50	106,2	140,3	14700	105,2
SRBIJA	56,40	43,60	81,6	107,8	5909	42,3
RUMUNJSKA	52,80	47,20	83,7	110,5	10796	77,2
AUSTRIJA	67,70	32,30	102,6	135,5	45291	324,0
GRČKA	61,40	38,60	82,0	108,3	19140	136,9

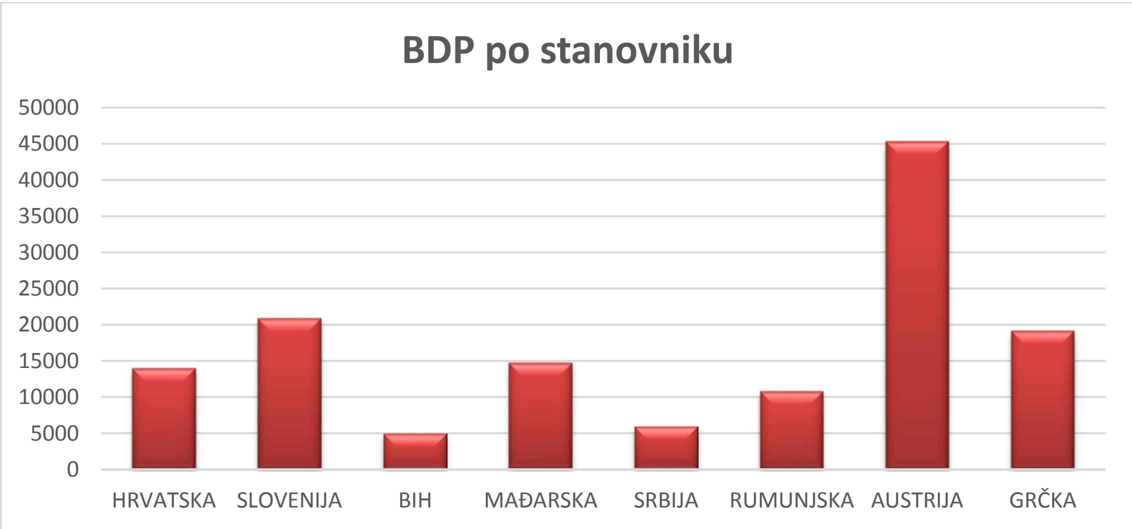
Tabela 2. Relativni pokazatelji za odabrane zemlje regije

Za detaljniju analizu naseljenosti i gospodarske razvijenosti navedenih zemalja promatrat ćemo i relativne pokazatelje gustoće naseljenosti te BDP po stanovniku,

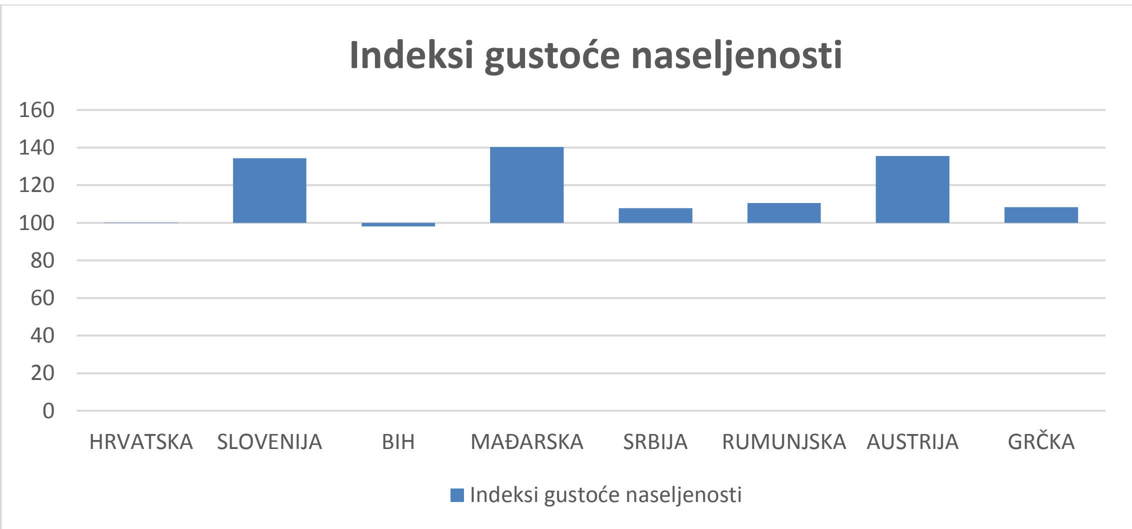
$$\text{gustoća naseljenosti} = \frac{\text{broj stanovnika}}{\text{površina}}, \qquad \text{BDP po stanovniku} = \frac{\text{BDP}}{\text{broj stanovnika}},$$

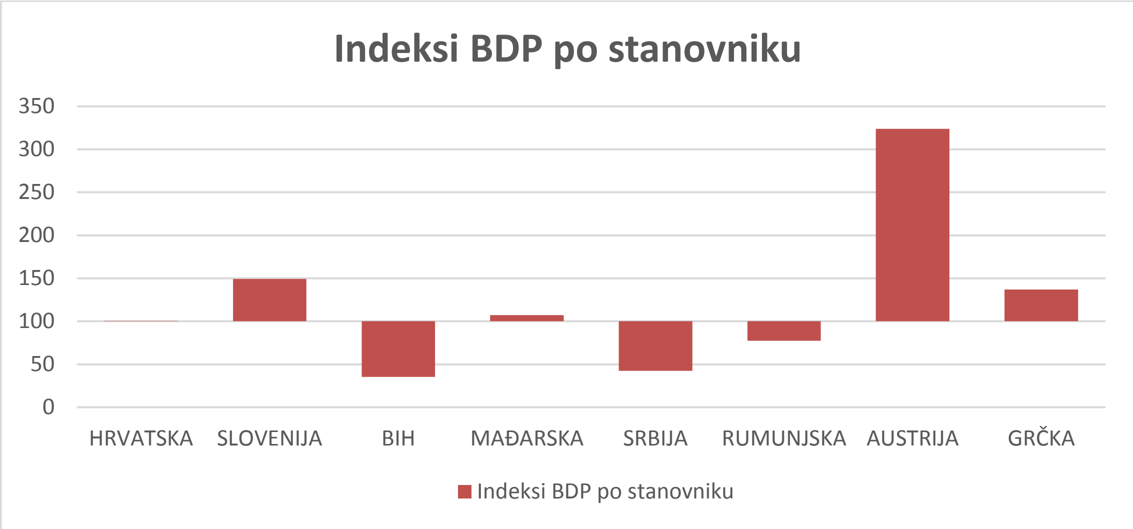
a koji su također navedeni u tabeli 2. Prikazujemo ih dijagramom jednostavnih stupaca (slika 4) i *dijagramom indeksa* (slika 5). Za bazni indeks uzeli smo podatak za Hrvatsku (B) te na temelju podataka za pojedine zemlje (f_i) izračunali indekse (I_i) formulom (7) (tabela 2).





Slika 4. Dijagrami jednostavnih stupaca gustoće naseljenosti i BDP po stanovniku





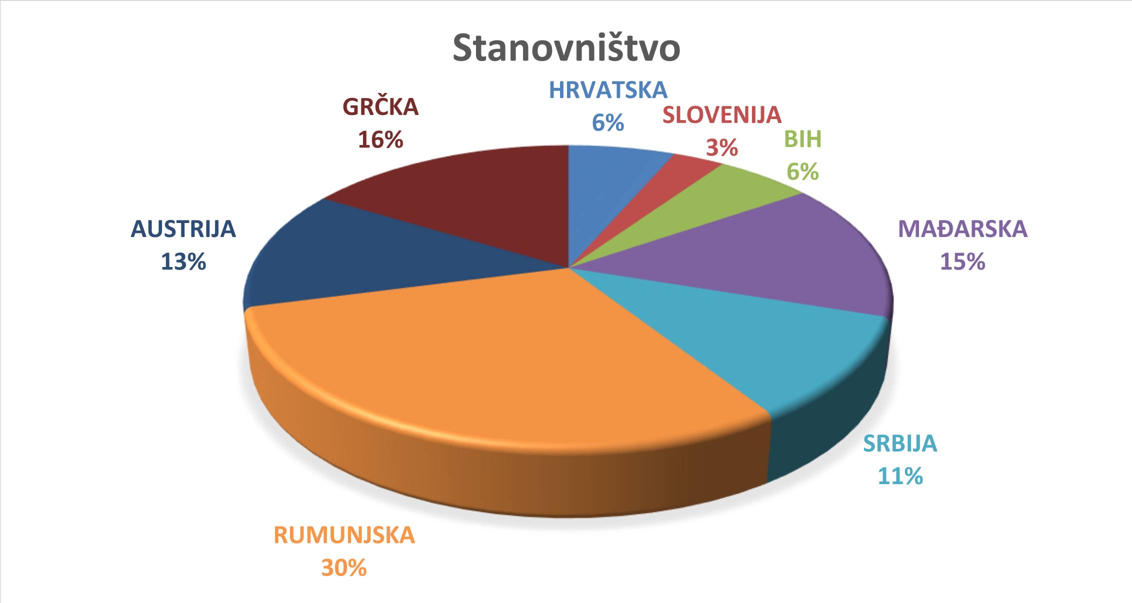
Slika 5. Dijagrami indeksa gustoće naseljenosti i BDP po stanovniku

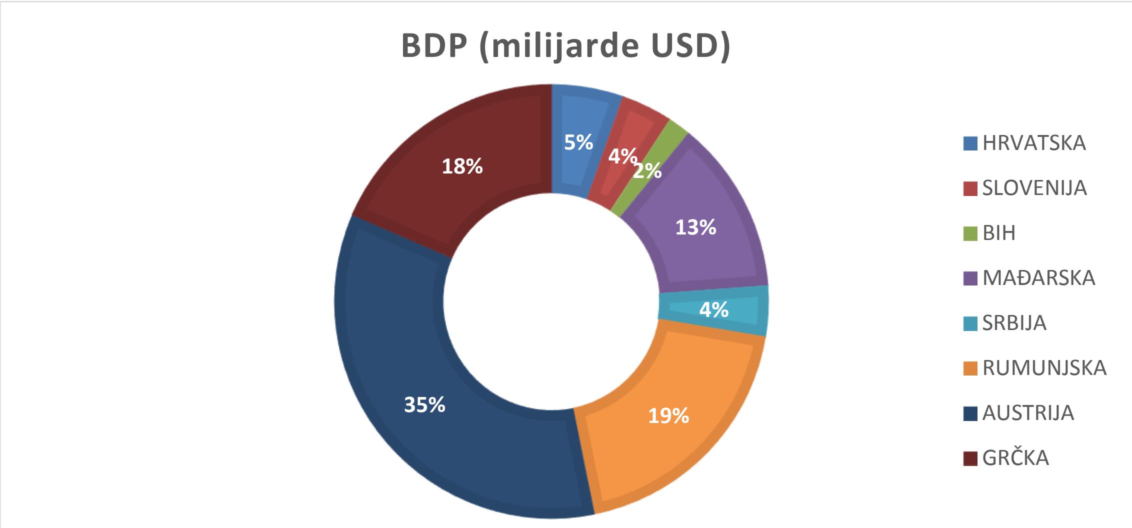
Ovakvi različiti načini prikazivanja omogućuju nam analizu sa različitih aspekata čime dobivamo cjeloviti uvid u promatranu situaciju. Tako na primjer iz slike 5 možemo uočiti lošiju gustoću naseljenosti i prosječnu ekonomsku razvijenost Hrvatske u odnosu na ostale promatrane zemlje regije. Za navedene međusobne odnose imamo točne kvantitativne pokazatelje. Primijetimo da stupci na dijagramu indeksa u stvari predočuju relativno odstupanje (udaljenost) vrijednosti promatrane veličine od bazne vrijednosti pa ga nazivamo i *dijagram relativnih odstupanja*. Pri tome stupci sa gornje strane bazne razine predočuju pozitivno a sa donje strane negativno odstupanje. Napomenimo da se na isti način mogu prikazati i apsolutna odstupanja (*dijagram apsolutnih odstupanja*).

Ako broj stanovnika ili BDP pojedine zemlje (f_i) promatramo u odnosu na ukupan broj stanovnika (66664697) ili ukupan BDP (1121.7) svih promatranih zemalja (N) analizu možemo provesti i pomoću *strukturnih geometrijskih likova* (krugovi, pravokutnici i sl.) ili *tijela* (valjci, prizme i sl.). Tu je frekvencija obilježja proporcionalna površini ili volumenu onog dijela lika ili tijela koji tu frekvenciju predstavlja. Ako je L dimenzija strukturnog lika ili tijela u smjeru koje se struktura prikazuje (duljina, visina, opseg, površina, volumen i sl.) a l_i dimenzija onog dijela kome pripada frekvencija f_i tada je, u skladu sa ranijim oznakama,

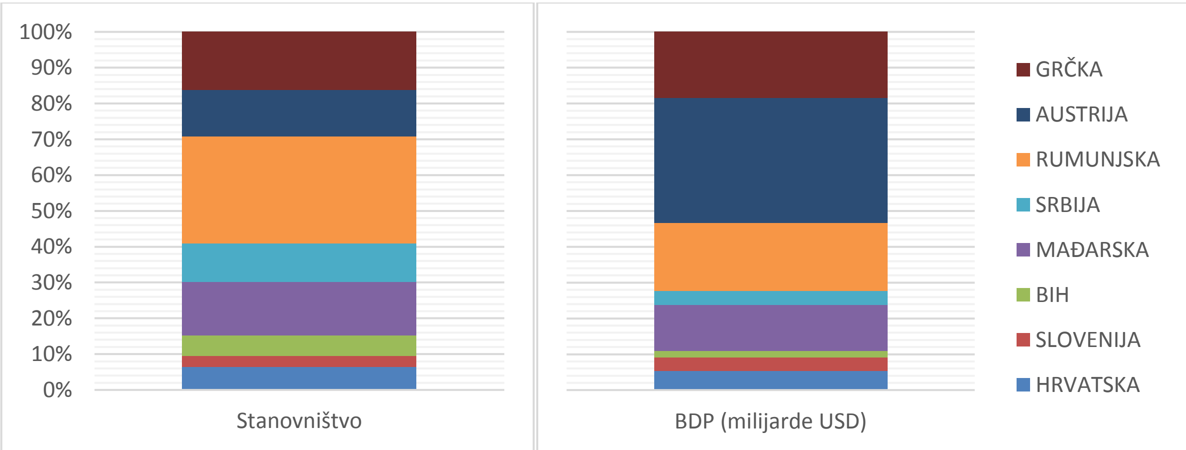
$$l_i : L = f_i : N \;\Rightarrow\; l_i = \frac{f_i}{N} \cdot L = p_i \cdot L, \quad i = 1, 2, \ldots, k. \tag{8}$$

Najčešće se koriste strukturni krugovi ili valjci. Pri tome svakoj pojedinoj frekvenciji odgovara kružni isječak pa je u formuli (8) l_i središnji kut tog isječka a $N = 360^o$.





Slika 6. Strukturni krugovi broja stanovnika i BDP-a



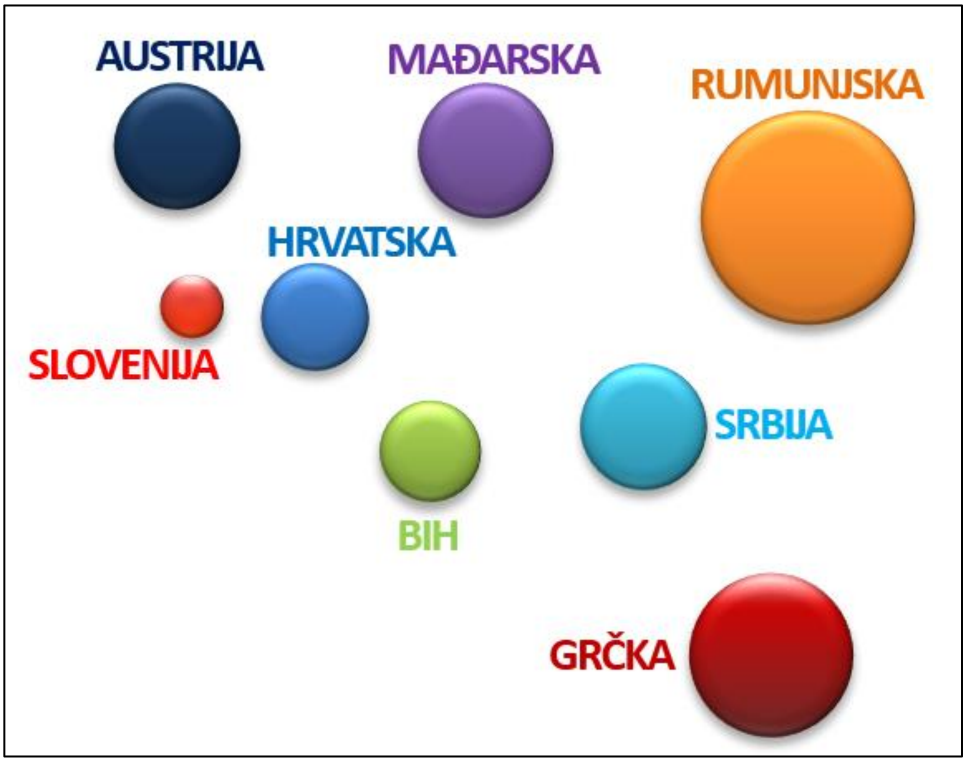
Slika 7. Strukturni pravokutnici broja stanovnika i BDP-a

Strukturni likovi i tijela koriste se i za vizualnu i kvantitativnu usporedbu struktura dviju ili više veličina. Tako na primjer iz slike 6 ili 7 uočavamo visok stupanj ekonomske razvijenosti Austrije (mali udio stanovništva a veliki udio BDP-a) u odnosu na ostale promatrane zemlje regije (osobito Rumunjsku, Srbiju i BiH) kod kojih je odnos udjela stanovništva i BDP-a suprotan.

Od ostalih načina prikazivanja navodimo *dijagram proporcionalnih krugova* (ili nekih drugih likova – kvadrata, pravilnih trokuta, šesterokuta i sl.). Njihove površine su proporcionalne frekvencijama, $r_1^2\pi : r_2^2\pi : \ldots : r_k^2\pi = f_1 : f_2 : \ldots : f_k$ pa je $r_1 : r_2 : \ldots : r_k = \sqrt{f_1} : \sqrt{f_2} : \ldots : \sqrt{f_k}$. Ako fiksiramo jedan polumjer, npr. r_j (obično je to polumjer pridružen najmanjoj ili najvećoj frekvenciji), tada za ostale vrijedi

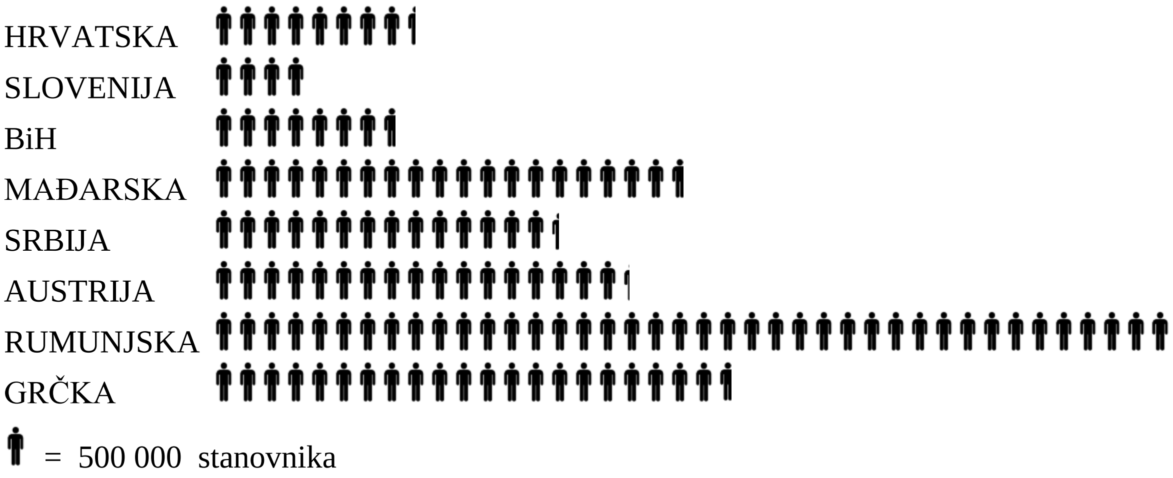
$$r_i : r_j = \sqrt{f_i} : \sqrt{f_j} \quad \Rightarrow \quad r_i = r_j \cdot \sqrt{\frac{f_i}{f_j}}, \quad i = 1, 2, \ldots, k.$$

Napomenimo da ova formula vrijedi kako za polumjere krugova tako i za stranice bilo kojeg pravilnog mnogokuta (trokuta, kvadrata, peterokuta, šesterokuta itd.). Na ovaj je način na slici 8 prikazana površina promatranih zemalja u regiji gdje smo zadali polumjer najmanjeg kruga koji predstavlja površinu Slovenije (4 mm).



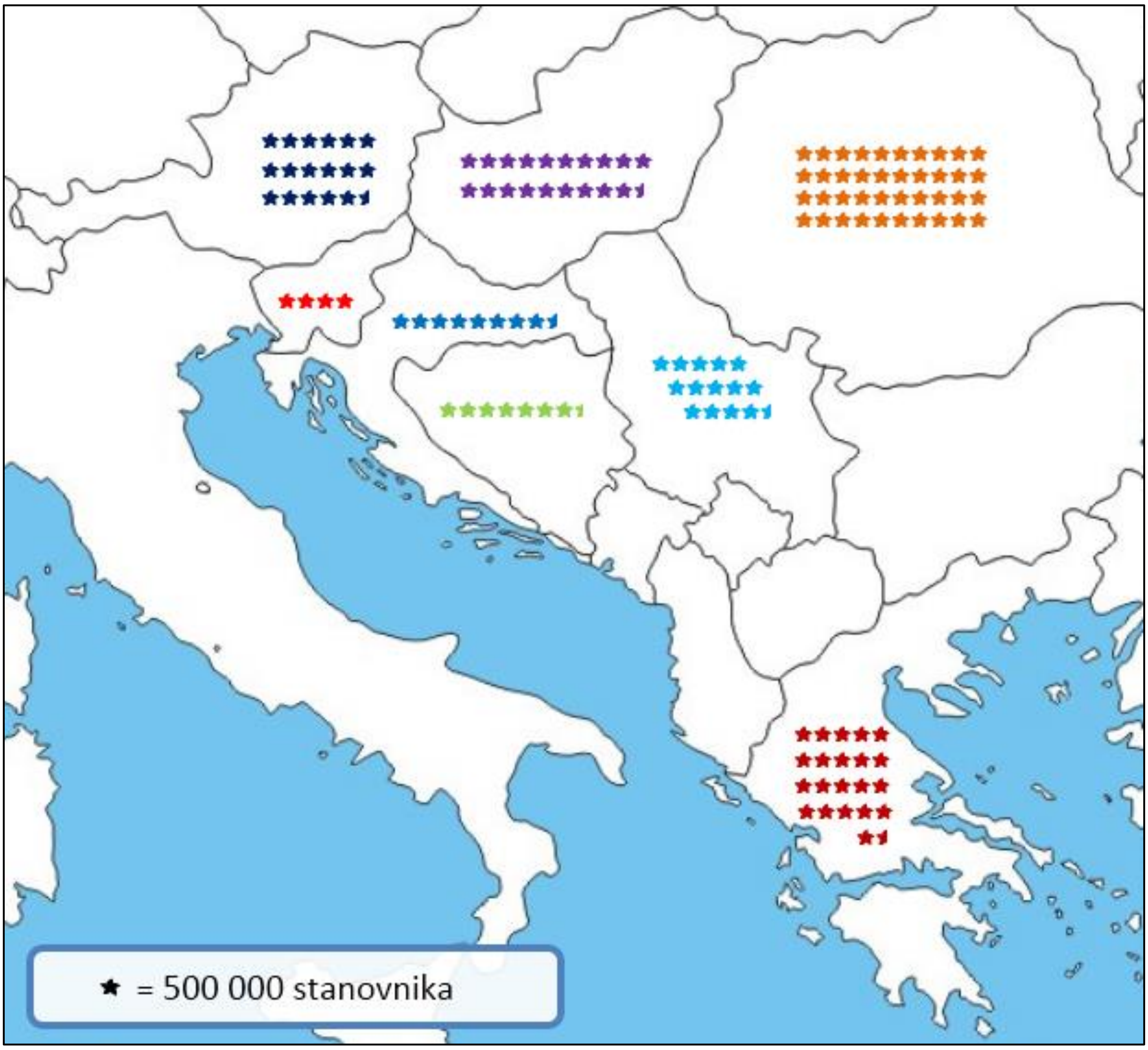
Slika 8. Prikaz površine proporcionalnim krugovima

Ponekad se za prikazivanje kvalitativnih podataka koriste i *slikovni dijagrami (piktografi)* koje često nalazimo u medijima. Odaberemo sličicu (znak ili simbol) koja podsjeća na podatke koje prikazujemo. Svaka takva sličica predočuje točno određeni broj podataka definiran u tumaču znakova.



Slika 9. Slikovni dijagram broja stanovnika

Dijagrami proporcionalnih geometrijskih likova i slikovni dijagrami mogu se ukomponirati u zemljopisne karte čime nastaju *statističke karte (kartogrami ili dijagramske karte).*

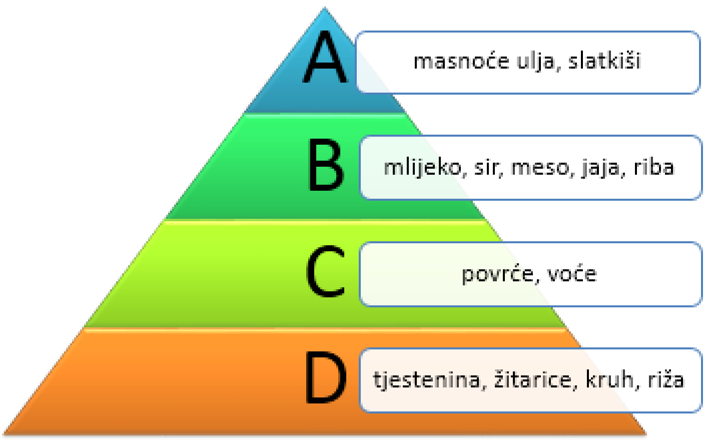


Slika 10. Statistička karta broja stanovnika

One prikazuju geografska obilježja promatranih veličina (struktura stanovništva, gospodarski pokazatelji, klimatski parametri, resursi itd.). Osim likova i sličica za prikaz frekvencija mogu se koristiti i točke, sjenčanja ili boje.

ZADACI

1. U proračunu nekog grada predviđeno je 260 milijuna kuna za investicije u tekućoj godini. Od toga je 117 milijuna namijenjeno izgradnji cesta. Ako investicije prikazujemo strukturnom prizmom visine 8 cm, kolika visina pripada iznosu namijenjenom za izgradnju cesta? Isto pitanje za središnji kut kružnog isječka ako investicije prikazujemo strukturnim krugom ili valjkom.
2. Izmjerite površine pojedinih prostorija u vašem stanu ili kući (hodnik, sobe, kuhinja, kupaonica itd.) te ih prikažite strukturnim krugom (valjkom). Usporedite međusobno dobivene prikaze.
3. Promatranu pojavu koja ima dva obilježja jednakih frekvencija prikazujemo strukturnim trokutom. Na kojoj visini trebamo povući razdjelnicu paralelnu sa osnovicom trokuta?
4. Prikupite podatke o općem uspjehu učenika vašeg razreda na kraju 8. razreda osnovne škole te 1. i 2. razreda srednje škole. Strukturnim pravokutnicima vizualno usporedite te uspjehe.
5. Režim zdrave prehrane prikazan je površinskim dijelovima trokuta iste visine na slici 11. Odredite postotne udjele pojedinih vrsta namirnica u tom režimu prehrane.



Slika 11. Struktura zdrave prehrane

6. Režim zdrave dijetne prehrane prikazan je volumnim dijelovima piramide iste visine na analogni način kao u prethodnom zadatku. Odredite postotne udjele pojedinih vrsta namirnica u tako definiranoj dijetnoj prehrani.
7. Dijagramom jednostavnih stupaca prikažite distribuciju datuma rođenja učenika vašeg razreda po mjesecima u godini.
8. Prikupite podatke o broju stanovnika nekoliko gradova ili mjesta u vašoj općini ili županiji i prikažite ih dijagramom jednostavnih stupaca i dijagramom indeksa (za baznu veličinu uzmite broj stanovnika vašeg mjesta).

9. Dijagramom dvostrukih stupaca prikažite odnos cijene i broja stranica nekoliko izabranih časopisa.
10. Jednim strukturnim krugom prikažite površine kontinenata a drugim mora i oceana. Površine krugova neka budu proporcionalne kopnenoj i vodenoj površini Zemlje.
11. Dijagramom indeksa prikažite gustoće nekoliko tvari npr. zrak, pluto, led, voda, zemlja, aluminij, željezo, olovo, živa, zlato, platina i sl. Za baznu veličinu uzmite gustoću vode 1000 kg/m^3.
12. Odaberite nekoliko istovrsnih proizvoda te dijagramom višestrukih stupaca usporedite relativni udio glavnih sastojaka, npr. uzmite 3-5 vrsta čokolade, pogledajte sastav te dijagramom trostrukih stupaca prikažite relativni (postotni) udio za kakao, šećer i mlijeko.
13. Ova knjiga ima pet poglavlja. Odredite relativni udio (relativnu frekvenciju) broja stranica svakog poglavlja u odnosu na ukupan broj stranica knjige te prikažite rezultate jednostavnim stupcima ili nekim strukturnim likom (tijelom).
14. Omjer $1:2:4:8$ prikažite površinama proporcionalnih: (a) koncentričnih krugova, (b) kvadrata sa jednim zajedničkim kutom.
15. Ako 5 cm na zemljopisnoj karti predstavlja 40 km u prirodi odredite: (a) u kojem mjerilu je napravljena karta? (b) koliku površinu u prirodi predstavlja krug polumjera 1 cm na karti?
16. Srednji promjer Zemlje je 12735 km a Mjeseca 3476 km. Odredite: (a) kolika dužina predstavlja promjer Zemlje ako 1 cm predstavlja promjer Mjeseca? (b) što predstavljaju površine krugova čiji su polumjeri dužine iz zadatka (a) te koliki je njihov odnos (omjer)? (c) koliki je odnos volumena Zemlje i Mjeseca?
17. Prikupite podatke o veličini pojedinih planeta i Sunca te napravite usporedbu sa Zemljom (promjer Zemlje = 1 cm) kao u zadatku 10.
18. Ako udaljenost od Zemlje do Sunca (150 milijuna km) predstavimo dužinom od 1 mm, kolika dužina predstavlja 1 svjetlosnu godinu (1 ly – light year)? Kako bi u tom mjerilu izgledao Sunčev sustav (najudaljeniji planet Neptun udaljen je 4.5 milijardi km od Sunca) a gdje bi bila nama najbliža zvijezda Proxima Centauri koja je od Sunca udaljena 4.24 ly?
19. Promjer atoma je oko 10^{-10} m a promjer atomske jezgre 10^{-15} m. Ako atomsku jezgru predočimo kuglicom promjera 1 mm kolika kugla predočuje atom?
20. Jedan proizvod proizvodi se u tri dislocirana pogona A, B i C. Opseg proizvodnje (u komadima) i ukupni troškovi proizvodnje (u kunama) za prošli mjesec dani su u tabeli.

Pogon	***Opseg proizvodnje***	***Ukupni troškovi***
A	16520	379960
B	19380	416670
C	14100	313725

Odredite slijedeće statističke omjerne brojeve.

(a) Ukupni opseg proizvodnje i ukupne troškove.

(b) Postotni udio (relativne frekvencije) količina i troškova pojedinih pogona.

(c) Prosječne troškove po komadu za pojedine pogone.

(d) Indekse prosječnih troškova po komadu u odnosu na pogon C.
(e) Prosječne troškove po komadu za čitavu proizvodnju.
(f) Prikažite neke od dobivenih rezultata proizvoljno izabranim grafičkim prikazom.

21. Kako se izračunaju slijedeći omjerni brojevi: (a) prosječna plaća, (b) gustoća naseljenosti, (c) natalitet (u promilima), (d) broj putnika po letu, (e) pokrivenost uvoza izvozom, (f) broj liječnika na 5000 stanovnika, (g) prinos kukuruza po hektaru ?
22. Slikovnim dijagramom i statističkom kartom prikažite distribuciju stanovništva:
(a) RH po županijama, (b) svjetskog stanovništva po kontinentima.

RJEŠENJA

1. Koristimo formulu (8): $l = \frac{117}{260} \cdot 8 = 0.45 \cdot 8 = 3.6$ cm odnosno $l = 0.45 \cdot 360^o = 162^o$.
2. Koristite podatke dobivene mjerenjem. Površina kruga odgovara ukupnoj površini stana (kuće).
3. Odnos površina manjeg i većeg trokuta je $1:2$ pa je odnos visina $\sqrt{1}:\sqrt{2} \approx 0.7071$. Razdjelnicu treba povući na 70.71% visine od vrha ili na 29,29% visine od osnovice.
4. Koristite stvarne podatke.
5. Odnos visina četiri trokuta je $1:2:3:4$ pa je odnos njihovih površina $1^2:2^2:3^2:4^2 = 1:4:9:16$. Udio grupe namirnica A u prehrani je $1/16 = 6.25\%$, grupe namirnica B je $4/16 - 1/16 = 3/16 = 18.75\%$, grupe C je $9/16 - 4/16 = 5/16 = 31.25\%$ i grupe D je $16/16 - 9/16 = 7/16 = 43.75\%$.
6. Odnos volumena piramida je $1^3:2^3:3^3:4^3 = 1:8:27:64$ pa na sličan način dobivamo: A $1/64 = 1.5625\%$, B $7/64 = 10.9375\%$, C $19/64 = 29.6875\%$, D $37/64 = 57.8125\%$.
7. Koristite stvarne podatke.
8. Koristite stvarne podatke.
9. Koristite izabrane podatke. Za prikaz koristite dvije ordinatne osi, lijevu i desnu, jednu za cijenu a drugu za broj stranica.
10. Od 510 milijuna km^2 ukupne Zemljine površine kopno zauzima 149 milijuna km^2 a mora i oceani 361 milijun km^2. Omjer polumjera (promjera) strukturnih krugova treba biti $\sqrt{361}:\sqrt{149} = 1.55654...:1$ (krug koji pokazuje strukturu vodene površine ima za oko 56% veći polumjer od strukturnog kruga za kopno). Strukturu krugova napravite prema stvarnim površinama pojedinih kontinenata, mora i oceana.
11. Prikupite podatke o stvarnoj gustoći izabranih tvari.
12. Koristite izabrane podatke.
13. Podatke o broju stranica odredite iz sadržaja na početku knjige.
14. Ako je r polumjer najmanjeg kruga (stranica najmanjeg kvadrata), tada su polumjeri (stranice) redom: $r,\ \sqrt{2}\cdot r \approx 1.4r,\ 2r,\ \sqrt{8}\cdot r = 2\sqrt{2}\cdot r \approx 2.8r$.
15. (a) $5\,\text{cm} : 40\,\text{km} = 5 : (40 \cdot 10^5) = 1 : 800\,000$,
(b) $(800\,000\,\text{cm})^2 \cdot \pi = (8\,\text{km})^2 \cdot \pi = 64\pi\,\text{km}^2 \approx 201\,\text{km}^2$.

16. (a) $x:1=12735:3476 \Rightarrow x=3.6637$ cm, (b) Površina kruga, čiji je polumjer jednak promjeru kugle, jednaka je površini kugle, $(2r)^2\pi=4r^2\pi$. Dakle, površine navedenih krugova predstavljaju površine Zemlje i Mjeseca a omjer im je $x^2:1=13.423:1$. (c) $x^3:1=49.18:1$.
17. Na primjer za Sunce (promjer 1392000 km) je: (a) 109.3 cm, (b) 11947.6 : 1, (c) 1305933 : 1.
18. Kako je brzina svjetlosti u vakuumu $299792.458 \approx 3\cdot 10^5$ km/s, duljina svjetlosne godine je $1\text{ly}=(365\cdot 24\cdot 60\cdot 60)\cdot(3\cdot 10^5)=9.46\cdot 10^{12}$ km (9.46 bilijuna km), pa je duljina tražene dužine $(9.46\cdot 10^{12}):(150\cdot 10^6)=63067$ mm ≈ 63 m. U navedenom mjerilu planeti kruže oko Sunca približno po koncentričnim kružnicama a najveća (Neptunova putanja) ima polumjer $(4.5\cdot 10^9):(150\cdot 10^6)=30$ mm = 3 cm, dok je najbliža zvijezda udaljena $4.24\cdot 63\approx 267$ metara. Uočite ogromnu prazninu svemirskog prostora !
19. Ako je atomska jezgra kuglica promjera 1 mm tada je čitav atom kugla promjera $10^{-10}:10^{-15}=100\,000$ mm = 100 m. Uočite ogromnu prazninu unutar atoma !
20. (a) 50000 komada, 1 110 355 kuna. (b) Količine: A 33.04%, B 38.76%, C 28.2%. Troškovi: A 34.22%, B 37.53%, C 28.25%. (c) A 23.00, B 21.50, C 22.25 kuna po komadu. (d) A 103.37, B 96.63, C 100. (e) $22.2071\approx 22.21$ kuna po komadu.
21. (a) ukupna masa plaća / broj zaposlenika, (b) broj stanovnika / površina, (c) (broj novorođenih / broj stanovnika) x 1000, (d) broj prevezenih putnika / broj letova, (e) ukupna vrijednost izvoza / ukupna vrijednost uvoza, (f) (broj liječnika / broj stanovnika) x 5000, (g) ukupna ubrana količina kukuruza / površina zemljišta u hektarima.
22. Za mjerilo možete uzeti npr. 1 sličica (simbol) = 25000 stanovnika za (a) odnosno 250 milijuna stanovnika za (b).

KVANTITATIVNI PODACI

Numeričko obilježje predočeno je kvantitativnim podacima (brojevima) a može biti diskretno (prekidno) ili kontinuirano (neprekidno). Diskretno obilježje koje ima relativno mali broj vrijednosti (oblika) grupira se u *numeričke grupe* rastućim ili padajućim redoslijedom čime dobivamo *distribuciju frekvencija* tog obilježja. Numeričku grupu čine svi podaci iste vrijednosti a predočuje se vrijednošću i pripadnom frekvencijom. Ako dakle, diskretno numeričko obilježje (prekidna varijabla X) poprima k različitih vrijednosti $x_1, x_2, ..., x_k$ poredanih u rastućem (padajućem) redoslijedu sa frekvencijama $f_1, f_2, ..., f_k$, tada je nizom uređenih parova,

$$(x_1, f_1), (x_2, f_2), \dots, (x_k, f_k), \quad x_i < x_{i+1}, \; i=1,2,\dots,k-1, \tag{9}$$

dana distribucija frekvencija tog obilježja (varijable X). Numerička obilježja možemo prikazivati tabelarno ili grafički (dijagram s točkama, stablo-list dijagram, linijski, površinski i kumulativni dijagram). Navedene prikaze ilustriramo kroz primjere.

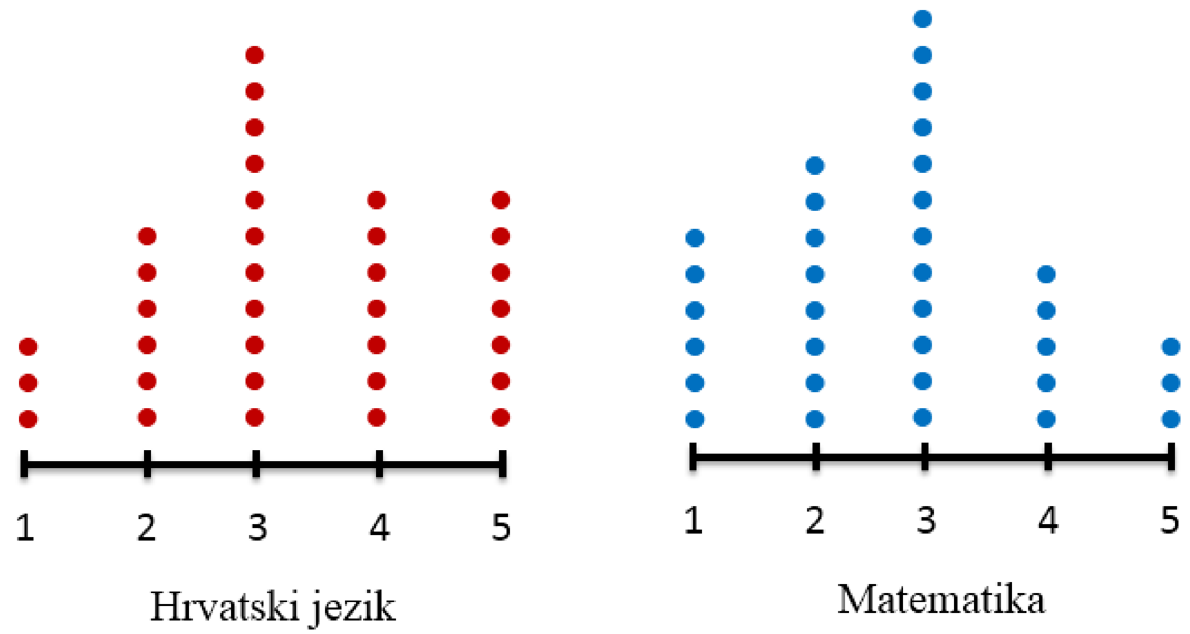
Primjer 6. Uspjeh učenika jednog trećeg razreda na kraju školske godine iz predmeta Hrvatski jezik i Matematika dani su u tabeli 6. U prvoj koloni je abecedni redni broj učenika po imeniku. U preostalim kolonama su prosječne ocjene dobivene temeljem pojedinačnih ocjena tokom godine te zaokružene konačne ocjene iz ta dva predmeta za svakog učenika posebno. Prikažimo i analizirajmo grafički, na različite načine, ocjene iz tih predmeta.

	HRVATSKI		**MATEMATIKA**	
R.br.	Prosjek	Kon.	Prosjek	Kon.
1.	3.4	3	2.2	2
2.	2.5	3	1.3	1
3.	3.2	3	3.0	3
4.	4.1	4	1.9	2
5.	2.1	2	1.0	1
6.	3.9	4	2.4	2
7.	4.8	5	4.1	4
8.	1.4	1	3.2	3
9.	2.8	3	2.7	3
10.	4.5	5	3.4	3
11.	3.3	3	2.0	2
12.	5.0	5	4.0	4
13.	3.0	3	1.4	1
14.	4.4	4	5.0	5
15.	4.9	5	2.4	2
16	1.2	1	1.8	2
17.	2.8	3	3.2	3

	HRVATSKI		**MATEMATIKA**	
R.br.	Prosjek	Kon.	Prosjek	Kon.
18.	2.3	2	2.7	3
19.	3.6	4	3.6	4
20.	4.1	4	3.2	3
21.	1.9	2	1.2	1
22.	4.8	5	2.7	3
23.	4.6	5	3.6	4
24.	2.1	2	1.4	1
25.	5.0	5	4.0	4
26.	3.9	4	4.6	5
27.	3.3	3	2.2	2
28.	1.2	1	2.0	2
29.	3.3	3	2.8	3
30.	3.3	3	3.0	3
31.	2.5	3	2.5	3
32.	2.4	2	1.3	1
33.	4.0	4	4.8	5
34.	2.4	2	2.7	3

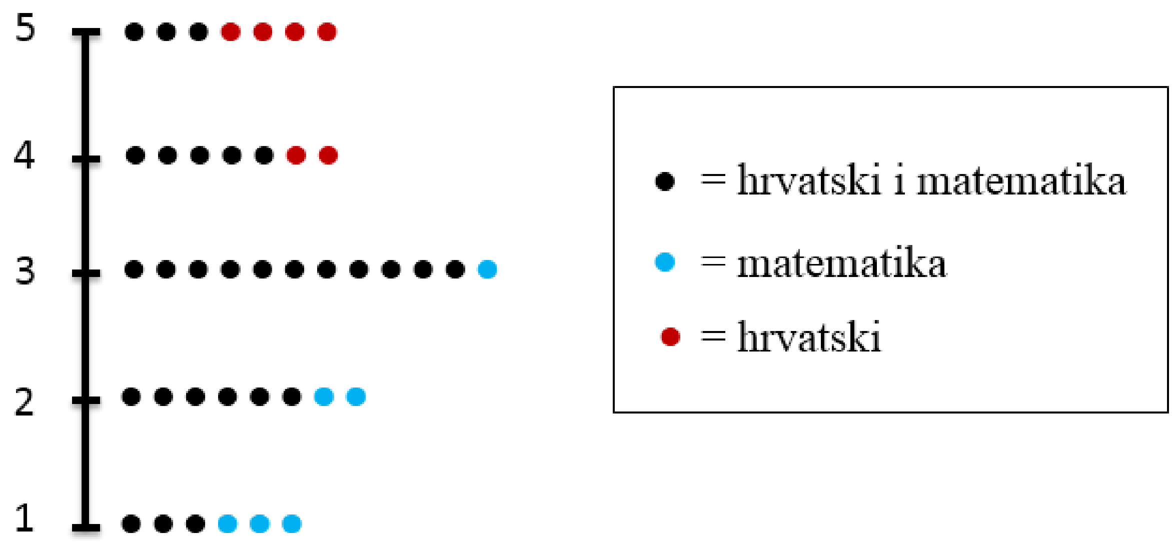
Tabela 6. Podaci o uspjehu na kraju školske godine

Jedan slikoviti način grafičkog prikazivanja je *dijagram sa točkama*. Broj točaka za svaku numeričku vrijednost varijable pokazuje njenu frekvenciju. Ovaj dijagram je prikladan za prikaz manjeg broja podataka (slika 12).



Slika 12. Prikaz konačnih ocjena dijagramom sa točkama

Dijagram može biti uspravan ili položen a može poslužiti i za kombinirani prikaz dviju vrsta obilježja (slika 13).



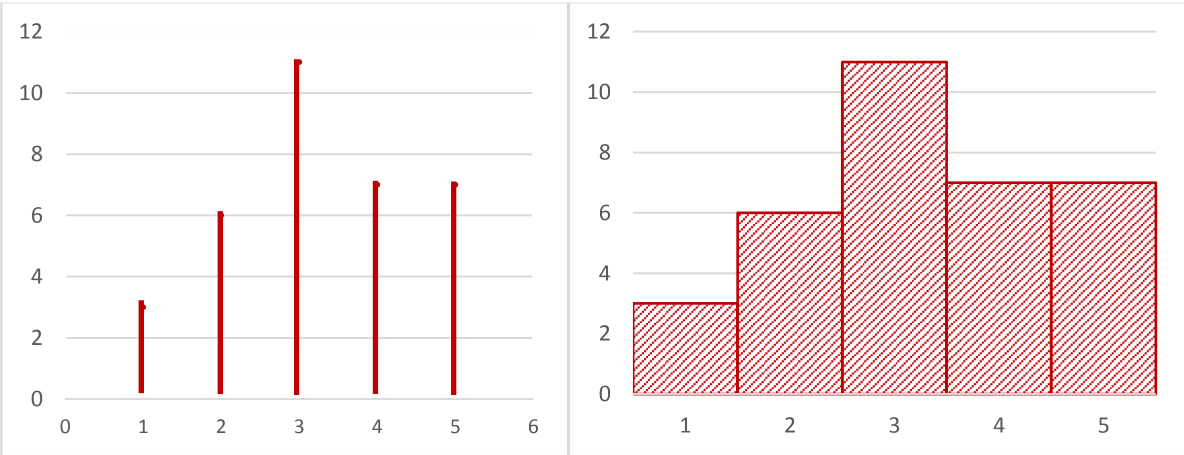
Slika 13. Usporedni prikaz konačnih ocjena položenim dijagramom sa točkama

Slijedeći vrlo ilustrativan prikaz manjeg broja podataka je *stablo-list dijagram*. Znamenke svakog numeričkog podatka podijele se na dva dijela, stablo i list. Vrijednosti stabla poredaju se po veličini u stupcu tabele a uz stablo se na isti način u recima tabele poredaju listovi. Dijagram može biti jednostavan ili dvostruki za dvije vrste obilježja (slika 14).

Broj pod.	**HRVATSKI**		**MATEMATIKA**	***Broj pod.***
4	9422	**1**	02334489	***8***
9	885544311	**2**	002244577778	***12***
10	9964333320	**3**	00222466	***8***
9	988654110	**4**	00168	***5***
2	00	**5**	0	***1***
34				***34***

Slika 14. Prikaz prosječnih ocjena dvostrukim stablo-list dijagramom

Za grafički prikaz distribucije frekvencija manjeg ili većeg broja podataka sa manjim brojem numeričkih obilježja prikladni su *linijski dijagrami* i *površinski dijagrami (histogrami)*. Kod linijskog dijagrama u koordinatnom sustavu spajamo točke (x_i, f_i) i $(x_i, 0)$ (vidjeti (9)) pa duljina dobivenih vertikalnih linija predstavlja frekvencije (slika 15, lijevo). Kod površinskih dijagrama frekvencija je predočena površinom pravokutnika. Kako pravokutnici imaju jednake baze, frekvencije u stvari predočuju njihove visine (slika 15, desno). Uočimo sličnost sa dijagramom jednostavnih stupaca kod kvalitativnih podataka samo što se ovdje između stupaca ne ostavlja razmak.



Slika 15. Prikaz konačnih ocjena iz hrvatskog jezika linijskim i površinskim dijagramom (histogramom)

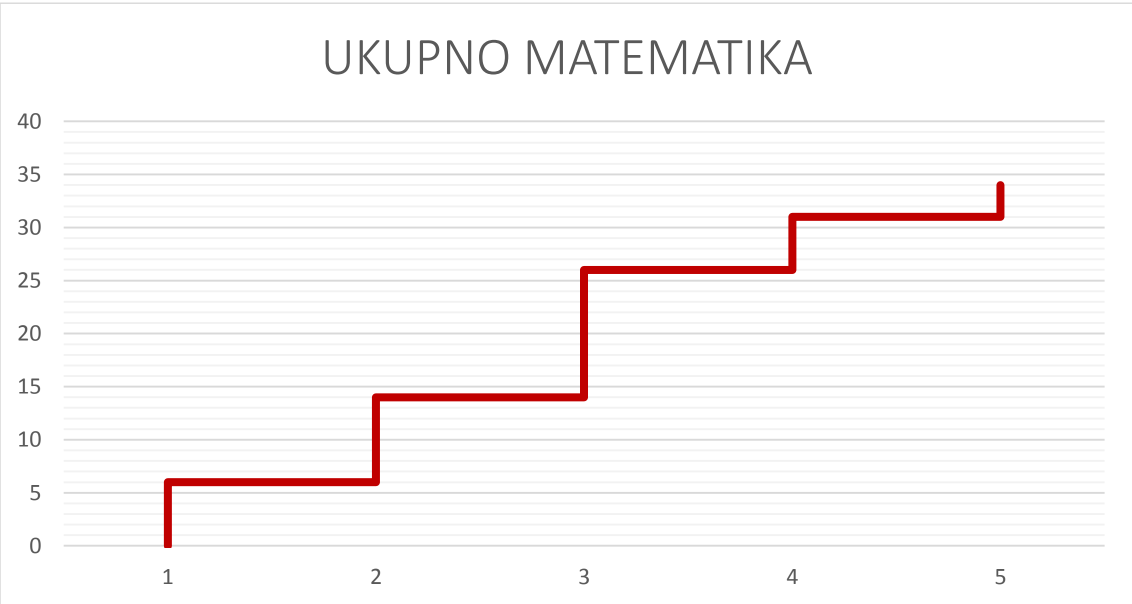
Međusobnim spajanjem točaka (x_i, f_i) po redoslijedu dobivamo linijski dijagram koji se još naziva i *poligon frekvencija* (slika 16).

Slika 16. Usporedni poligon frekvencija

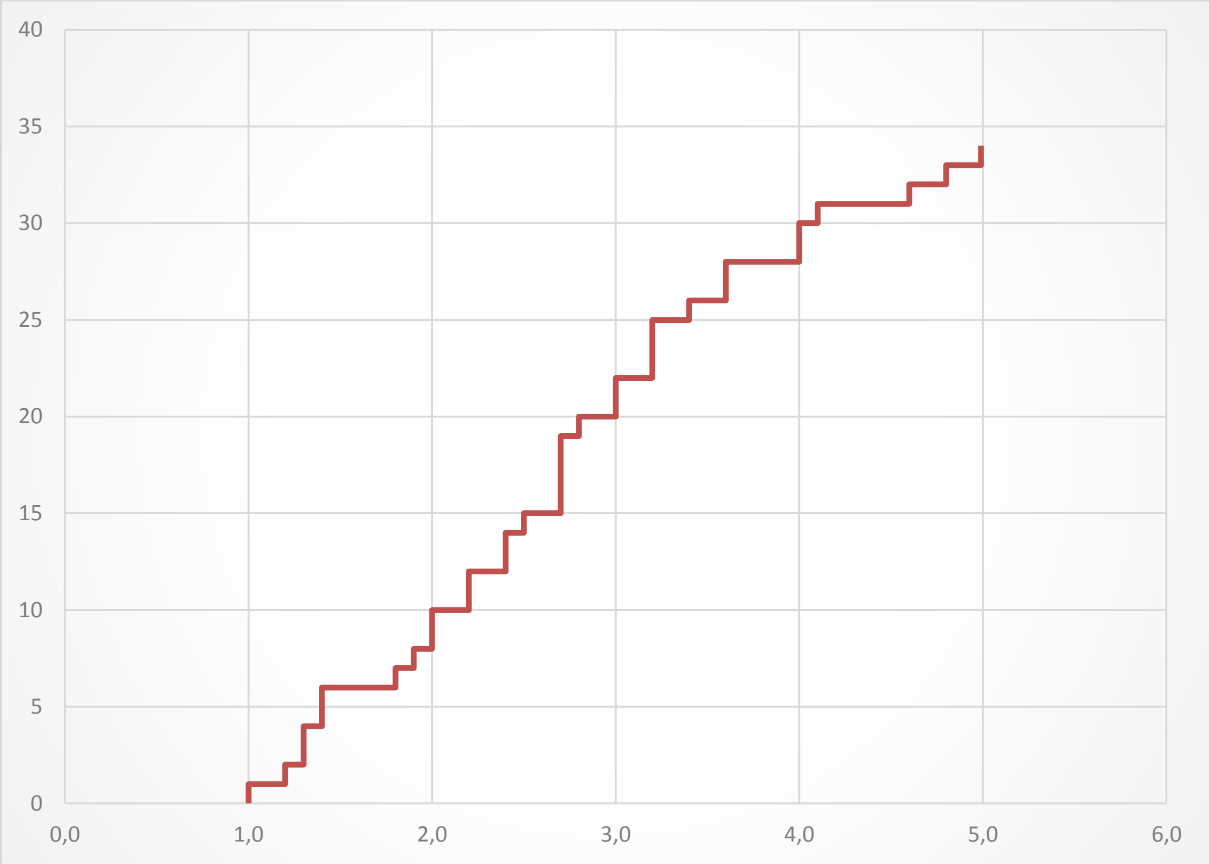
Iz distribucije frekvencija (9) zbrajanjem frekvencija prva dva, zatim prva tri pa onda prva četiri itd. obilježja dobivamo *niz kumulativnih frekvencija*,

$$(x_1, f_1), (x_2, f_1+f_2), (x_3, f_1+f_2+f_3), \dots, (x_k, f_1+f_2+\dots+f_k), \quad x_i < x_{i+1}, \; i=1,2,\dots,k-1, \qquad (10)$$

koji grafički prokazujemo *stepenastim linijskim dijagramom (kumulantom)* (slike 17 i 18). Za što lakšu konstrukciju slike 17 (18) možemo npr. koristiti podatke sa slike 12 (14).



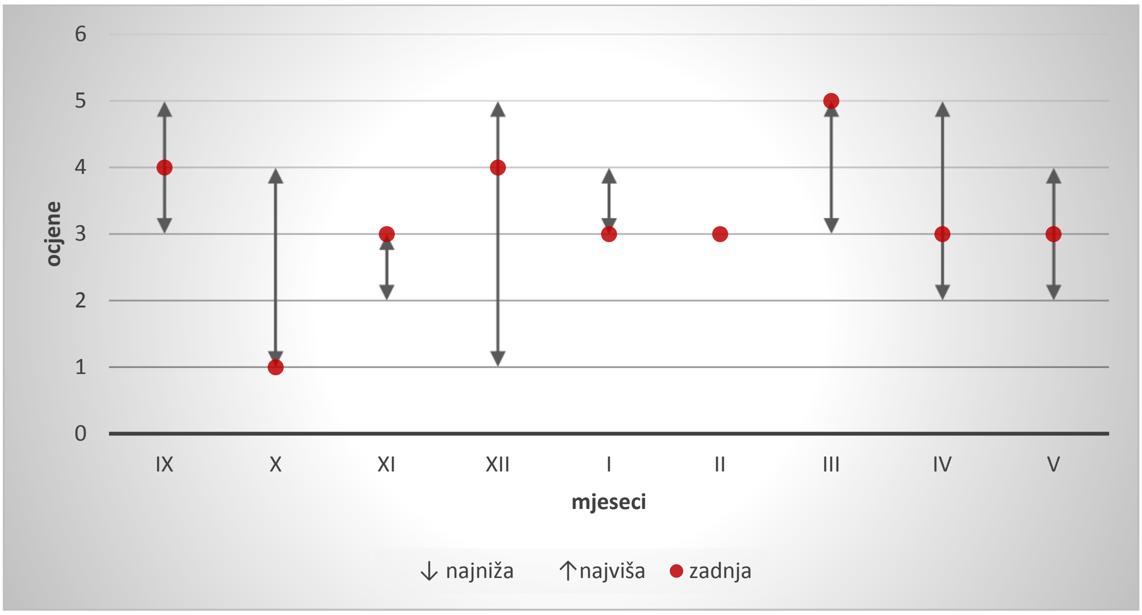
Slika 17. Kumulativa konačnih ocjena iz matematike



Slika 18. Kumulativa prosječnih ocjena iz matematike

Jedan specifični način prikazivanja numeričkih podataka čine *burzovni dijagrami*. Oni služe prvenstveno za prikaz stanja i promjena cijena i prometa dionica na burzama a mogu poslužiti i u druge svrhe. Ovisno o namjeni ima ih više vrsta.

Navodimo primjer burzovnog dijagrama najniža-najviša-zadnja koji za svako promatrano (vremensko) razdoblje prikazuje raspon vrijednosti obilježja: najmanju, najveću i posljednju (najnoviju ili najvažniju) vrijednost. Za primjer na slici 19 prikazan je raspon ocjena koje je učenik dobio iz jednog (ili više) predmeta tokom školske godine. Primijetimo da se iz ovog dijagrama ne vidi broj dobivenih ocjena (frekvencija) već samo raspon. Tako npr. vidimo da je učenik u IX mjesecu dobio bar tri različite ocjene, u X bar dvije itd. dok je u II mjesecu dobivao samo trojke (ne znamo koliko).



Slika 19. Burzovni dijagram ocjena po mjesecima (najniža-najviša-zadnja)

Do sada navedeni dijagrami prikladni su za prikazivanje podataka sa relativno malim brojem obilježja. Ako je broj obilježja (broj oblika ili broj vrijednosti obilježja) velik, dakle ako je numerička varijabla prekidna i ima mnogo vrijednosti ili ako je neprekidna, ovakav prikaz bio bi složen i nepregledan. U tom slučaju podatke grupiramo u razrede. To su disjunktni intervali dobiveni podjelom duljine raspona (intervala) između najmanje i najveće vrijednosti promatrane varijable. Na taj način podaci koji pripadaju jednom razredu nemaju potpuno ista već samo slična svojstva. Tako frekvencija razreda predstavlja broj istih i sličnih vrijednosti varijable.

Na primjer stanovnike nekog mjesta možemo grupirati po starosti u razrede: 0-10 godina, 10-25 godina, 25-45 godina itd. Tako u ovaj zadnji spomenuti razred spadaju i osobe starosti 27 i 30 i 40 godina. Svaki razred ima donju i gornju granicu. Njihov razmak (razlika) je veličina razreda. Razredi mogu biti različite veličine. Gornja granica jednog razreda (ako nije najviši) jednaka je donjoj granici slijedećeg (ako to nije tako razredi se prošite tako da bude), Podaci koji pripadaju toj zajedničkoj granici svrstavaju se u jedan od tih razreda ovisno o prirodi promatrane pojave i cilju analize (obično u viši razred). Svaki razred ima svoju sredinu (srednju vrijednost) koja ima jednaku apsolutnu udaljenost od donje i od gornje granice razreda.

Tako grupirane podatke možemo prikazivati grafički linijskim dijagramima (poligonom frekvencija ili kumulantom) kao i histogramom. Pri tome, u slučaju razreda različitih veličina, frekvencije treba korigirati na fiksnu baznu veličinu. Točke koje koristimo za prikazivanje na linijskim dijagramima su razredne sredine. Pogledajmo primjere.

Primjer 7. Konačna ocjena u tabeli 6 dobivena je zaokruživanjem prosječne ocjene na najbliži cijeli broj. Tako sve prosječne ocjene iz intervala $[1,1.5)$ daju konačnu ocjenu 1, iz intervala $[1.5,2.5)$ ocjenu 2, iz intervala $[2.5,3.5)$ ocjenu 3, iz intervala $[3.5,4.5)$ ocjenu 4 te iz intervala $[4.5,5]$ ocjenu 5. Navedeni intervali su razredi na koje smo podijelili ukupni raspon prosječnih ocjena $[1,5]$. Primijetimo da su prosječne ocjene na zajedničkoj granici dvaju razreda svrstane u viši razred. Primijetimo također da su ocjene 2, 3 i 4 sredine svojih razreda.

Grupiranje podataka iz tabele 6 možemo napraviti i na druge načine ovisno o vrsti i ciljevima analize. Za primjer navodimo još dva načina.

Primjer 8. Grupirajmo prosječne ocjene iz matematike iz tabele 6 na slijedeće načine: (a) u skladu sa dijagramom stablo-list na slici 14, na razrede jednakih veličina: $[1,2)$, $[2,3)$, $[3,4)$ i $[4,5]$ (tabela 7), (b) na razrede različitih veličina: $[1,1.5)$, $[1.5,3.5)$ i $[3.5,5]$ (tabela 8).

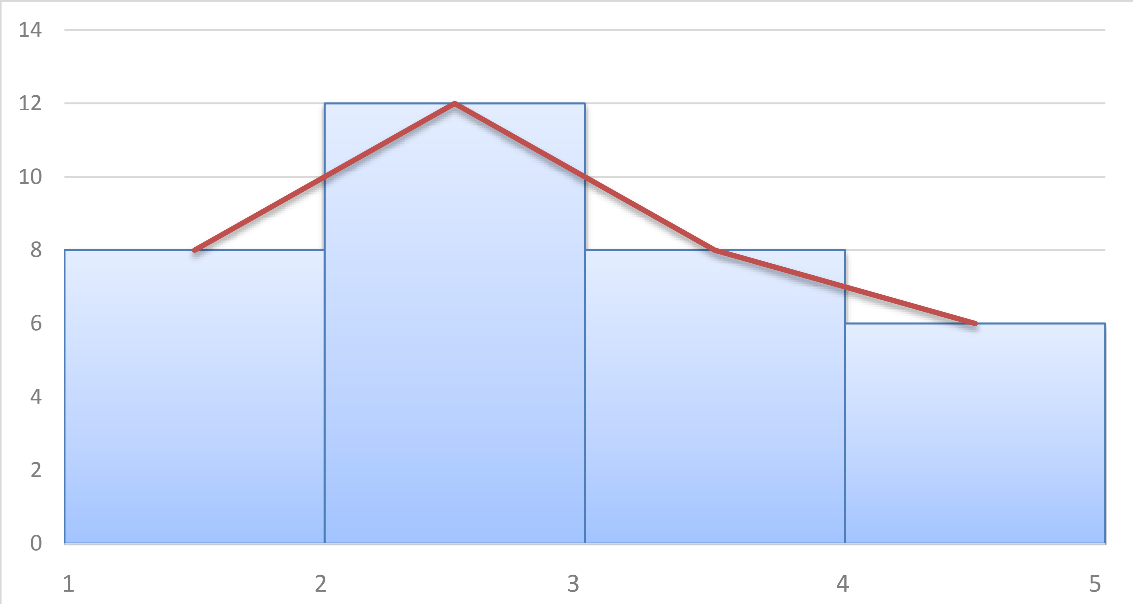
RAZREDI	**Apsolutne frekvencije**	**Relativne frekvencije**	**Kumulativne aps. frekv.**	**Kumulativne rel. frekv.**	**Razredne sredine**
1-2	8	23.53	8	23.53	1.5
2-3	12	35.29	20	58.82	2.5
3-4	8	23.53	28	82.35	3.5
4-5	6	17.65	34	100	4.5
Ukupno	34	100			

Tabela 7. Grupiranje prosječnih ocjena iz matematike u razrede jednakih veličina

RAZREDI	**Veličina razreda**	**Razredni omjeri**	**Apsolutne frekvencije**	**Korigirane aps. frekv.**	**Razredne sredine**
1-1.5	0.5	1	6	6	1.25
1.5-3.5	2	4	20	5	2.5
3.5-5	1.5	3	8	2.67	4.25

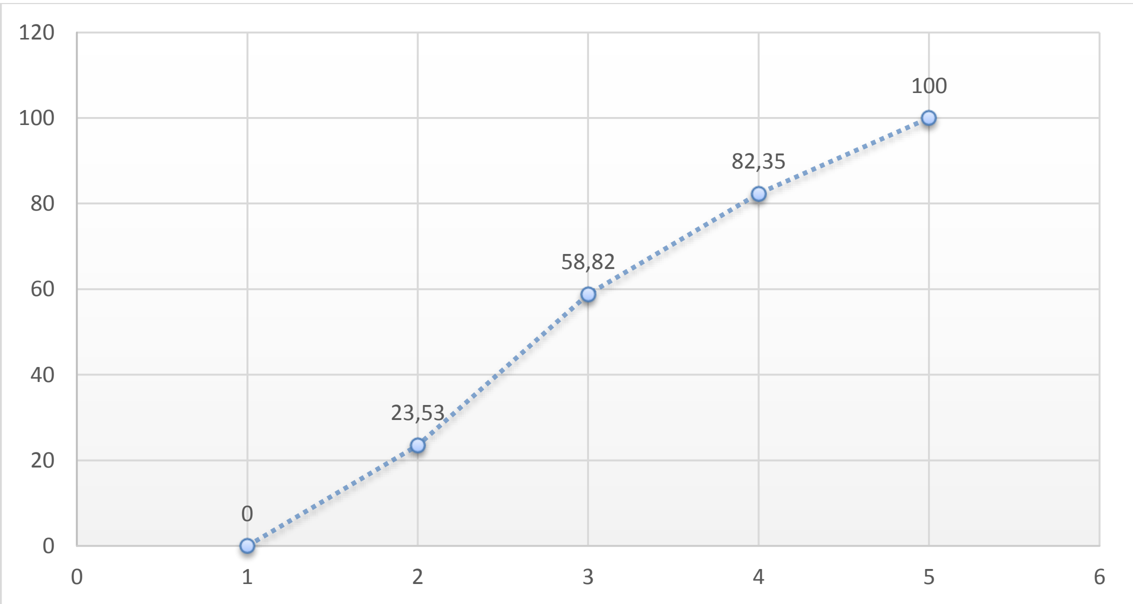
Tabela 8. Grupiranje prosječnih ocjena iz matematike u razrede različitih veličina

Grupirane podatke iz ovih tabela možemo prikazivati linijskim ili površinskim dijagramima koristeći apsolutne ili relativne frekvencije. Tako na slici 20 imamo kombinirani prikaz apsolutnih frekvencija podataka iz tabele 7 histogramom i linijskim dijagramom – poligonom frekvencija gdje se za prikaz koriste razredne sredine.



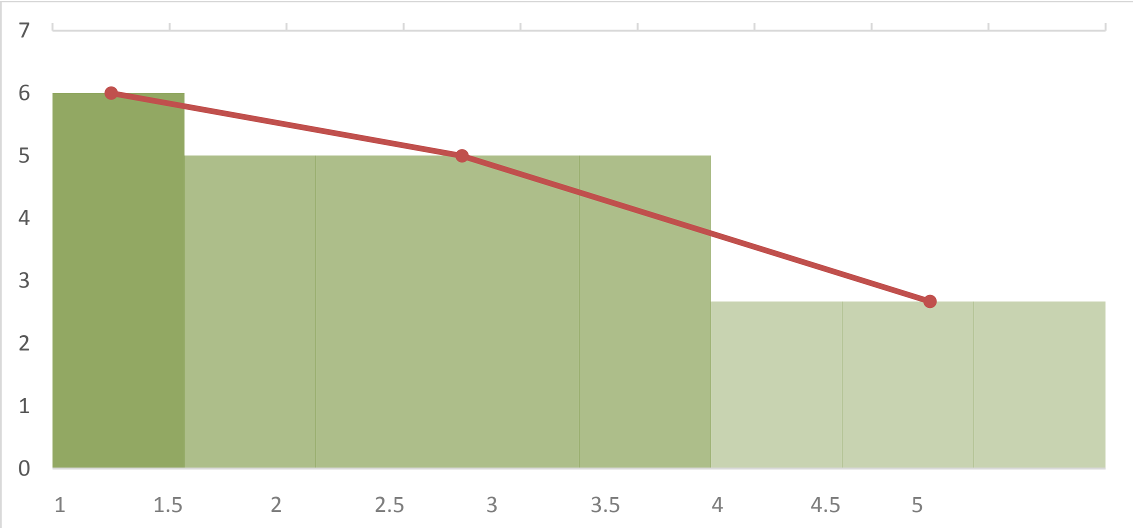
Slika 20. Histogram i poligon apsolutnih frekvencija prosječnih ocjena iz matematike po statističkim razredima jednakih veličina (tabela 7)

Na slici 21 prikazana je kumulanta relativnih frekvencija istih podataka.



Slika 21. Prikaz kumulativnog niza relativnih frekvencija prosječnih ocjena iz matematike po statističkim razredima jednakih veličina (tabela 7)

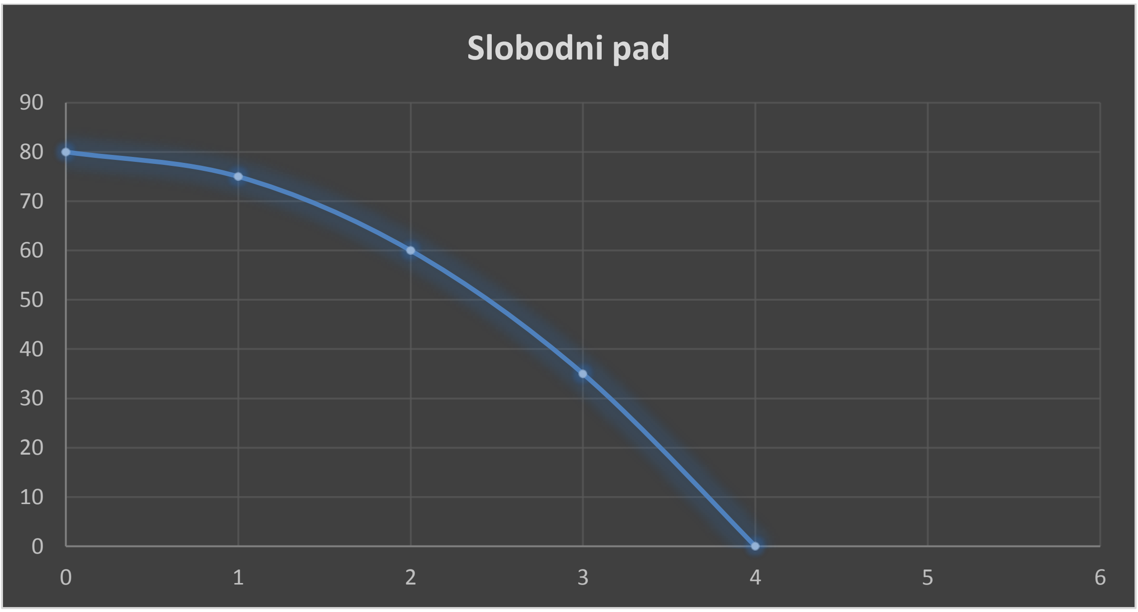
Za grafički prikaz podataka grupiranih u razrede različitih veličina (tabela 8) treba korigirati apsolutne frekvencije na istu (baznu) veličinu. Kako je frekvencija predočena površinom pravokutnika na histogramu, da bi površina ostala ista visinu pravokutnika treba smanjiti (povećati) onoliko puta koliko je osnovica pravokutnika – veličina razreda veća (manja) od bazne osnovice – veličine baznog razreda. U tu svrhu veličinu svakog razreda dijelimo sa veličinom baznog razreda čime dobivamo razredne omjere (tabela 8), Dijeljenjem frekvencije svakog razreda sa pripadnim omjerom dobivamo korigirane frekvencije. Na slici 22 prikazan je histogram i linijski dijagram tako dobivenih korigiranih frekvencija na temelju podataka iz tabele 8 gdje je bazni razred veličine 0.5 (razred 1–1.5). Ako podatke želimo prikazati kumulantom tada za prikaz koristimo originalne (nekorigirane) frekvencije (vidjeti zadatak 1).



Slika 22. Histogram i poligon apsolutnih frekvencija prosječnih ocjena iz matematike po statističkim razredima različitih veličina (tabela 8)

Iz slike 22 možemo uočiti dominaciju broja nedovoljnih ocjena (iz baznog razreda 1–1.5) u odnosu na ostale ocjene čije su frekvencije sada normirane na veličinu baznog razreda.

Sličan postupak grupiranja možemo provesti i za neprekidnu statističku varijablu koja poprima beskonačno mnogo vrijednosti. Ako je poznat njezin analitički oblik možemo je prikazati preciznije, grafom u koordinatnom sustavu.



Slika 23. Grafički prikaz kontinuiranog obilježja

Na slici 23 vidimo prikaz neprekidne varijable udaljenosti u metrima od površine podloge kod slobodnog pada (ordinata) u ovisnosti o vremenu u sekundama od početka pada (analitički oblik $y=80–5x^2$).

Uočimo na kraju da smo kroz primjer 6 analizirali prikazivanje kvantitativnih (numeričkih) obilježja – prosječnih i konačnih ocjena za koje smo na grafikonima koristili os apscisu. Da smo prikazivali ocjene po učenicima tada bi za njih koristili os apscisu pa bi to bio prikaz kvalitativnog obilježja. Napomenimo također da neki načini prikazivanja kvalitativnih obilježja (npr. dvostruki ili višestruki stupci, dijagram indeksa, strukturni likovi i tijela) mogu biti prikladni i za prikazivanje pojedinih kvantitativnih obilježja

ZADACI

1. Prikažite podatke iz tabele 8 kumulantom.
2. Koja od slijedećih numeričkih obilježja su prekidna (diskretna) a koja neprekidna (kontinuirana): prosječna ocjena, konačna ocjena, potrošnja vode, prodani automobili, kamatna stopa, broj nezaposlenih, stopa nezaposlenosti, potrošnja goriva na 100 km, konfekcijski broj, šećer u posudici, kocke šećera na tanjuru ?
3. U jednom restoranu praćen je broj izdanih računa na dnevnoj bazi. Podaci za 27 radnih dana tokom mjeseca prikazani su stablo-list dijagramom.

Stablo	***List***
2	0447
3	245777
4	01335555699
5	024449

   (a) Ovako grupirane podatke prikažite površinskim i linijskim dijagramom.
   (b) Grupirajte podatke u razrede 20-24, 25-34. 35-49, 50-59 te njih prikažite površinskim i linijskim dijagramom.
4. Prikupite podatke o broju učenika po odjeljenjima i vašoj školi te prikažite podatke stablo-list dijagramom. Grupirajte podatke u razrede, npr. 10-19, 20-24, 25-29, 30-39, te ih prikažite grafički proizvoljno odabranim načinom.
5. Grupirajte učenike iz tabele 6 u dvije jednake grupe (npr. 1–17 i 18–34) te usporedite uspjeh (konačne ocjene) iz hrvatskog (koristeći relativne frekvencije) i matematike (koristeći apsolutne frekvencije) pomoću strukturnih likova ili tijela.
6. Robna razmjena RH sa inozemstvom u zadnjih šest godina iznosila je:

Godina	***2009***	***2010***	***2011***	***2012***	***2013***	***2014***
Izvoz	55.2	64.9	71.2	72.2	72.6	79.1
Uvoz	111.8	110.3	121.0	121.5	125.1	130.7

Tabela 9. Vrijednost izvoza i uvoza RH u milijardama kuna

   (a) Prikažite ove podatke dijagramom dvostrukih stupaca.

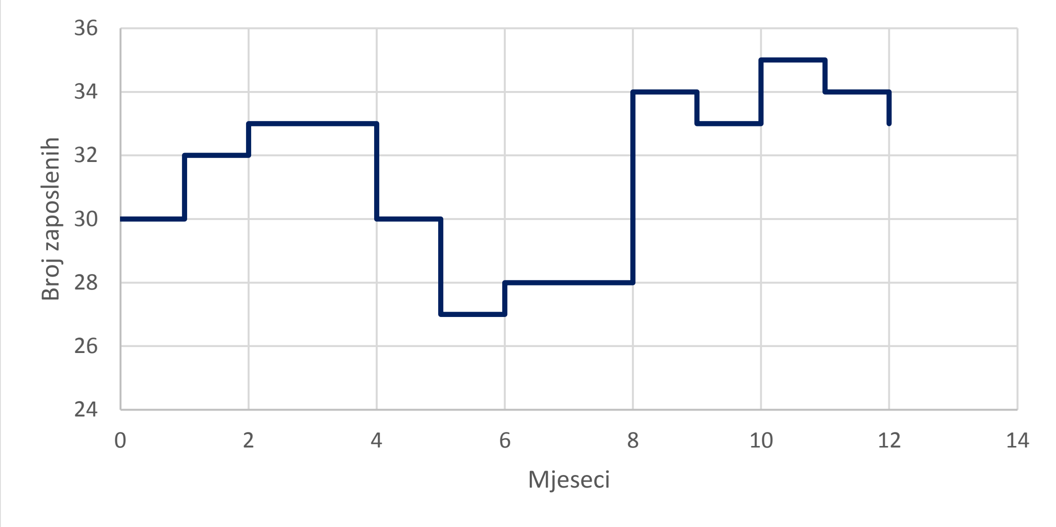
(b) Prikažite ove podatke poligonima frekvencija na istom dijagramu.
(c) Prikažite vanjskotrgovinski deficit linijskim dijagramom i kumulativom.
(d) Prikažite apsolutna odstupanja izvoza i uvoza u odnosu na baznu 2011. godinu na odvojenim dijagramima apsolutnih odstupanja.
(e) Prikažite relativna odstupanja izvoza i uvoza u odnosu na baznu 2011. godinu na odvojenim dijagramima indeksa.

7. Dva građevinska zemljišta površina 350 i 700 m^2 prikazana su na mapi proporcionalnim kvadratima. Za koliko % je stranica većeg kvadrata veća od stranice manjeg kvadrata?
8. Prikupite podatke o visini (cijeli broj cm) i težini (cijeli broj kg) učenika vašeg razreda te ih prikažite dijagramom sa točkama. Zatim ih grupirajte u prikladne razrede (npr. 150-155, 155-160, ... za visinu) i prikažite histogramom.
9. Podaci o strukturi zaposlenih prema dobi u jednoj firmi dani su u tabeli,

Starosna dob	*18-25*	*25-30*	*30-40*	*40-50*	*50-65*
Broj radnika	21	36	57	27	9

Pri čemu zajedničke granice pripadaju gornjim razredima. Prikažite strukturu zaposlenih histogramom relativnih frekvencija i kumulantom.

10. Grupirajte učenike vašeg razreda po udaljenosti od kuće do škole u razrede: 0-1 km, 1-3 km. 3-5 km, 5-10 km, 10-20 km te prikažite dobivene podatke površinskim i linijskim dijagramom.
11. Na odabranom skupu osoba ispitajte koliko prosječno potroše na subotnjem izlasku. Grupirajte podatke u razrede: 0-10 kn, 10-50 kn, 50-100 kn, 100-200 kn itd. po potrebi. Prikažite podatke grafički načinom koji sami odaberete.
12. Da bi se jedan komad proizvoda A proizveo na stroju potrebno je 6 minuta. Proizvodni pogon ima 10 takvih strojeva i narudžbu za 2520 komada proizvoda A. Histogramom i poligonom frekvencija prikažite potrebno vrijeme proizvodnje ako se istovremeno koristi samo 1 stroj, 2, 3, ..., 9 ili svih 10 strojeva. Da li je moguće realizirati proizvodnju unutar 48 sati? Ako da, kako?
13. Cijena nekog proizvoda je 25 kuna. Kumulantom prikažite tok dnevnog prometa ako je kroz dan prodano 10 proizvoda: (a) pojedinačno (jedan po jedan), (b) prva 3 pojedinačno zatim 4 zajedno i ponovo 3 pojedinačno, (c) prva 2 zajedno pa onda 5 zajedno i na kraju 3 zajedno.
14. Ako je 10000 EUR uloženo uz 10% godišnjih dekurzivnih kamata, prikažite stanje glavnice nakon 0, 1, 2, ..., 10 godina histogramom i poligonom frekvencija za: (a) jednostavne, (b) složene kamate.
15. Broj zaposlenih radnika u jednoj firmi po mjesecima tekuće godine prikazan je kumulantom,



Napravite tabelu kadrovskih promjena (zapošljavanja i otpuštanja radnika) po mjesecima.

16. Prikažite strukturnim krugom i polukrugom općenito veličinu kutova slijedećim geometrijskim likovima: (a) paralelogramu, (b) pravokutnom trokutu, (c) pravilnom mnogokutu, (d) proizvoljno odaberite još neke likove.
17. Prikupite podatke o broju članova obitelji učenika vašeg razreda te ih prikažite: (a) dijagramom sa točkama, (b) strukturnim krugom. Napomena: broj članova obitelji je numeričko obilježje a broj obitelji sa istim brojem članova je frekvencija.
18. Frekvencije triju vrijednosti jednog obilježja prikazane su linijskim dijagramom, kumulantom i strukturnim krugom. Dokument sa navedenim prikazima je vrlo oštećen tako da su ostali vidljivi slijedeći podaci. Linijski dijagram počinje sa frekvencijom 250, kumulanta završava sa 800 a srednji isječak strukturnog kruga je 42%. Možemo li rekonstruirati sve tri frekvencije ovog obilježja?
19. Dnevne cijene jedne dionice neke tvrtke po tjednima su bile, I tjedan: 74, 75, 74, 78, 75, 73, 74, II tjedan: 73, 73, 76, 71, 73, 70, 70, III tjedan: 72, 74, 76, 76, 76, 74, 76. Prikažite cijene po tjednima burzovnim dijagramom te zatim učestalost pojedinih cijena dijagramom sa točkama.
20. Knjigu od 146 stranica čitamo tokom 7 dana tako da 2., 3. i 4. dan pročitamo 10 stranica više nego prethodni dan a 5., 6. i 7. dan 10 stranica manje nego prethodni dan. Prikažite histogramom tempo čitanja.
21. Automobil je prvo 4 minute vozio prosječno 60 km/h, zatim je 3 minute stajao pa 2 minute vozio 90 km/h i onda stajao još 2 minute. Nakon toga se okrenuo i vozio natrag do polazne točke 70 km/h. Grafički prikažite put u ovisnosti o vremenu te očitajte prijeđeni put i ukupno vrijeme putovanja.

RJEŠENJA

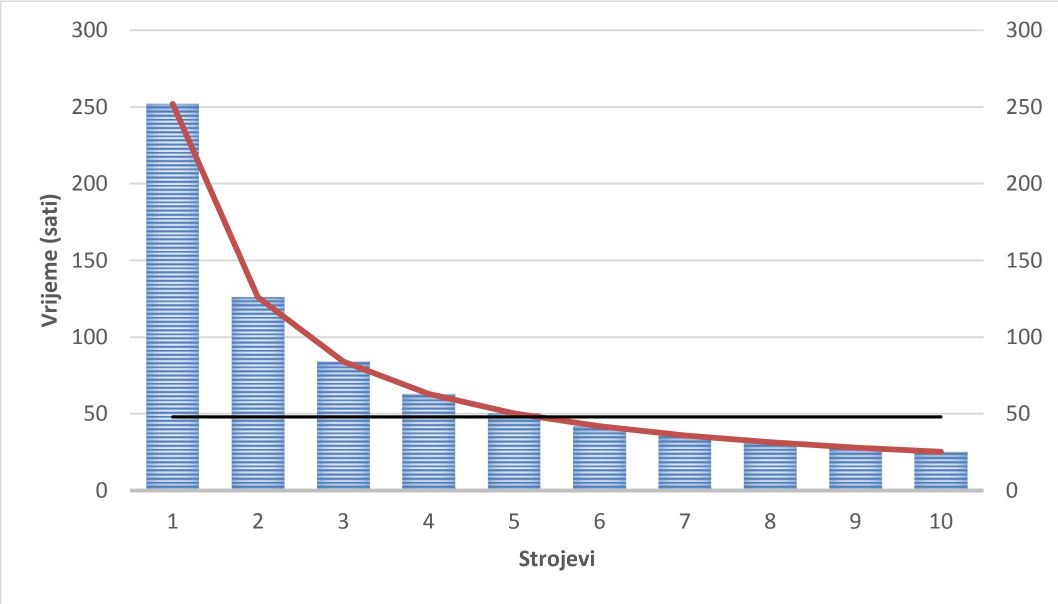
1. Za prikaz koristimo zadane (nekorigirane) apsolutne frekvencije. Kumulanta spaja točke (1, 0), (1.5, 6), (3.5, 26) i (5, 34).

2. Prekidna obilježja su: konačna ocjena, prodani automobili, broj nezaposlenih, konfekcijski broj i kocke šećera na tanjuru a ostala su neprekidna.
3. (a) Razredi su 20-29, 30-39, 40-49 i 50-59, veličine svih razreda su 10, razredne sredine 24.5, 34.5, 44.5 i 54.5 i frekvencije 4, 6, 11 i 6. Kod linijskog dijagrama spajamo točke (24.5, 4), (34.5, 6), (44.5, 11) i (54.5, 6). (b) Razredne sredine su 22, 29.5, 42 i 54.5, veličine razreda 5, 10, 15 i 10, razredni omjeri 1, 2, 3 i 2, frekvencije 3, 3, 15 i 6, pa su korigirane frekvencije 3, 1.5, 5 i 3. Na dijagramima prikazujemo korigirane frekvencije. Primijetimo da se radi o diskretnom obilježju (cijeli brojevi) pa npr. razred 20-29 ima veličinu 10 a da je obilježje neprekidno veličina bi bila 9.
4. Koristite stvarne podatke.
5. Za hrvatski jezik za grupu 1–17: 1 (11.76%), 2 (5.88%), 3 (41.18%), 4 (17.65%), 5 (23.53%), te za grupu 18–34: 1 (5.88%), 2 (29.41%), 3 (23.53%), 4 (23.53%), 5 (17.65%). Za matematiku za grupu 1–17: 1 (3), 2 (6), 3 (5), 4 (2), 5 (1), te za grupu 18–34: 1 (3), 2 (2), 3 (7), 4 (3), 5 (2). Za prikazivanje možemo odabrati strukturni krug, valjak, pravokutnik, prizmu i sl. (slike 6 i 7).
6. (a) Analogno prvom dijagramu na slici 3. (b) Analogno slici 16.

   (c) Analogno dijagramu na slici 15 lijevo prikazujemo podatke: 56.6, 45.4, 49.8, 49.3, 52.5, 51.6, te analogno slici 17 ili 18 podatke: 56.6, 102, 151.8, 201.1, 253.6, 305.2.

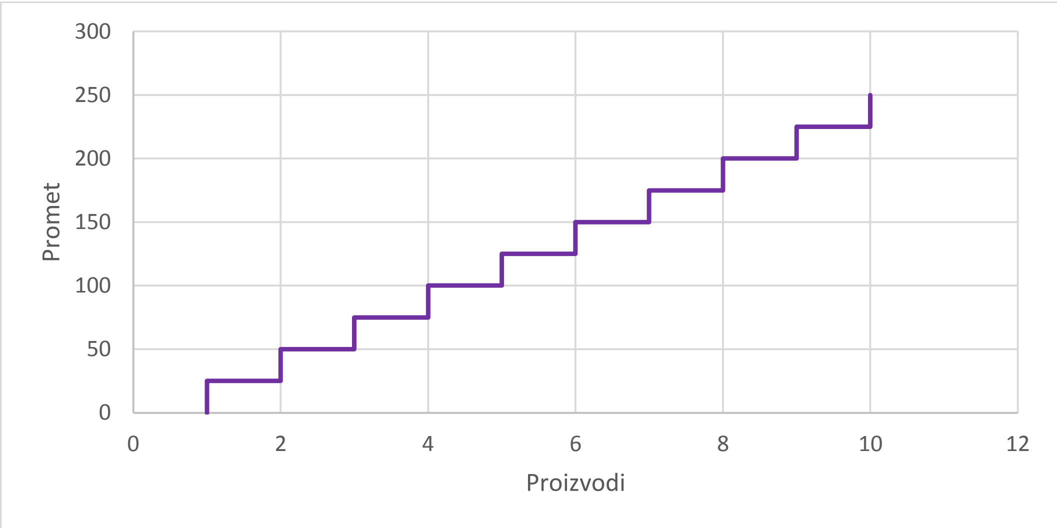
   (d) Analogno slici 5 (sa ili bez razmaka između stupaca) prikazujemo podatke za izvoz: –16, –6.3, 0, 1, 1.4, 7.9, te za uvoz: –9.2, –10.7, 0, 0.5, 4.1, 9.7 milijardi kuna. Pri tome na ordinati kod 0 možemo dodati u zagradi (71.2) za izvoz a (121.0) za uvoz ili u legendi iznad ili ispod dijagrama navesti: Bazna vrijednost (0) iznosi ... .

   (e) Kao pod (d) prikazujemo podatke za izvoz: 77.53, 91.15, 100, 101.4, 102, 111.1, te za uvoz: 92.4, 91.2, 100, 100.4, 103.4, 108.
7. $P_2 / P_1 = 700/350 = 2 = a_2^2 / a_1^2 \;\Rightarrow\; a_2 = a_1\sqrt{2} \approx 1.4142 a_1 \;\Rightarrow$ za približno 41.42%.
8. Koristite stvarne podatke.
9. Kako je ukupno 150 zaposlenih, relativne frekvencije su (u %) 14, 24, 38, 18 i 6, veličine razreda su 7, 5, 10, 10 i 15, razredni omjeri 0.7, 0.5, 1, 1 i 1.5, pa su korigirane relativne frekvencije 20, 48, 38, 18 i 4. Za prikaz histogramom koristimo korigirane frekvencije a za kumulantu originalne. Kumulanta spaja točke (18, 0), (25, 14), (30, 38), (40, 76), (50, 94) i (65, 100).
10. Koristite prikupljene podatke.
11. Koristite prikupljene podatke.
12. Potrebno vrijeme je $(2520 \cdot 6) : 60 = 252$ sata, pa je: $\mathit{broj\ strojeva} \times \mathit{vrijeme} = 252$. Imamo prikaz zavisnosti dviju veličina konstantnog umnoška – dijagram obrnute proporcionalnosti (recipročni dijagram).

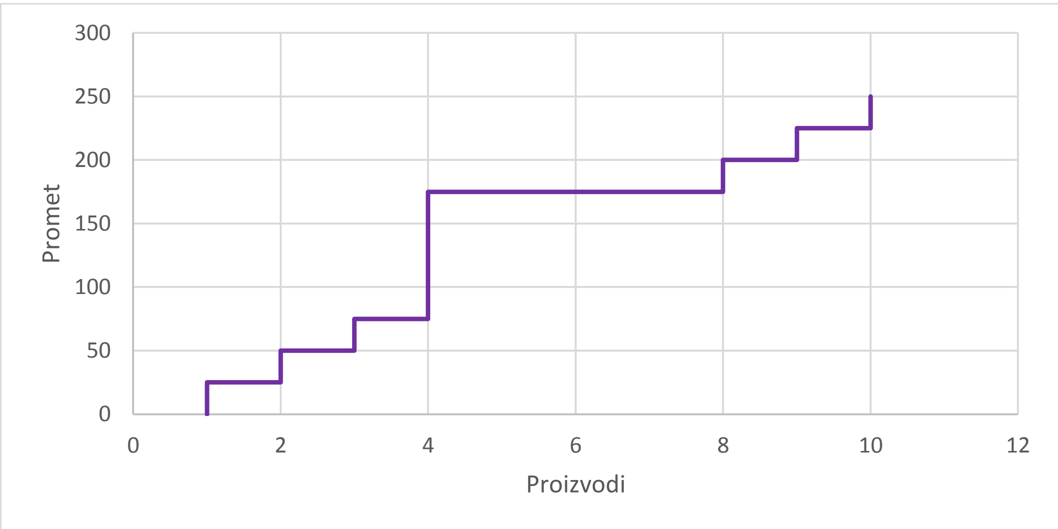


Vidimo da se proizvodnja može realizirati unutar 48 sati korištenjem barem 6 strojeva.

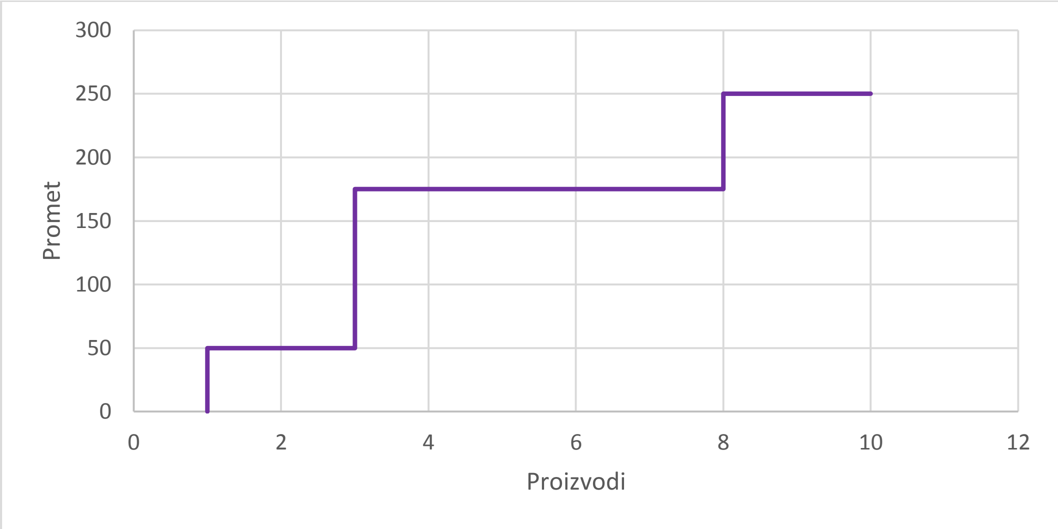
13. (a)



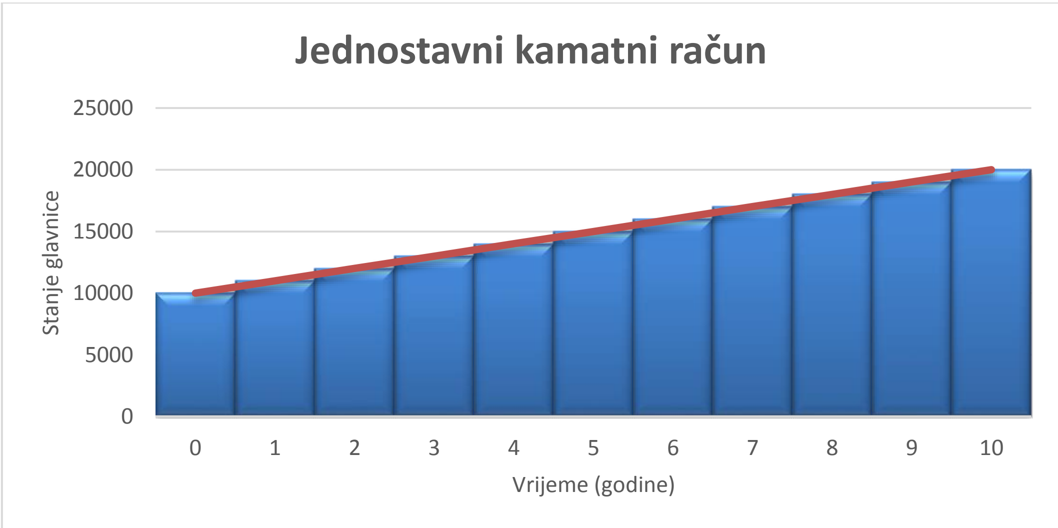
(b)



(c)



14. (a) Prikazujemo podatke $C_k = C_0(1+pk) = 10000 \cdot (1+0.1k)$ za $k = 0,1,2,\ldots,10$ što daje linearni ili afini dijagram.

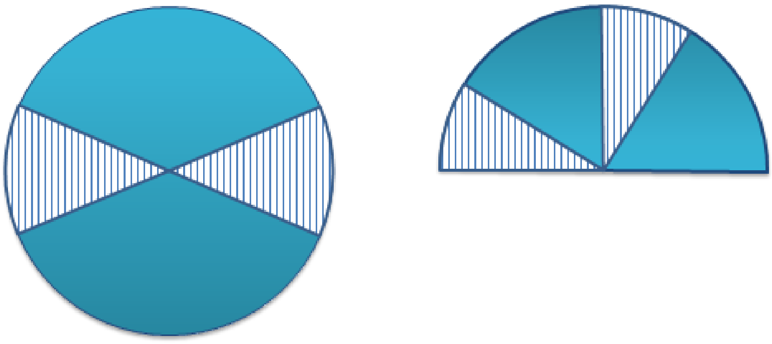


(b) Prikazujemo podatke $C_k = C_0(1+p)^k = 10000 \cdot 1.1^k$ za $k = 0,1,2,\ldots,10$ što daje eksponencijalni dijagram.

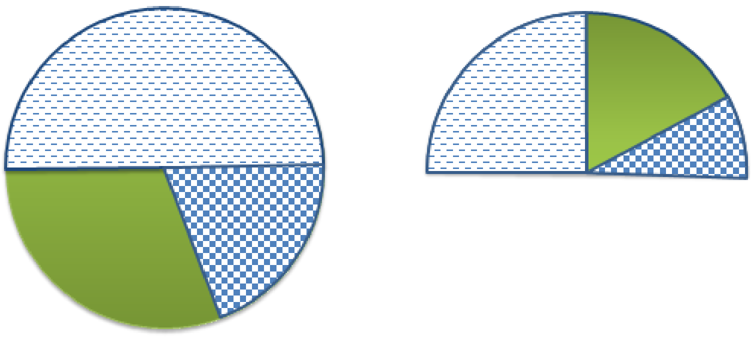
15. Iz dijagrama možemo očitati razliku između broja novozaposlenih i otpuštenih radnika po mjesecima. Ako su mjesečno radnici ili zapošljavani ili otpuštani, imamo

Mjeseci	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*Σ*
Zaposleno	2	1	-	-	-	1	-	6	-	2	-	-	12
Otpušteno	-	-	-	3	3	-	-	-	1	-	1	1	9

16. (a)



(b)



(c) Strukturni krug (polukrug) podijelimo na $n = 3, 4, 5, 6, \ldots$ jednakih isječaka.

17. Koristite stvarne podatke. (a) Umjesto točaka možete koristiti bilo koje znakove (zvjezdice, kvadratići, slova i sl.). U istom dijagramu (čak i za istu vrijednost

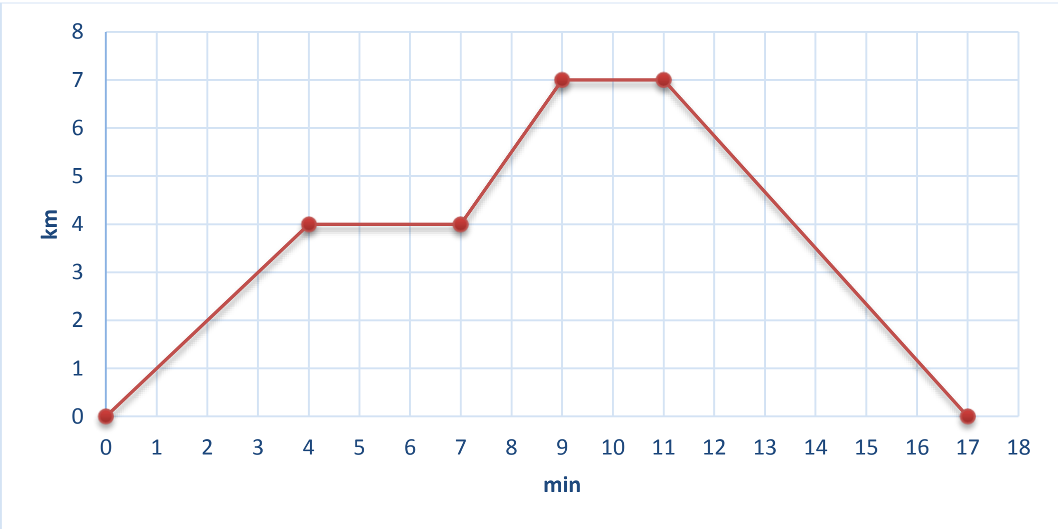
obilježja) znakovi mogu biti različiti, npr. zvjezdice za djevojke (tj. njihove obitelji), kvadratići za dečke itd.

18. Imamo $f_1 = 250$, $f_1 + f_2 + f_3 = 800$, $f_2 / (f_1 + f_2 + f_3) = 0.42$ iz čega slijedi $f_1 = 250$, $f_2 = 336$, $f_3 = 214$.
19. Burzovni dijagram najniža-najviša-zadnja prikazujemo analogno slici 19. U dijagramu sa točkama prikazujemo frekvencije (broj dana) za svaku cijenu: 70(2), 71(1), 72(1), 73(4), 74(5), 75(2), 76(5), 77(0), 78(1).
20. Ako je x broj pročitanih stranica u prvom danu, tada imamo

$$x + (x+10) + (x+20) + (x+30) + (x+20) + (x+10) + x = 146$$

iz čega slijedi $x = 8$. Histogramom prikazujemo broj pročitanih stranica redom po danima: 1. 8, 2. 18, 3. 28, 4. 38, 5. 28, 6. 18, 7. 8.

21. Brzine pretvaramo u km/min: 60 km/h = 1 km/min, 90 km/h = 1.5 km/min. Prve 4 minute prelazi 4 km, 3 minute stoji, 2 minute prelazi 3 km (ukupno 7 km), 2 minute stoji, vraća se natrag 7 km brzinom 70 km/h za što mu treba 1/10 sata = 6 min. Ukupno prelazi 7+7=14 km a put traje 4+3+2+2+6=17 minuta. Grafički prikaz puta:



METODE ANALIZIRANJA PODATAKA

Grafička analiza podataka daje nam osim vizualnog uvida i neke kvantitativne pokazatelje o svojstvima i strukturi promatrane pojave. U ovom poglavlju upoznajemo se sa parametrima za temeljitiju kvantitativnu analizu podataka. Najznačajniji su srednje vrijednosti (mjere centralne tendencije) te mjere raspršenosti (disperzije), distribucije i koncentracije. Osim za kvantitativne podatke neki od njih mogu se koristiti i za analizu kvalitativnih podataka. Svaki parametar u određenom smislu sažima informacije više različitih podataka u jednu.

SREDNJE VRIJEDNOSTI

Srednja vrijednost je konstanta koja predočuje niz (skup) međusobno različitih vrijednosti. Ona time sažima vrijednosti svih članova niza u jednu, prosječnu vrijednost. Članovi niza su, u određenom smislu, ravnomjerno raspoređeni oko srednje vrijednosti. Drugim riječima, ako zadani niz zamijenimo sa konstantnim nizom,

$$x_1, x_2, \dots, x_n \quad \to \quad x, x, \dots, x \ \ (n \text{ članova}),$$

gdje je x srednja vrijednost a n broj članova zadanog niza, tada oba niza imaju isto, unaprijed dogovoreno svojstvo. Dakle, srednja vrijednost x ovisi o traženom svojstvu (zahtjevu). Za numeričke nizove možemo postaviti slijedeće zahtjeve.

Ako tražimo da zbroj članova oba niza bude isti, $x_1 + x_2 + \dots + x_n = x + x + \dots + x = nx$, dobivamo *aritmetičku sredinu* (oznaka $x = \overline{x}$ ili $x = A$). Ako tražimo da umnožak članova oba niza bude isti, $x_1 \cdot x_2 \cdot \dots \cdot x_n = x \cdot x \cdot \dots \cdot x = x^n$, dobivamo *geometrijsku sredinu* (oznaka $x = G$). Ako postavimo zahtjev da je zbroj recipročnih vrijednosti članova oba niza isti. $\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{1}{x} + \frac{1}{x} + \dots + \frac{1}{x} = \frac{n}{x}$, dobivamo *harmonijsku sredinu* (oznaka $x = H$). Dakle, imamo slijedeću definiciju

DEFINICIJA 2. ZA NUMERIČKI NIZ (SKUP) $x_1, x_2, \dots, x_{n-1}, x_n$ JE

$$A = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(aritmetička sredina)},$$

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad \text{(geometrijska sredina)},$$

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad \text{(harmonijska sredina)}.$$

Istaknimo ponovo da je

$$\sum_{i=1}^{n} x_i = nA, \qquad \prod_{i=1}^{n} x_i = G^n, \qquad \sum_{i=1}^{n} \frac{1}{x_i} = \frac{n}{H}.$$

Iz navedenih definicija, odnosno postavljenih zahtjeva koji su rezultirali ovakvim definicijama, proizlaze osnovna svojstva navedenih sredina. Tako za aritmetičku sredinu, ako u izrazu $x_1 + x_2 + \dots + x_n = A + A + \dots + A$ prvo sve članove prebacimo na lijevu stranu a zatim podijelimo sa A, dobivamo

$$\sum_{i=1}^{n}(x_i - A) = 0, \qquad \sum_{i=1}^{n}\frac{x_i - A}{A} = \sum_{i=1}^{n}\left(\frac{x_i}{A} - 1\right) = 0. \tag{11}$$

Vidimo da je zbroj apsolutnih ali i relativnih odstupanja svih članova niza od aritmetičke sredine jednak nuli. To znači da su u navedenom smislu članovi niza ravnomjerno (simetrično) raspoređeni oko aritmetičke sredine. Ako slično postupimo u izrazu iz kojeg je dobivena harmonijska sredina, $1/x_1 + 1/x_2 + \ldots + 1/x_n = 1/H + 1/H + \ldots + 1/H$, dobivamo

$$\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{H}\right) = 0, \qquad \sum_{i=1}^{n}\frac{\frac{1}{x_i} - \frac{1}{H}}{\frac{1}{H}} = \sum_{i=1}^{n}\frac{H - x_i}{x_i} = \sum_{i=1}^{n}\left(\frac{H}{x_i} - 1\right) = 0.$$

Dakle, zbroj apsolutnih ali i relativnih odstupanja recipročnih vrijednosti svih članova niza od recipročne vrijednosti harmonijske sredine jednak je nuli. Slično tome iz izraza $x_1 \cdot x_2 \cdot \ldots \cdot x_n = G \cdot G \cdot \ldots \cdot G$ slijedi

$$\prod_{i=1}^{n}\frac{x_i}{G} = 1 \quad \text{ili} \quad \prod_{i=1}^{n}\frac{x_i}{G} - 1 = 0,$$

što znači da je relativno odstupanje niza $x_1, x_2, \ldots, x_n$ kao cjeline od konstantnog niza $G, G, \ldots, G$ jednako nuli. Navedimo još neka važna svojstva ovih sredina.

Aritmetička sredina je vrijednost od koje je zbroj kvadrata odstupanja svih članova niza minimalan (najmanji mogući),

$$\sum_{i=1}^{n}(x_i - A)^2 \quad \text{je minimalno.} \tag{12}$$

Naime, ako uzmemo neku drugu vrijednost $B \neq A$, imamo

$$\sum_{i=1}^{n}(x_i - B)^2 = \sum_{i=1}^{n}\left[(x_i - A) + (A - B)\right]^2 = \sum_{i=1}^{n}(x_i - A)^2 + 2(A - B)\sum_{i=1}^{n}(x_i - A) + \sum_{i=1}^{n}(A - B)^2 =$$
$$= \sum_{i=1}^{n}(x_i - A)^2 + 2(A - B)\cdot 0 + n(A - B)^2 > \sum_{i=1}^{n}(x_i - A)^2.$$

Za geometrijsku srednju vrijednost logaritmiranjem dobivamo

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \ldots + \log x_n) = \frac{1}{n}\sum_{i=1}^{n}\log x_i.$$

Dakle, logaritam geometrijske sredine numeričkog niza je aritmetička sredina logaritama njegovih članova.

Ako u promatranom nizu ima jednakih članova, možemo ih grupirati kao distribuciju frekvencija,

$$(x_1, f_1), (x_2, f_2), ..., (x_k, f_k), \qquad f_1 + f_2 + ... + f_k = n, \tag{13}$$

gdje je k broj različitih članova a n ukupan broj članova niza. U tom slučaju izraze za A, G, H možemo pisati i na slijedeći način:

$$A = \frac{f_1x_1 + f_2x_2 + ... + f_kx_k}{f_1 + f_2 + ... + f_k} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \sum_{i=1}^{k} p_i x_i ,$$

$$G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot ... \cdot x_k^{f_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}} = \prod_{i=1}^{k} x_i^{p_i} ,$$

$$H = \frac{f_1 + f_2 + ... + f_k}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + ... + \frac{f_k}{x_k}} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{1}{\sum_{i=1}^{k} \frac{p_i}{x_i}}, \quad p_i = \frac{f_i}{n}, \ i = 1, 2, ..., k.$$

Ovako izražene srednje vrijednosti nazivaju se *ponderirane (vagane)* pri čemu su apsolutne frekvencije f_i odnosno relativne p_i ponderi (težine) za vrijednosti x_i. Ako su podaci uređeni kao distribucija frekvencija sa razredima u navedenim formulama za vrijednosti x_i uzimaju se razredne sredine.

Sve tri navedene srednje vrijednosti izražene su u mjernim jedinicama pojave (varijable) za koju se računaju. Budući da su dobivene korištenjem svih članova niza spadaju među *potpune srednje vrijednosti*. Sredine ne moraju biti jednake niti jednom članu niza. Koju sredinu ćemo primijeniti u konkretnom slučaju ovisi o prirodi i vrsti obilježja kao i o svrsi i cilju analize. Aritmetička sredina je prikladna za obilježja aditivnih svojstava (gdje se vrijednosti obilježja dobivaju uglavnom zbrajanjem ili oduzimanjem) i najčešće se primjenjuje u praksi. Tako je koristimo za određivanje velikog broja prosječnih veličina kao što su: prosječni troškovi, prihodi, izdaci, prosječna plaća, brzina, temperatura, potrošnja, cijena itd. Služi i za određivanje normi za pojedine aktivnosti, normativa i parametara za različita mjerenja (npr. u medicini: tlak, šećer, kolesterol, idealna težina, visina i sl.). Na taj je način aritmetička sredina osnova za čitave sustave izračuna (npr. račun smjese, diobe i sl.) Geometrijska sredina je prikladna za multiplikativna svojstva (kod kojih su množenje i dijeljenje glavne operacije među vrijednostima obilježja) a harmonijska za obilježja gdje se promatraju recipročne vrijednosti varijabli. Osim toga za obilježja koja poprimaju negativne vrijednosti ili nulu izračunavanje geometrijske i/ili harmonijske sredine ponekad nije moguće (matematički izraz za te sredine nije definiran – dijeljenje sa nulom, parni korijen iz negativnog broja). Za numeričke nizove sa pozitivnim članovima, a što je u praksi i najčešći slučaj, sve tri sredine su uvijek dobro definirane. Za njihov međusobni odnos tada vrijedi slijedeća relacija

$$\min\{x_1, x_2, ..., x_n\} \le H \le G \le A \le \max\{x_1, x_2, ..., x_n\}.$$

Primjer 9. Odredimo srednje vrijednosti konačnih ocjena iz hrvatskog jezika iz primjera 6 (tabele 6) te provjerimo njihov međusobni odnos.

Možemo koristiti sliku 12, 13, 14, 15 ili 16 gdje je niz već uređen po veličini i grupiran pa vidimo frekvencije ocjena. Imamo

$$A=\frac{3\cdot 1+6\cdot 2+11\cdot 3+7\cdot 4+7\cdot 5}{3+6+11+7+7}=\frac{111}{34}=3.2647...$$

$$G=\sqrt[34]{1^3\cdot 2^6\cdot 3^{11}\cdot 4^7\cdot 5^7}=2.9877...$$

$$H=\frac{34}{\frac{3}{1}+\frac{6}{2}+\frac{11}{3}+\frac{7}{4}+\frac{7}{5}}=2.6527...$$

Vidimo da vrijedi $A\geq G\geq H$. Razlike među sredinama mogu biti male ali i dosta značajne (kao u ovom primjeru) ovisno o distribuciji podataka. Kako se ovdje radi o aditivnom obilježju, u praksi se koristi aritmetička sredina. Napomenimo da su i prosječne ocjene navedene u tabeli 6 aritmetičke sredine pojedinačnih ocjena tokom godine.

Osim navedenih potpunih srednjih vrijednosti imamo i *položajne srednje vrijednosti*. To može biti jedan član niza ili se određuju na temelju jednog dijela članova niza. Najpoznatije su *mod i kvantili (medijan, kvartili, decili i percentili)*. Općenito kvantili su vrijednosti koje dijele niz na zadani broj dijelova sa jednakim brojem članova.

DEFINICIJA 3. U STATISTIČKOM NIZU MOD JE NAJČEŠĆA VRIJEDNOST (VRIJEDNOST NAJVEĆE FREKVENCIJE). MEDIJAN JE VRIJEDNOST KOJA DIJELI NIZ NA 2 DIJELA SA JEDNAKIM BROJEM ČLANOVA, KVARTILI GA DIJELE NA 4, DECILI NA 10 A PERCENTILI NA 100.

Vidimo da su navedene sredine definirane za kvantitativna ali mogu biti i za kvalitativna obilježja (važno je imati uređene podatke po nekom redoslijedu – niz). Mod (oznaka Mo) je definiran ako niz ima bar dvije iste vrijednosti. Ako je mod samo jedan imamo *unimodalni niz*. Ako ima više vrijednosti iste najveće frekvencije imamo i više modalnih vrijednosti (*multimodalni niz*). Mod (ako postoji) je uvijek neki od članova niza dok ostale sredine ne moraju biti.

U nizu sa n članova, $x_1,x_2,...,x_n$ medijan (oznaka Me) se određuje ovisno o tome da li je $n/2$ cijeli broj ili nije (da li je n paran ili neparan). Neka je $r=\text{int}(n/2)$ cjelobrojni dio broja $n/2$ tj. najveći cijeli broj sadržan u $n/2$ (oznaka int = integer). Imamo

$$r=\frac{n}{2}\ (n\ \text{paran})\Rightarrow Me=\frac{x_r+x_{r+1}}{2},\qquad r\neq\frac{n}{2}\ (n\ \text{neparan})\Rightarrow Me=x_{r+1}.$$

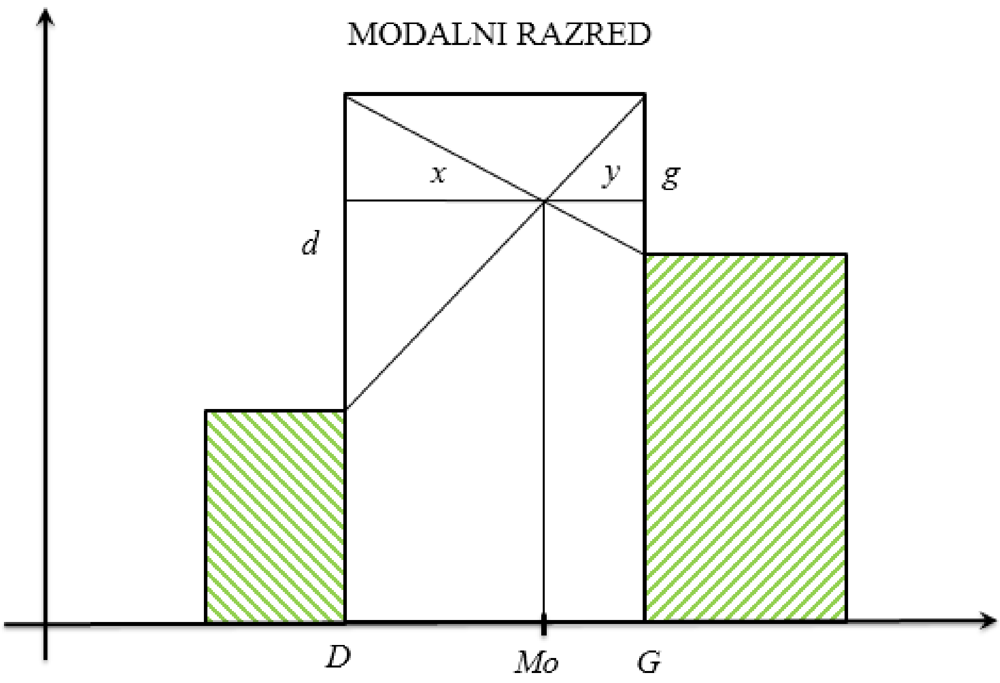
Na analogni način određuju se kvartili Q_1,Q_2,Q_3. Za prvi kvartil Q_1 uzimamo $r=\text{int}(n/4)$ a za treći Q_3 uzimamo $n=\text{int}(3n/4)$ dok je drugi kvartil u stvari medijan, $Q_2=Me$. Decili

$D_i,\ i=1,2,\ldots,9$ se definiraju analogno uzimajući $r=\text{int}(in/10)$ a percentili $P_i,\ i=1,2,\ldots,99$ uzimajući $r=\text{int}(in/100)$. Decili a osobito percentili se koriste za nizove sa puno članova. Za lakše određivanje ovih sredina kod grupiranih nizova dobro je koristiti kumulativne frekvencije.

Primjer 10. Za konačne ocjene iz hrvatskog jezika iz primjera 6 (tabele 6) odredimo mod, medijan i kvartile.

Kako je $n=34$ imamo $n/2=17=\text{int}(34/2)=r$ pa je medijan sredina između 17. i 18. člana niza 1,1,1,2,2,2,2,2,**2**,3,3,3,3,3,3,3,**3**,**3**,3,3,4,4,4,4,4,**4**,4,5,5,5,5,5,5,5 i iznosi (**3**+**3**)/2=3 (ali nije član niza). Nadalje $n/4=8.5\neq\text{int}(8.5)=8=r \Rightarrow r+1=9$ pa je prvi (donji) kvartil Q_1 deveti član niza (**2**). Slično tome $3n/4=25.5\neq\text{int}(25.5)=25=r \Rightarrow r+1=26$ pa je treći (gornji) kvartil Q_3 dvadeset šesti član niza (**4**). Mod je ocjena 3 koja ima najveću frekvenciju (11). Primijetimo iz primjera da su ove srednje vrijednosti određene svojim položajem u nizu podataka. Ako na primjer na proizvoljan način povećavamo vrijednosti članova niza desno od medijana ili smanjujemo lijevo od njega, medijan se ne mijenja. Slično i za ostale kvantile (kvartile, decile i percentile).

Ako su podaci grupirani u razrede tada u pravilu nisu poznate frekvencije pojedinačnih obilježja veće samo čitavih razreda. U tom slučaju razred sa najvećom frekvencijom nazivamo *modalni razred*. Unutar njega mod se definira aproksimativno kao podatak najčešće frekvencije u odnosu na susjedne razrede uz pretpostavku da je distribucija frekvencija unutar modalnog razreda linearna (slika 24).

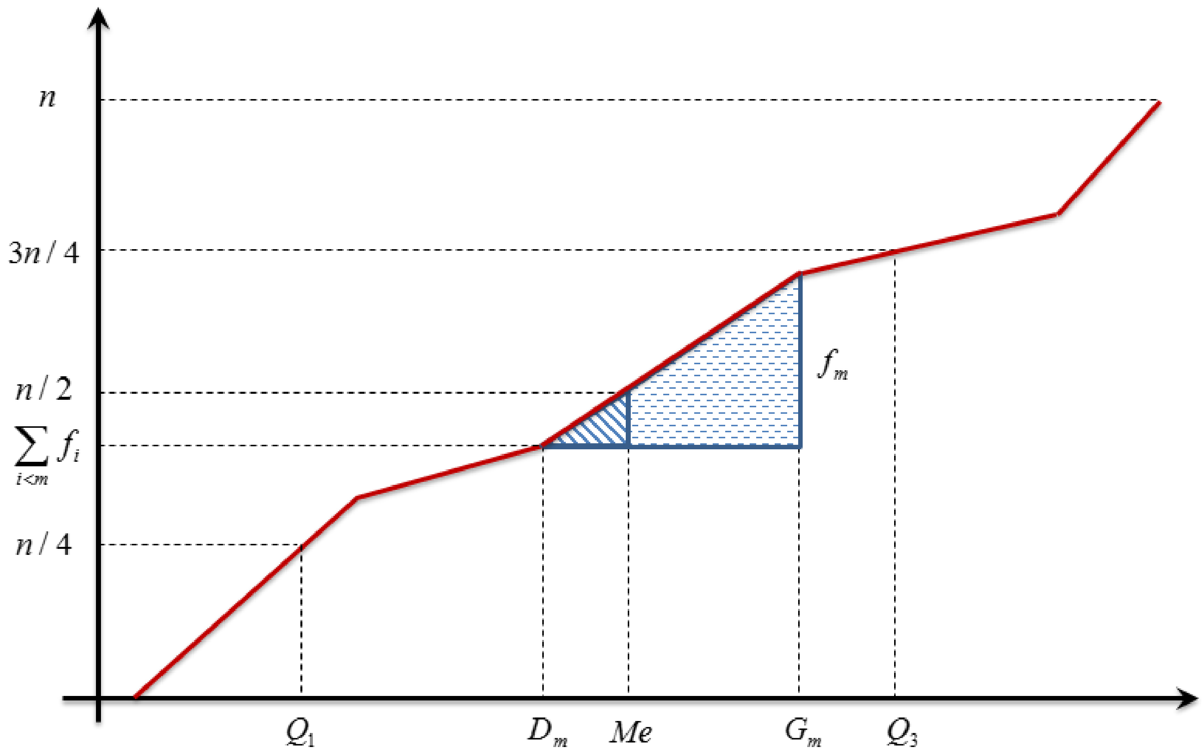


Slika 24. Mod podataka grupiranih u razrede

Ako su D i G donja i gornja granica modalnog razreda a d i g razlike frekvencija modalnog razreda i susjednih razreda na donjoj i gornjoj granici tada iz sličnosti dvaju trokuta (slika 24) imamo $x:y=d:g$ i $x+y=G-D=$ veličina modalnog razreda, iz čega dobivamo x i y. Sada je $Mo=D+x=G-y$, dakle

$$Mo=D+\frac{d}{d+g}(G-D) \quad ili \quad Mo=G-\frac{g}{d+g}(G-D).$$

Za određivanje ostalih sredina formiramo kumulativni niz frekvencija po razredima. Ako je n podataka grupirano u k razreda sa frekvencijama $f_i,\ i=1,2,\ldots,k$ tada se razred čija kumulativna frekvencija prvi put sadrži broj $n/2$ naziva *medijalni razred*. Slično je za $n/4$ *prvi* ili *donji kvartilni razred*, za $3n/4$ *treći* ili *gornji kvartilni razred*, za $n/10$ *prvi decilni razred* itd.). Ako je $m-$ ti razred medijalni, medijan određujemo tako da unutar medijalnog razreda frekvenciju proširimo linearno do $n/2$ (slika 25).



Slika 25. Kvantili podataka grupiranih u razrede

Slično ako je q_1 (q_3) donji (gornji) kvartilni razred frekvenciju proširujemo do $n/4$ ($3n/4$), kod decila do $n/10$, $2n/10$ itd. Dakle, ako su $D_m, D_{q_1}, D_{q_3}, D_{p_1}, \ldots$ i $G_m, G_{q_1}, G_{q_3}, G_{p_1}, \ldots$ donje i gornje granice medijalnog, donjeg kvartilnog, gornjeg kvartilnog, prvog decilnog, … razreda, imamo (slika 25)

$$Me = D_m + \frac{\frac{n}{2} - \sum_{i<m} f_i}{f_m}(G_m - D_m), \qquad Q_1 = D_{q_1} + \frac{\frac{n}{4} - \sum_{i<q_1} f_i}{f_{q_1}}(G_{q_1} - D_{q_1}),$$

$$Q_3 = D_{q_3} + \frac{\frac{3n}{4} - \sum_{i<q_3} f_i}{f_{q_3}}(G_{q_3} - D_{q_3}), \quad P_1 = D_{p_1} + \frac{\frac{n}{10} - \sum_{i<p_1} f_i}{f_{p_1}}(G_{p_1} - D_{p_1}), \quad \ldots$$

Primijetimo da su $G_{\ldots} - D_{\ldots}$ veličine pojedinih razreda.

ZADACI

1. Izračunajte: (a) $\sum_{i=1}^{6}(2i+1) = ?$, (b) $\sum_{i=1}^{5} i^2 = ?$, (c) $\sum_{i=0}^{3} \frac{i^2}{i+1} = ?$, (d) $\sum_{i=-n}^{n} i^3 = ?$,
   (e) $\sum_{i=1}^{n-1}(x_{i+1} - x_i) = ?$, (f) $\sum_{i=1}^{n} i = ?$.
2. Izračunajte: (a) $\prod_{i=1}^{5}(2i-1) = ?$, (b) $\prod_{i=0}^{3}(i^2 + 2i - 2) = ?$, (c) $\prod_{i=1}^{4} \sqrt[i]{x} = ?$,
   (d) $\prod_{i=1}^{n} i = ?$, (e) $\prod_{i=1}^{n-1} \frac{x_i}{x_{i+1}} = ?$, (f) $\prod_{i=1}^{n} 4^i = ?$.
3. Riješite primjere 9 i 10 za matematiku i usporedite rezultate. Iste primjere možete riješiti za bilo koji predmet u vašem razredu.
4. Definirajte obilježja svojih aktivnosti tokom 24 sata te ih prikažite strukturnim krugom (npr. vrijeme provedeno u spavanju, u školi, za učenje, jelo, odmor, zabavu i sl.). Izračunajte aritmetičke sredine svih obilježja te napravite jedinstveni strukturni krug za vaš razred. Kako on reprezentira dnevnu aktivnost vašeg razreda kao cjeline, istaknite ga na vidljivo mjesto.
5. Dokažite da vrijedi $H \le G \le A$ za niz sa dva pozitivna člana, x_1, x_2. Kada će nejednakost postati jednakost?
6. Odredite geometrijsku sredinu srednjih vrijednosti H, G, A za dva pozitivna člana.
7. Odredite prosječnu godišnju stopu pokrivenosti uvoza izvozom za RH za razdoblje 2009-2014. na temelju podataka iz tabele 9 (poglavlje *Kvantitativni podaci*).
8. Po gregorijanskom kalendaru prijestupna je svaka godina djeljiva sa 4 a izuzetak su godine djeljive sa 100 među kojima su prijestupne samo one djeljive sa 400. Na temelju tih pravila odredite stvarno trajanje kalendarske godine.
9. U jednom novoizgrađenom stambenom kompleksu tokom godine je prodano: 9 stanova površine 20-30 m^2 po cijeni 1700 €/m^2, 22 stana površine 30-50 m^2 po cijeni 1800 €/m^2, 12 stanova površine 50-80 m^2 po cijeni 1650 €/m^2 i 7 stanova površine 80-120 m^2 po cijeni 1550 €/m^2. Odredite prosječnu površinu i vrijednost kupljenog stana te prosječnu cijenu po m^2.
10. Odredite stranicu pravilnog n-terokuta čiji je opseg jednak opsegu proizvoljnog n-terokuta sa stranicama $a_1, a_2, \ldots, a_n$.

11. (a) Odredite stranicu kvadrata čija je površina jednaka površini proizvoljnog pravokutnika sa stranicama a i b.

    (b) Odredite brid kocke čiji je volumen jednak volumenu proizvoljnog kvadra sa bridovima a, b i c.
12. Koji dio cjeline najbolje reprezentira niz: polovina, trećina, četvrtina, šestina, desetina?
13. Za koji x izraz

$$\left(\frac{16}{3}-x\right)^2+\left(\frac{1-\sqrt{2}}{2}-x\right)^2+\left(\frac{1}{6}-x\right)^2+\left(\frac{1}{\sqrt{2}}-x\right)^2$$

    poprima najmanju moguću vrijednost i koliko ona iznosi?
14. Sa kojom točnošću procjenjujete trajanje jedne minute? Prikladno za rad u parovima. Od dogovorenog startnog trenutka jedan učenik mjeri vrijeme na satu a drugi, kad je procijenio da je protekla jedna minuta, izjavljuje stop. Učenici bilježe procijenjeno vrijeme (u sekundama) i zamjenjuju uloge. Cijeli postupak treba provesti bar pet puta. Za svakog učenika odredite aritmetičku sredinu njegovih procjena te apsolutno i relativno odstupanje od stvarne vrijednosti (60 sekundi). Za mjerenje umjesto jedne minute može se uzeti bilo koji drugi vremenski interval.
15. Ako napravimo leguru od 15 grama 12-karatnog zlata i 5 grama 20-karatnog zlata kolika je finoća legure? Napomena: finoća 1 karat označava 1/24 težinskog udjela čistog zlata u leguri (smjesi).
16. Bazen se puni kroz tri cijevi istovremeno. Ako kroz prvu cijev svake sekunde ulazi 16 litara vode temperature 15°C, kroz drugu 10 litara temperature 25°C, kojom brzinom treba puniti kroz treću cijev vodu temperature 40°C da bi dobili temperaturu bazena 29°C?
17. Raspolažemo sa dvije vrste kave: lošije kvalitete po cijeni od 35 kuna za kilogram i bolje kvalitete po cijeni od 70 kuna. U kojem količinskom omjeru treba pomiješati te dvije vrste da bi dobili cijenu 45 kuna za kilogram? Poopćite rezultat ako je p cijena neke robe lošije kvalitete, q cijena istovrsne robe bolje kvalitete a r željena cijena smjese ($p<r<q$). Napomena: osim za cijenu rezultat vrijedi i za svako drugo svojstvo različitog intenziteta (stupnja) koje se može mjeriti kao što je koncentracija (%), toplina (temperatura), finoća, gustoća i sl.
18. Na tahografu autobusa izvanredne linije očitani su slijedeći podaci: prvih 19 minuta vozio je prosječnom brzinom 45 km/h, narednih 13 minuta 70 km/h, zatim 25 minuta 95 km/h, 17 minuta 80 km/h i 26 minuta 50 km/h. Odredite prosječnu brzinu na toj liniji i duljinu linije (prijeđeni put).
19. U brodogradilištu za gradnju broda formirana je ekipa od 200 radnika: 50 pomoćnih radnika (P), 90 bravara (B), 40 zavarivača (Z) i 20 inženjera (I). Plaća radnika po kategorijama P, B, Z, I je redom 20, 32, 42, 50 kuna po satu. Odredite:

    (a) Ukupnu vrijednost jednog sata rada čitave ekipe.

    (b) Prosječnu vrijednost (plaću) jednog sata rada po radniku. Interpretirajte rezultat.

(c) Koliko radnika ima plaću višu od prosječne (pozitivno odstupanje) a koliko nižu (negativno odstupanje)?
(d) Koliko iznosi ukupno pozitivno i negativno odstupanje? Obrazložite.
(e) Koliko iznosi najveće i najmanje pozitivno i negativno odstupanje po radniku?
(f) Koliko iznosi prosječno pozitivno i negativno odstupanje po radniku?
(g) Koliko iznosi prosječno pozitivno i negativno relativno odstupanje po radniku?
(h) Prikažite vrijednost jednog sata rada čitave ekipe kumulantom po kategorijama (razredima) redoslijedom P, B, Z, I.
(i) Prikažite kumulantu pod (h) redoslijedom I, Z, B, P.

20. Gdje se nalazi prvi, 21., 77. i zadnji percentil u nizu od 7286 članova?
21. Gdje su medijan, kvartili i decili u nizu sa: (a) 40, (b) 42, (c) 43 člana?
22. Odredite medijan i kvartile niza učenika vašeg razreda uređenog po abecedi.¨
23. Ako imamo rang listu osoba prema određenom kriteriju (uspjehu) od najlošijeg do najboljeg, kako ćemo izdvojiti: (a) 20% najlošijih, (b) 25% najboljih, (c) polovicu prosječnih (srednjih), (d) 10% prosječnih, (e) 1% najboljih? Isto pitanje ako je lista poredana od najboljeg do najlošijeg.
24. Broj turista u jednom primorskom mjestu bio je: 36000 (2013.), 40000 (2014.) i 38000 (2015.). Koje godine i približno u kojem mjesecu: (a) je došlo najviše turista, (b) je kumulativni broj turista prešao polovicu ukupnog broja turista u ove tri godine?

RJEŠENJA

1. (a) 48, (b) 55, (c) $49/12$, (d) 0 jer je $(-x)^3+x^3=0$ za svaki x, (e) x_n-x_1, (f) $n(n+1)/2$.
2. (a) 945, (b) -156, (c) $x^{25/12}=x^2\cdot\sqrt[12]{x}$, (d) $n!$, (e) x_1/x_n, (f) $4^{1+2+\ldots+n}=2^{n(n+1)}$.
3. $A=2.735\ldots$, $G=2.451\ldots$, $H=2.145\ldots$. $Mo=3$, $Me=3$ (između 17. i 18. člana - nije član niza), $Q_1=2$ (9. član), $Q_3=3$ (26. član).
4. Koristite stvarne podatke.
5. $\left(\sqrt{x_1}-\sqrt{x_2}\right)^2\geq 0 \Rightarrow x_1-2\sqrt{x_1x_2}+x_2\geq 0 \Rightarrow (x_1+x_2)/2\geq\sqrt{x_1x_2}$, tj. $A\geq G$. Na sličan način iz $\left(\dfrac{1}{\sqrt{x_1}}-\dfrac{1}{\sqrt{x_2}}\right)^2\geq 0$ slijedi $G\geq H$. Jednakost nastupa za $x_1=x_2$.
6. $\sqrt[3]{AGH}=\sqrt[3]{\dfrac{x_1+x_2}{2}\cdot\sqrt{x_1x_2}\cdot\dfrac{2}{\frac{1}{x_1}+\frac{1}{x_2}}}=\sqrt[3]{x_1x_2\sqrt{x_1x_2}}=\sqrt{x_1x_2}=G \Rightarrow G^3=AGH \Rightarrow$ $G^2=AH \Rightarrow G=\sqrt{AH}$. Geometrijska sredina dvaju pozitivnih brojeva jednaka je geometrijskoj sredini njihove aritmetičke i harmonijske sredine.
7. Pokrivenost uvoza izvozom = ukupna vrijednost izvoza / ukupna vrijednost uvoza,

$$R=\frac{1}{6}\left(\frac{55.2}{111.8}+\frac{64.9}{110.3}+\frac{71.2}{121}+\frac{72.2}{121.5}+\frac{72.6}{125.1}+\frac{79.1}{130.7}\right)=0.575\ldots\approx 57.5\%.$$

8. Kako u 400 godina ima 100–3=97 prijestupnih, koristeći ponderiranu aritmetičku sredinu, imamo

$$\frac{97\cdot 366+303\cdot 365}{400}=365.2425 \text{ dana.}$$

Napomenimo da je stvarno trajanje astronomske tropske godine 365.242189 dana.

9. Podaci za površinu su sredine razreda. Za ukupnu površinu P i vrijednost V imamo

$$P=25\cdot 9+40\cdot 22+65\cdot 12+100\cdot 7=2585 \text{ m}^2,$$
$$V=25\cdot 9\cdot 1700+40\cdot 22\cdot 1800+65\cdot 12\cdot 1650+100\cdot 7\cdot 1550=4338500 \text{ EUR.}$$

Kako je prodano 50 stanova, prosječna površina stana je $P/50=51.7\,\text{m}^2$, prosječna vrijednost je $V/50=86770$ EUR a cijena $V/P=1678.34$ EUR po m^2.

10. Stranica je aritmetička sredina, $a=(a_1+a_2+...+a_n)/n$.

11. Geometrijska sredina: (a) $x=\sqrt{ab}$, (b) $x=\sqrt[3]{abc}$.

12. Harmonijska sredina jer imamo recipročno obilježje, $H=5/(2+3+4+6+10)=1/5$.

13. Prema relaciji (12) x je aritmetička sredina brojeva $16/3$, $(1-\sqrt{2})/2$, $1/6$ i $1/\sqrt{2}$. Dakle, najmanja vrijednost je za $x=6/4=1.5$ i iznosi $373/18-\sqrt{2}/2=20.0151...$.

14. Koristite vaše podatke. Dobiveno relativno odstupanje pokazuje za koliko % prema vašem subjektivnom dojmu vrijeme protječe brže (negativno odstupanje) ili sporije (pozitivno odstupanje) od stvarnog protoka.

15. Ponderirana aritmetička sredina: $(15\cdot 12+5\cdot 20)/(15+5)=14$ karata.

16. $(16\cdot 15+10\cdot 25+x\cdot 40)/(16+10+x)=29 \;\Rightarrow\; x=24$ litara u sekundi.

17. $(35x+70y)/(x+y)=45 \;\Rightarrow\; 10x=25y \;\Rightarrow\; x/y=5/2$ ili $x:y=5:2$. Općenito je $(px+qy)/(x+y)=r \;\Rightarrow\; (r-p)x=(q-r)y \;\Rightarrow\; x:y=(q-r):(r-p)$.

18. $(19\cdot 45+13\cdot 70+25\cdot 95+17\cdot 80+26\cdot 50):(19+13+25+17+26)=6800:100=68$ km/h. Put je $68\cdot(100/60)\approx 113.3$ km.

19. (a) $50\cdot 20+90\cdot 32+40\cdot 42+20\cdot 50=6560$ kuna.

(b) $6560:200=32.80$ kuna. Uz istu ukupnu masu plaće za bilo koji vremenski period, po ovoj vrijednosti sata svi radnici bi dobili jednak iznos.

(c) 60 ima višu a 140 nižu.

(d) Pozitivno odstupanje: $40\cdot(42-32.8)+20\cdot(50-32.8)=712$ kuna po satu. Negativno odstupanje: $50\cdot(20-32.8)+90\cdot(32-32.8)=-712$ kuna po satu. Odstupanja su jednaka ali suprotnih predznaka jer im je zbroj nula (zbroj apsolutnih odstupanje podataka od aritmetičke sredine je nula).

(e) Pozitivno 9.2 i 17.2 a negativno –0.8 i –12.8.

(f) $712:60=11.87$ i $(-712):140=-5.09$ kuna po satu i po radniku.

(g) $11.87:32.8\approx 36.2\%$ i $(-5.09):32.8\approx -15.5\%$.

(h) Na apscisi prikazujemo razrede P(0-50), B(50-140), Z(140-180), I(180-200) a zatim poligonalnom linijom spajamo točke T_0,T_1,T_2,T_3,T_4 gdje su: $T_0=(0,0)$,

$T_1 = (50, 50 \cdot 20) = (50, 1000)$, $T_2 = (140, 1000 + 90 \cdot 32) = (140, 3880)$,

$T_3 = (180, 3880 + 40 \cdot 42) = (180, 5560)$, $T_4 = (200, 5560 + 20 \cdot 50) = (200, 6560)$.

(i) Na apscisi prikazujemo razrede I(0-20), Z(20-60), B(60-150), P(150-200) a zatim poligonalnom linijom spajamo točke S_0, S_1, S_2, S_3, S_4 gdje su sada: $S_0 = (0, 0)$, $S_1 = (20, 1000)$, $S_2 = (60, 2680)$, $S_3 = (150, 5560)$, $S_4 = (200, 6560)$.

20. $n = 7286$ pa je $n/100 = 72.86 \Rightarrow P_1 = 73.$ član, $21n/100 = 1530.06 \Rightarrow P_{21} = 1531.$ član, $77n/100 = 5610.22 \Rightarrow P_{77} = 5611.$ član, $99n/100 = 7213.14 \Rightarrow P_{99} = 7214.$ član.

21. (a) Medijan je sredina 20. i 21. člana (20-21.), kvartili su 10-11. i 30-31. a decili 4-5., 8-9., 12-13., …, 36-37. (b) Medijan je 21-22., kvartili su 11. i 32. član a decili 5., 9., 13., 17., 21-22., 26., 30., 34. i 38. član. (c) Medijan je 22. član, kvartili su 11. i 33. član a decili 5., 9., 13., 18., 22., 26., 31., 35. i 39. član.

22. Koristite stvarne podatke.

23. (a) Ispred D_2 (od početka niza do D_2 ali bez D_2), (b) iza Q_3 (od Q_3 do kraja niza ali bez Q_3), (c) između Q_1 i Q_3, (d) između P_{45} i P_{55}, (e) iza P_{99} (od P_{99} do kraja niza ali bez P_{99}). Za suprotno poredanu listu: (a) iza D_8, (b) ispred Q_1, (c) i (d) isti odgovor, (e) ispred P_1.

24. U zadatku (a) tražimo mod. Kako je modalni razred 2014., prema slici 24 imamo

$$Mo = 2013 + \frac{4}{4+2} \cdot (2014 - 2013) = 2013 + \frac{4}{6} = 2013 + \frac{8}{12},$$

pa je modalni mjesec 8. mjesec 2014.

U zadatku (b) tražimo medijan. Kako je $n = 114000$ (ukupan broj turista u sve tri godine), $n/2 = 57000$ pa je medijalni razred 2014. Prema slici 25 imamo

$$Me = 2013 + \frac{57000 - 36000}{40000} \cdot (2014 - 2013) = 2013 + \frac{21}{40} = 2013 + \frac{6.3}{12},$$

pa se medijan ostvario u 7. mjesecu (6.3) 2014.

MJERE RASPRŠENOSTI

Vidjeli smo da su srednje vrijednosti podaci koji sažimaju informacije niza podataka u jednu vrijednost. Ako želimo detaljnije analizirati raspored podataka unutar niza koristimo mjere raspršenosti (disperzije). Ima ih više vrsta a pokazuju kolike su varijacije među podacima te kako su podaci raspoređeni u odnosu na neku srednju vrijednost (kolika su njihova odstupanja od te sredine). Raspršenost podataka ovisi o njihovoj različitosti. Ako su svi podaci međusobno jednaki nema raspršenosti pa je njena mjera jednaka nuli. Što je različitost veća to je mjera raspršenosti veća. Možemo ih izraziti u mjernim jedinicama obilježja ili relativnim iznosima, koristeći sve ili samo neke podatke iz promatranog niza. Upoznajemo ih preko slijedećeg primjera.

Primjer 11. Prosječne dnevne temperature u Labinu i Ivancu tokom jednog tjedna izražene u ^{o}C dane su u tabeli 10. Analizirajmo varijacije temperature koristeći mjere raspršenosti.

Dan	*PON*	*UTO*	*SRI*	*ČET*	*PET*	*SUB*	*NED*
Labin	13	15	12	13	16	19	17
Ivanec	8	5	10	15	22	25	20

Tabela 10. Temperature tokom promatranog tjedna

Jedan od načina mjerenja raspršenosti je koristeći neke karakteristične vrijednosti (članove) niza. Tu su najvažniji *raspon varijacije, interkvartilni raspon* i *koeficijent kvartilne devijacije.*

DEFINICIJA 4. ZA NUMERIČKI NIZ $x_1, x_2, ..., x_{n-1}, x_n$ JE

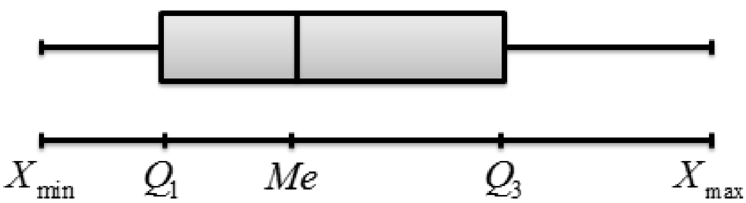
$$R = x_{\max} - x_{\min} \quad \text{(raspon varijacije),}$$

$$I_Q = Q_3 - Q_1 \quad \text{(interkvartilni raspon),}$$

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad \text{(koeficijent kvartilne devijacije),}$$

GDJE SU Q_1 I Q_3 KVARTILI, TE $x_{\min} = \min\{x_1, x_2, ..., x_n\}$, $x_{\max} = \max\{x_1, x_2, ..., x_n\}$.

Vidimo da svaka od ove tri mjere raspršenosti koristi samo dva podatka iz niza. Pri tome su R i I_Q izražene u mjernim jedinicama obilježja dok je V_Q relativna mjera pa se može koristiti za usporedbu različitih obilježja. Raspon varijacije je razmak između najveće i najmanje vrijednosti niza koje mogu biti netipične za niz (ekstremno male ili velike u odnosu na većinu ostalih članova niza). Zbog toga ova mjera ne daje uvijek pravi uvid u varijabilnost članova niza. Interkvartilni raspon isključuje prvu četvrtinu najmanjih i zadnju četvrtinu najvećih članova niza pa pokazuje varijabilnost srednje polovice niza. Time su krajnje, možda netipične, vrijednosti niza isključene. Koeficijent kvartilne devijacije je relativna mjera (%) tog raspona koja je za niz pozitivnih članova uvijek između 0 i 1. Stupanj raspršenosti podataka često se slikovito prikazuje *B-P dijagramom* (box-plot) koji uključuje pet temeljnih pokazatelja: minimalnu i maksimalnu vrijednost, donji i gornji kvartil i medijan (slika 26).



Slika 26. B-P dijagram

U primjeru 11 formirat ćemo niz temperatura po rastućem redoslijedu:

$$\text{Labin: } 12, 13, 13, 15, 16, 17, 19. \qquad \text{Ivanec: } 5, 8, 10, 15, 20, 22, 25.$$

Vidimo da je srednja tjedna temperatura (aritmetička sredina) u oba mjesta bila 15^{o}C a to je i medijalna temperatura (medijan) što znači da je u oba mjesta tri dana u tjednu temperatura bila manja a tri dana veća od 15^{o}C. Međutim usporedimo li varijacije temperature u ovim mjestima, razlika je očita. To će pokazati i navedene mjere. Kako je donji kvartil drugi član niza a gornji šesti član, imamo

$$\text{Labin}: \quad R = 19 - 12 = 7, \quad I_Q = 17 - 13 = 4, \quad V_Q = \frac{17-13}{17+13} = 0.1333... \approx 13.3\%,$$

$$\text{Ivanec}: \quad R = 25 - 5 = 20, \quad I_Q = 22 - 8 = 14, \quad V_Q = \frac{22-8}{22+8} = 0.4666... \approx 46.7\%.$$

Uočavamo znatno veće vrijednosti mjera raspršenosti temperatura za Ivanec u odnosu na Labin. Dobivene vrijednosti možemo usporediti i interpretirati na gore opisani način.

U definiciji navedenih mjera raspršenosti korišteni su samo neki članovi niza. Najvažnije mjere u kojima se koriste svi članovi niza su *varijanca*, *standardna devijacija* i *koeficijent varijacije*.

DEFINICIJA 5. ZA NUMERIČKI NIZ $x_1, x_2, ..., x_{n-1}, x_n$ JE

$$\sigma^2 = \frac{(x_1 - A)^2 + (x_2 - A)^2 + ... + (x_n - A)^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 \qquad \text{(varijanca)},$$

$$\sigma = \sqrt{\frac{(x_1 - A)^2 + (x_2 - A)^2 + ... + (x_n - A)^2}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2} \qquad \text{(standardna devijacija)},$$

$$V = \frac{\sigma}{A} \qquad \text{(koeficijent varijacije)},$$

GDJE JE A ARITMETIČKA SREDINA.

Vidimo da je varijanca aritmetička sredina kvadrata odstupanja članova niza od njegove aritmetičke sredine. Standardna devijacija je ista mjera ali izražena u mjernim jedinicama obilježja (niza). Budući da je zbroj apsolutnih odstupanja svih članova niza od aritmetičke sredine jednak nuli (relacija (11)) a zbroj kvadrata odstupanja svih članova niza minimalan (relacija (12)), ovako definirane mjere raspršenosti su najbolje moguće.

Koeficijent varijacije je neovisan o mjernim jedinicama pa se može koristiti za usporedbu različitih obilježja. On pokazuje u kojoj mjeri aritmetička sredina A reprezentira niz podataka (što je manji to je reprezentativnost bolja). Općenito za vrijednosti veće od 30% smatra se slaba reprezentativnost odnosno velika raspršenost podataka oko prosjeka A.

Kako je $\sum_{i=1}^{n} x_i = nA$ i $A = \frac{1}{n}\sum_{i=1}^{n} x_i$, imamo

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\cdot 2A\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} A^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\cdot 2A\cdot nA + \frac{1}{n}\cdot nA^2,$$

pa se navedeni izraz može pisati u alternativnim oblicima,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - A^2 \quad \text{ili} \quad \sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]. \tag{14}$$

Ako su podaci grupirani kao distribucija frekvencija (13) tada izraz poprima oblike

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - A)^2 \quad \text{ili} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - A^2 \quad \text{ili} \quad \sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{k} f_i x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} f_i x_i\right)^2\right],$$

koje možemo izraziti i pomoću relativnih frekvencija $p_i = f_i / n$,

$$\sigma^2 = \sum_{i=1}^{k} p_i (x_i - A)^2 \quad \text{ili} \quad \sigma^2 = \sum_{i=1}^{k} p_i x_i^2 - A^2 \quad \text{ili} \quad \sigma^2 = \sum_{i=1}^{k} p_i x_i^2 - \left(\sum_{i=1}^{k} p_i x_i\right)^2.$$

Ako su podaci uređeni kao distribucija frekvencija sa razredima u navedenim formulama za mjere raspršenosti (kao i ranije za srednje vrijednosti) x_i su razredne sredine.

U navedenom primjeru 11 imamo $A = 15$ pa je

$$\text{Labin}: \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - A^2 = \frac{1}{7}(12^2 + 13^2 + 13^2 + 15^2 + 16^2 + 17^2 + 19^2) - 15^2 = \frac{38}{7},$$

$$\text{Ivanec}: \quad \sigma^2 = \frac{1}{7}(5^2 + 8^2 + 10^2 + 15^2 + 20^2 + 22^2 + 25^2) - 15^2 = \frac{348}{7},$$

odnosno

$$\text{Labin}: \quad \sigma = \sqrt{\sigma^2} = 2.329929..., \qquad V = \frac{\sigma}{A} = 0.155326... \approx 15.5\%,$$

$$\text{Ivanec}: \quad \sigma = 7.950835..., \qquad V = 0.470055... \approx 47\%.$$

Uočimo da 15°C prilično dobro reprezentira temperaturu tokom tjedna u Labinu (uz 15.5% prosječnog odstupanja) dok to nije slučaj za Ivanec gdje su prosječna temperaturna odstupanja od srednje vrijednosti temperature u tom tjednu bila visokih 47%.

Za još točniju usporedbu relativnog položaja podataka u različitim numeričkim nizovima koristi se *standardizirano obilježje*. Ako su A i σ aritmetička sredina i standardna devijacija numeričkog niza $x_1, x_2, ..., x_n$, tada se linearna transformacija tog niza u niz $z_1, z_2, ..., z_n$ definirana relacijom

$$z_i = \frac{x_i - A}{\sigma}, \quad i = 1, 2, ..., n, \tag{15}$$

naziva standardizacija. Dobiveni standardizirani niz ima aritmetičku sredinu 0 a varijancu i standardnu devijaciju 1. Standardizirana vrijednost z_i ($z -$vrijednost ili $z -$score) pokazuje za koliko se standardnih devijacija podatak x_i odstupa od aritmetičke sredine ($x_i = A + \sigma z_i$). To omogućuje usporedbu pojedinačnih podataka iz različitih obilježja (različitih mjernih jedinica i različitih vrijednosti).

Ako u primjeru 11 (tabeli 10) usporedimo temperature u srijedu, imamo

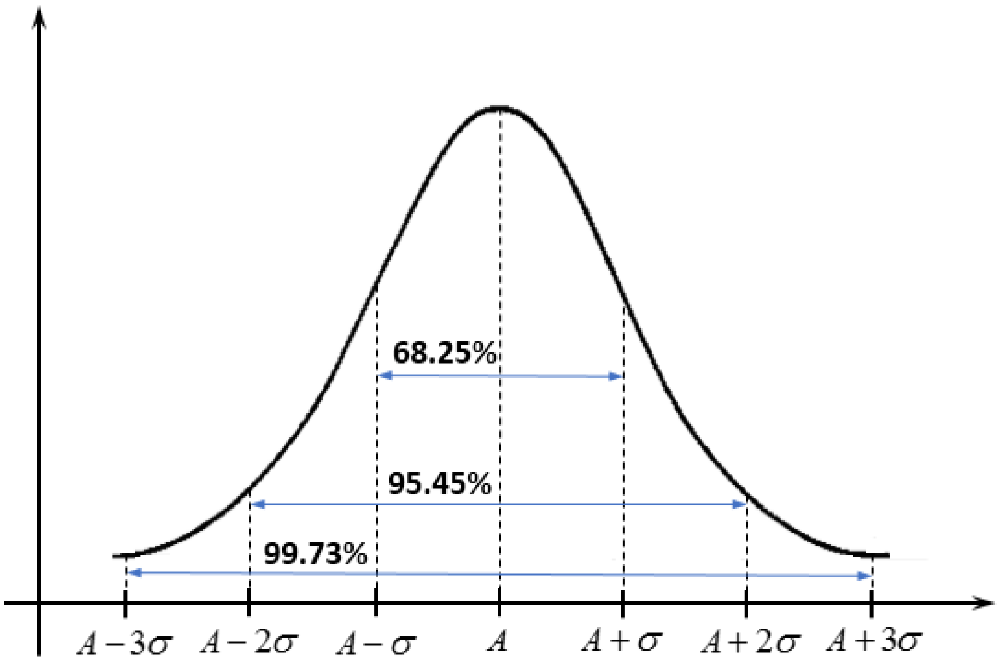
$$\text{Labin}: \quad z_{SRI} = \frac{12 - 15}{2.329929...} = -1.28759..., \quad \text{Ivanec:} \quad z_{SRI} = \frac{10 - 15}{7.950835...} = -0.62886...,$$

što pokazuje da je 5°C ispod prosječne temperature u Ivancu daleko manje (≈ dvostruko manje) odstupanje u odnosu na disperziju temperatura tog tjedna nego što je to 3°C u Labinu.

Jedna od najvažnijih oblika distribucije podataka je *normalna ili Gaussova distribucija* (prirodna distribucija). To je neprekidno numeričko obilježje (realna funkcija $f : R \to R_+$) čiji je analitički oblik

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-A}{\sigma}\right)^2}, \ A \in R, \ \sigma > 0 \quad \text{ili} \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \ \text{(standardizirani oblik)},$$

a graf je zvonolika simetrična krivulja (slika 27). Njena aritmetička sredina, mod i medijan su međusobno jednaki (A) a ona je simetrično raspoređena oko njih (pravac $x = A$ je os simetrije). Po ovoj distribuciji prirodno se raspoređuju obilježja sa velikim brojem podataka (vrijednosti) u prirodi i društvu pa je ona osnova za analizu pojava u teoriji vjerojatnosti. Tako na primjer ako nasumce (slučajnim odabirom) izaberemo 10000 ljudi tada se njihova obilježja, koja se mogu izraziti numeričkom varijablom – kvantitativna obilježja (visina, težina, inteligencija, fizičke i mentalne sposobnosti, medicinski parametri itd.), raspoređena po normalnoj distribuciji. Jedno od njenih važnih svojstava je da se u intervalu $(A-\sigma, A+\sigma)$ nalazi 68.25% svih podataka, u intervalu $(A-2\sigma, A+2\sigma)$ 95.45% te u intervalu $(A-3\sigma, A+3\sigma)$ 99.73% ili gotovo svi podaci. Ove vrijednosti služe za usporedbu proizvoljne distribucije sa normalnom. Općenito, vrijednosti koje od aritmetičke sredine odstupaju za više od dvije standardne devijacije, tj. koje ne pripadaju intervalu $(A-2\sigma, A+2\sigma)$ smatraju se netipične za promatrano obilježje.



Slika 27. Normalna ili Gaussova distribucija

ZADACI

1. Za primjer 11 odstupanja od prosječne temperature po danima prikažite dijagramom apsolutnih odstupanja. Dijagramom odvojenih stupaca prikažite kvadrate tih odstupanja te na dijagramu istaknite varijancu.
2. Svaki učenik neka na papirić napiše jedan proizvoljno izabran prirodni broj. Za dobivene podatke napravite B-P dijagram.
3. Starost članova jedne osmeročlane obitelji je 13, 14, 18, 30, 44, 45, 75 i 80 godina. Odredite raspon varijacije, interkvartilni raspon i koeficijent kvartilne devijacije te prosječnu starost članova te obitelji.
4. Odredite kvartile ako je $I_Q = 46$, $V_Q = 0.23$.
5. Ako je na jednoj provjeri znanja 10% ispitanika dobilo ocjenu nedovoljan, 20% dovoljan, 30% dobar, 25% vrlo dobar i 10% izvrstan, kolika je prosječna ocjena, prosječno odstupanje i stupanj varijabilnosti (koeficijent varijacije)?
6. Svaki učenik neka na papirić napiše jedan broj (ako želite više podataka tada dva ili tri ili više) iz intervala [0, 1]. Brojevi neka budu napisani barem sa tri decimale, npr. 0.765, 0.200, 0.590, 0.12345 i sl. Odredite aritmetičku sredinu, standardnu devijaciju i koeficijent varijacije za dobivene podatke. Prikažite histogramom distribuciju podataka grupiranih u razrede [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8) i [0.8, 1].
7. Koliki vam je prosječni tjedni džeparac koji potrošite kroz tjedan? Prikupite podatke za vaš razred te odredite raspon varijacije, aritmetičku sredinu, standardnu devijaciju i koeficijent varijacije.
8. Prosječno vrijeme čekanja do javljanja operatera na automatskoj telefonskoj sekretarici za 100 praćenih poziva bilo je: do 30 sekundi 6 poziva, od 30 sekundi do 1

minute 16 poziva, 1-2 minute 42 poziva, 2-4 minute 22 poziva i 4-8 minuta 14 poziva. Odredite prosječno vrijeme čekanja, standardnu devijaciju i koeficijent varijacije.

9. Svaki od vas neka zasebno metrom izmjeri duljinu zadane dužine (širinu, duljinu ili visinu nekog predmeta, prostorije, objekta, staze i sl.). Na temelju sakupljenih podataka mjerenja odredite raspon varijacije, aritmetičku sredinu, standardnu devijaciju i koeficijent varijacije.
10. U jednoj trgovini tokom tjedna (šest radnih dana) prodane količine proizvoda X bile su 44, 45, 41, 48, 55, 61 a proizvoda Y 105, 110, 99, 107, 96, 101. Čija prodaja ima bolju ujednačenost po danima?
11. Ako su A, σ i V aritmetička sredina, standardna devijacija i koeficijent varijacije za niz $x_1, x_2, ..., x_n$, kolike su navedene vrijednosti za niz: (a) $x_1+c,\ x_2+c,\ ...,\ x_n+c$, (b) $cx_1,\ cx_2,\ ...,\ cx_n$, gdje je c proizvoljna pozitivna konstanta?
12. Odredite A, σ i V za niz $x_1,\ x_2\ (n=2)$.
13. Ako je $\sum_{i=1}^{n} x_i = 55,\quad \sum_{i=1}^{n} x_i^2 = 385,\quad \sigma^2 = 8.25$, koliki je n ?
14. Najveća udaljenost Zemlje od Sunca je 152.2 a najmanja 147 milijuna km. Najveća udaljenost Mjeseca od Zemlje je 405504 a najmanja 363296 km. Odredite odstupanje putanje Zemlje oko Sunca i Mjeseca oko Zemlje od idealne kružne putanje. Čije odstupanje je veće?
15. U jednom proizvodnom pogonu 50 radnika iz dopremljenih dijelova sklapa gotove proizvode. Radnici koji sklope više proizvoda od prosjeka dobiju stimulaciju na osnovnu plaću a za manje destimulaciju (umanjenje). Tokom jednog mjeseca u pogonu je sklopljeno 20000 proizvoda X a tokom drugog 8000 proizvoda Y. Pri tome je srednje odstupanje (devijacija) od prosjeka po radniku bilo 50 proizvoda X i 10 Y. Ako je radnik Marko sklopio 420 proizvoda X i 158 Y a Tomo 390 X i 165 Y što zaključujemo o njihovim (de)stimulacijama?
16. Prosječna plaća u firmi X je 5000 kuna uz prosječno odstupanje 500 kuna a u firmi Y 7000 kuna uz prosječno odstupanje 1000 kuna. U kojoj firmi plaće više variraju? Ako jedan radnik u firmi X zarađuje 5600 kuna a drugi u Y 8000 kuna, koji od njih je bolje plaćen u odnosu na prosjek firme?
17. Dokažite da je za standardizirano obilježje (15) aritmetička sredina 0 a varijanca 1.
18. Dokažite da za bilo koji niz realnih brojeva $x_1,\ x_2,\ ...\ ,\ x_n$ vrijedi

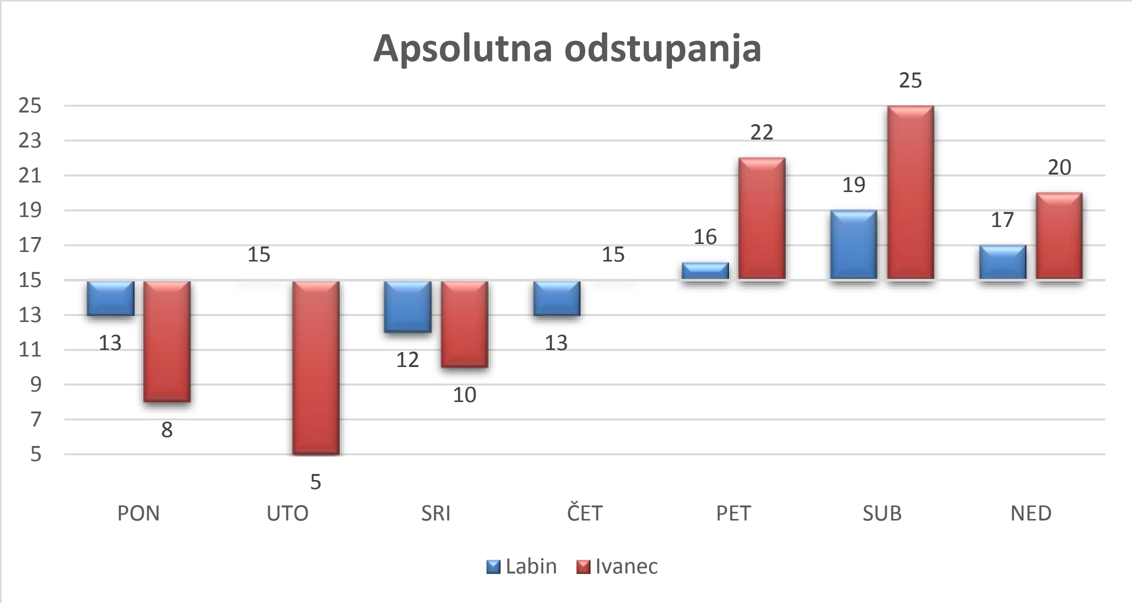
$$\left(x_1+x_2+...+x_n\right)^2 \ \le\ n\left(x_1^2+x_2^2+...+x_n^2\right).$$

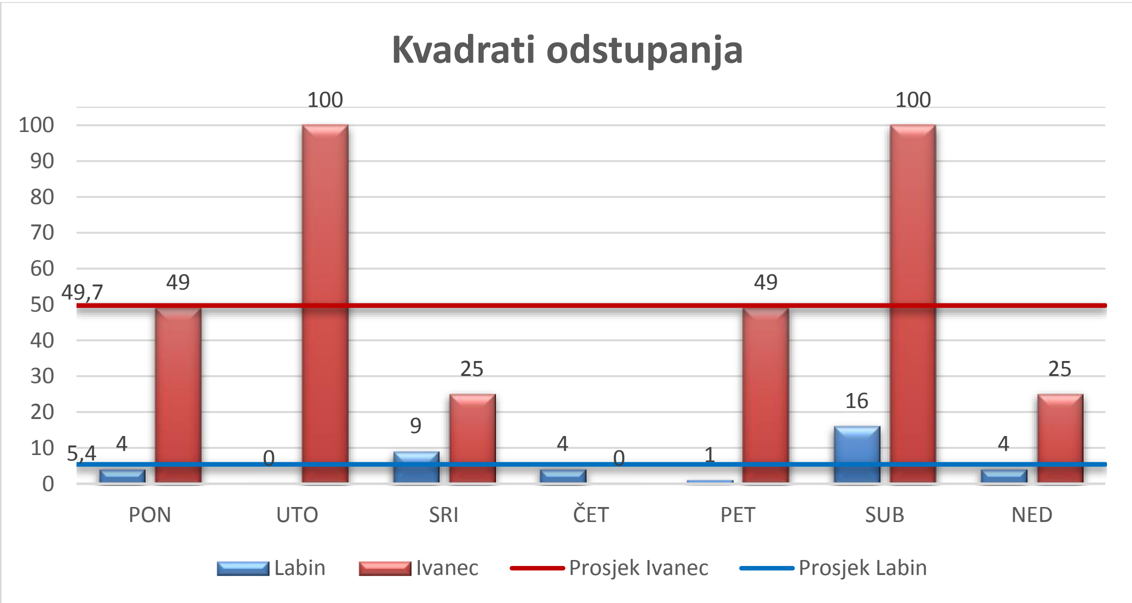
Kada ova nejednakost postaje jednakost?

19. Odredite modalnu frekvenciju standardiziranog oblika Gaussove distribucije za neprekidno numeričko obilježje. Za koju vrijednost varijable je frekvencija 10^{-1}?
20. Prosječan vijek trajanja jednog malog kućanskog aparata je 500 sati rada sa odstupanjem 10%. Ako je proizvedena serija od 3000 takvih aparata, za koliko njih možemo očekivati da: (a) neće raditi duže od 500 sati, (b) će raditi bar 450 sati, (c) će raditi više od 600 sati ?

RJEŠENJA

1. Dijagrami:





2. Brojeve poredajte po veličini, od najmanjeg do najvećeg te odredite svih pet pokazatelja za B-P dijagram (slika 26).
3. Imamo $R = 80 - 14 = 66$ godina. Kako je $Q_1 = (14+18)/2 = 16$ (sredina drugog i trećeg člana) te $Q_3 = (45+75)/2 = 60$ (sredina šestog i sedmog člana), imamo $I_Q = 44$, $V_Q = 19/11 \approx 1.727$, dok je prosječna starost $A = 39.875$ godina.
4. $Q_3 - Q_1 = 46$, $Q_3 + Q_1 = 46/0.23 = 200 \Rightarrow Q_1 = 77$, $Q_3 = 123$.
5. $A = \sum_{i=1}^{k} p_i x_i = 0.15 \cdot 1 + 0.2 \cdot 2 + 0.3 \cdot 3 + 0.25 \cdot 4 + 0.1 \cdot 5 = 2.95$,

$$\sigma=\sqrt{\sum_{i=1}^{k}p_ix_i^2-A^2}=0.15\cdot1^2+0.2\cdot2^2+0.3\cdot3^2+0.25\cdot4^2+0.1\cdot5^2-2.95^2=1.20312...,$$

$V=\sigma/A=0.4078376...\approx 40.8\%$, što je značajna varijacija uspjeha oko prosjeka.

6. Koristite prikupljene podatke. Histogramom prikazujete frekvencije pojedinih razreda (koliko brojeva pripada pojedinom razredu).
7. Koristite prikupljene podatke.
8. Podaci za vrijeme čekanja su razredne sredine (u minutama): 0.25, 0.75, 1.5, 3 i 6. Imamo $A=2.265\,\text{min}=2\,\text{min}\;15.9\,\text{s}$, $\sigma=1.71129337...$, $V=0.7555379...\approx 75.6\%$ a što pokazuje značajnu raspršenost podataka (neujednačeno vrijeme čekanja).
9. Koristite stvarne podatke dobivene mjerenjem. U praksi se često tako dobivena aritmetička sredina uzima kao točna vrijednosti za veličinu koju mjerimo.
10. Za proizvod X je $A=49$, $\sigma=6.9041$, $V=14.09\%$, a za proizvod Y je $A=103$, $\sigma=4.7958$, $V=4.656\%$, pa prodaja proizvoda Y po danima ima oko tri puta bolju ujednačenost od prodaje proizvoda X.
11. (a) $A'=A+c$, $\sigma'=\sigma$, $V'=\sigma'/A'=\sigma/(A+c)=(\sigma/A)/(1+c/A)=V/(1+c/A)$.

    (b) $A'=cA$, $\sigma'=c\sigma$, $V'=V$.
12. Imamo $A=(x_1+x_2)/2$, pa je

$$\sigma^2=\frac{1}{2}\left[\left(x_1-\frac{x_1+x_2}{2}\right)^2+\left(x_2-\frac{x_1+x_2}{2}\right)^2\right]=\frac{(x_1-x_2)^2}{4},\quad \sigma=\frac{|x_1-x_2|}{2},\quad V=\frac{|x_1-x_2|}{x_1+x_2}.$$

13. Iz relacije (14) slijedi $8.25=\frac{1}{n}\left[385-\frac{1}{n}\cdot55^2\right]$ iz čega dobivamo kvadratnu jednadžbu $8.25n^2-385n+55^2=0$ čije je cjelobrojno rješenje $n=10$.
14. Koristimo prethodni zadatak. Uz oznake $M-$Mjesec i $Z-$Zemlja imamo

$$V(Z)=\frac{|147-152.2|}{147+152.2}=0.017379...,\quad V(M)=\frac{|363296-405504|}{363296+405504}=0.054901....$$

Vidimo da je odstupanje Mjesečeve putanje od idealne kružnice 5.49% znatno veće od odstupanja Zemljine putanje 1.74%. Dobiveni rezultati su u stvari vrijednosti numeričkog ekscentriciteta elipse $\varepsilon=e/a$, $e=\sqrt{a^2-b^2}$ gdje su a i b veća i manja poluos elipse, pa je najveća udaljenost $x_1=a+e$ a najmanja $x_2=a-e$ (centar kruženja je u jednom žarištu $F(\pm e,0)$ elipse $x^2/a^2+y^2/b^2=1$). Provjerite.

15. Srednje vrijednosti su $A(\text{X})=20000:50=400$ i $A(\text{Y})=8000:50=160$. Kako su standardne devijacije $\sigma(\text{X})=50$ i $\sigma(\text{Y})=10$, imamo slijedeće $z-$vrijednosti (15).

	X	**Y**
Marko	0.4	– 0.2
Tomo	– 0.2	0.5

Vidimo da Marko treba dobiti stimulaciju za prvi mjesec a Tomo za drugi mjesec. Pri tome je Markova stimulacija manja (za 20%) od Tomove. Jednaku destimulaciju dobivaju Marko za drugi a Tomo za prvi mjesec.

16. Za X je $V = 10\%$ a za Y je $V = 14.2857\%$ pa su tu veće varijacije plaća. Za usporedbu radnika računamo standardizirane vrijednosti. Imamo $z_X = (5600-5000)/500 = 1.1$, $z_Y = (8000-7000)/1000 = 1$, pa je radnik u X relativno bolje plaćen.

17. $\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - A}{\sigma} = \frac{1}{n}\cdot\frac{1}{\sigma}\cdot 0 = 0,\quad \frac{1}{n}\sum_{i=1}^{n}(z_i - 0)^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - A)^2}{\sigma^2} = \frac{1}{n}\cdot\frac{1}{\sigma^2}\cdot n\sigma^2 = 1.$

18. Treba dokazati da je $n(x_1^2 + x_2^2 + ... + x_n^2) - (x_1 + x_2 + ... + x_n)^2 \geq 0$. Imamo

$$n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = n\cdot\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = n^2\cdot\frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = n^2\cdot\sigma^2 \geq 0,$$

gdje smo koristili relaciju (14). Jednakost nastupa za $x_1 = x_2 = ... = x_n$.

19. Kako funkcija $e^{-\frac{1}{2}z^2}$ poprima najveću vrijednost za $z = 0$, modalna frekvencija je $f(0) = (1/\sqrt{2\pi})\cdot e^0 = 1/\sqrt{2\pi} = 0.39894...$. Iz $f(z) = 0.1$ slijedi

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} = 0.1 \Rightarrow -\frac{1}{2}z^2 = \ln\left(0.1\sqrt{2\pi}\right) \Rightarrow z = \pm\sqrt{-2\ln\left(0.1\sqrt{2\pi}\right)} = \pm 1.6635... .$$

20. Kako je $A = 500, V = 0.1 \Rightarrow \sigma = A\cdot V = 50$, po normalnoj distribuciji (slika 27) je: (a) 1500 aparata (50%). (b) $450 = A - \sigma \Rightarrow 50 + 68.25/2 = 84.125\%$ što iznosi 2524 aparata, (c) $600 = A + 2\sigma \Rightarrow 50 - 95.45/2 = 2.275\%$ što iznosi 68 aparata.

MJERE OBLIKA DISTRIBUCIJE

Dok mjere raspršenosti pokazuju veličinu odstupanja (udaljenosti) podataka od neke srednje vrijednosti, mjere oblika distribucije podataka pokazuju način rasporeda podataka oko te vrijednosti. Pri tome se najčešće uzima aritmetička srednja vrijednost. Mjere oblika distribucije izražavaju se pomoću *momenata* numeričkog niza.

DEFINICIJA 6. NEKA SU ZADANI $c \in R$ I $r \in \{0,1,2,3,..\}$. ZA NUMERIČKI NIZ $x_1, x_2, ..., x_{n-1}, x_n$ JE

$$m_r(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^r \qquad (r - \text{ti moment oko konstante } c).$$

ZA $c = A$ (ARITMETIČKA SREDINA) IMAMO GLAVNI ILI CENTRALNI MOMENT $\mu_r = m_r(A)$, DOK ZA $c \neq A$ IMAMO POMOĆNE MOMENTE.

Ako su podaci grupirani kao distribucija frekvencija (13), imamo

$$m_r(c) = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - c)^r, \quad n = f_1 + f_2 + \ldots + f_k .$$

Vidimo da je $r-$ti moment definiran kao aritmetička sredina $r-$tih potencija odstupanja podataka od zadanog podatka c. Primijetimo da je nulti moment oko bilo koje vrijednosti jednak 1 ($m_0(c) = 1$), prvi pomoćni moment oko nule je aritmetička sredina ($m_1(0) = A$) dok je prvi glavni moment nula ($\mu_1 = 0$) a drugi glavni moment varijanca ($\mu_2 = \sigma^2$). Mjere oblika distribucije definiramo na slijedeći način.

DEFINICIJA 7. ZA NUMERIČKI NIZ $x_1, x_2, \ldots, x_{n-1}, x_n$ JE

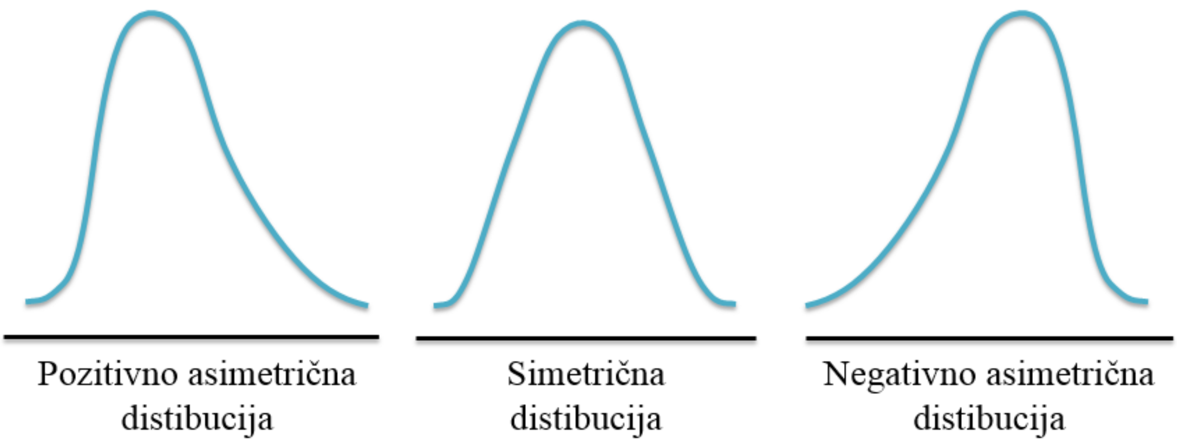
$$\alpha_3 = \frac{\mu_3}{\sigma^3} \quad \text{(koeficijent asimetrije)}, \qquad \alpha_4 = \frac{\mu_4}{\sigma^4} \quad \text{(koeficijent zaobljenosti)},$$

GDJE SU μ_3 I μ_4 GLAVNI MOMENTI A σ STANDARDNA DEVIJACIJA.

Navedene koeficijente možemo definirati i za sve ostale glavne momente,

$$\alpha_r = \frac{\mu_r}{\sigma^r}, \quad r = 0, 1, 2, \ldots .$$

Kako je $\alpha_0 = 1$, $\alpha_1 = 0$ i $\alpha_2 = 1$, α_3 je prvi koeficijent koji ukazuje na smjer asimetrije jer sadrži treće potencije odstupanja,. Ako dominiraju pozitivna odstupanja od srednje vrijednosti distribucija je *pozitivno asimetrična* ($\alpha_3 > 0$), u protivnom je *negativno asimetrična* ($\alpha_3 < 0$), a za simetričnu distribuciju nema asimetrije pa je $\alpha_3 = 0$ (slika 28).



Slika 28. Distribucije s obzirom na asimetriju

Vrijednost koeficijenta asimetrije α_3 općenito se interpretira na slijedeći način: ako je $|\alpha_3| > 1$ distribucija je jako asimetrična, za $0.5 \le |\alpha_3| \le 1$ umjereno asimetrična a za $|\alpha_3| < 0.5$ približno simetrična. Tip asimetrije kod unimodularnih distribucija neprekidnog obilježja možemo utvrditi i po međusobnom položaju srednjih vrijednosti: aritmetičke sredine A, medijana Me i moda Mo. Za simetričnu distribuciju one su međusobno jednake, za pozitivno asimetričnu je $A > Me > Mo$ a za negativno asimetričnu distribuciju vrijedi $A < Me < Mo$.

Koeficijent zaobljenosti α_4 ukazuje na oblik vrha i repova distribucije. Što je koeficijent veći (manji) to je vrh šiljastiji (širi) a repovi deblji (tanji). Obično se vrijednost koeficijenta zaobljenosti uspoređuje sa vrijednošću za neprekidnu normalnu distribuciju ($\alpha_4 = 3$).

Primijetimo da su navedeni koeficijenti neimenovani brojevi pa ih možemo koristiti za usporedbu oblika distribucije nizova podataka različitih obilježja i mjernih jedinica.

ZADACI

1. Odredite koeficijent asimetrije i zaobljenosti za podatke iz primjera 11.
2. Odredite prvi, drugi, treći, četvrti i peti moment oko nule za niz od prvih pet prirodnih brojeva.
3. Za materijalnu točku mase m na udaljenosti d od neke fiksne osi definira se statički moment, $M = md$ i moment inercije ili tromosti, $I = md^2$. Za više točaka ukupni moment dobije se zbrajanjem pojedinačnih. Kakva je veza ovih momenata sa momentima iz definicije 6 ? Ako su zadane točke $A(3,4)$, $B(0,6)$ i $C(8,2)$ čije su mase $m(A) = 40$, $m(B) = 100$ i $m(C) = 50$, odredite statički moment i moment inercije skupa $\{A, B, C\}$ u odnosu na svaku koordinatnu os (os x i os y).
4. Na temelju definicije iz prethodnog zadatke odredite statički moment i moment inercije kvadrata oko osi koja prolazi središtem kvadrata okomito na ravninu kvadrata ako je njegova masa koncentrirana ravnomjerno na njegovim vrhovima.
5. Za niz od p jedinica i q nula odredite pomoćne momente oko nule i jedinice.
6. Odredite aritmetičku sredinu, varijancu i standardnu devijaciju niza iz prethodnog zadatka.
7. Izrazite drugi i treći glavni moment pomoću pomoćnih momenata.
8. Eksperiment „zbrojite brojeve na tri istovremeno bačene kocke" izvedite u parovima ili grupama barem 20 puta. Svaka grupa neka tabelarno i grafički prikaže dobivene rezultate te izračuna aritmetičku sredinu, standardnu devijaciju, koeficijent asimetrije i zaobljenosti. Zatim to isto napravite sumarno za čitav razred (eksperiment izvesti bar 200 puta) te usporedite rezultate sa teorijskom distribucijom za ovaj eksperiment.
9. Svaki od vas neka zasebno procijeni težinu predmeta kojeg ste izabrali (bez vaganja). Zatim izvažite predmet te ispitajte odstupanja vaših procjena od stvarne težine tako da izračunate prvi i drugi moment.
10. Odredite koeficijent asimetrije i zaobljenosti za niz x_1, x_2.
11. Što možemo zaključiti o distribuciji neprekidnog obilježja za koju je: (a) $A = 56$, $Mo = 62$, (b) $Mo = 8$, $Me = 9$?

12. Kada je distribucija ocjena iz nekog testa pozitivno odnosno negativno asimetrična?

RJEŠENJA

1. Labin: $\alpha_3 = 0.33884\ldots$, $\alpha_4 = 1.87119\ldots$, Ivanec: $\alpha_3 = 0$, $\alpha_4 = 1.505846\ldots$. Vidimo da je distribucija temperatura tokom promatranog tjedna u Ivancu simetrična a u Labinu približno simetrična sa blagom pozitivnom asimetrijom. Zaobljenost distribucije za Ivanec je manja (spljoštenost veća) što znači da su podaci u širem pojasu oko srednje vrijednosti u odnosu na podatke za Labin.
2. Iz $m_r(0) = \frac{1}{n}\sum_{i=1}^{n} x_i^r = \frac{1}{5}\sum_{i=1}^{5} i^r$, $r = 1,2,3,4,5$, slijedi $m_1(0) = 3$, $m_2(0) = 11$, $m_3(0) = 45$, $m_4(0) = 195.8$ i $m_5(0) = 885$.
3. Prosječni statički moment je prvi a prosječni moment inercije je drugi moment oko nule pri čemu je masa predstavlja frekvenciju a udaljenost od osi vrijednost podatka. Za zadani skup točaka ukupni momenti oko osi x su $M = 40\cdot 4 + 100\cdot 6 + 50\cdot 2 = 840$ i $I = 40\cdot 4^2 + 100\cdot 6^2 + 50\cdot 2^2 = 4440$, a oko osi y su $M = 40\cdot 3 + 100\cdot 0 + 50\cdot 8 = 520$ i $I = 40\cdot 3^2 + 100\cdot 0^2 + 50\cdot 8^2 = 3560$.
4. Ako je m masa kvadrata i a duljina njegove stranice tada svaki vrh ima masu $m/4$ i udaljen je od osi $a/\sqrt{2}$. Traženi momenti su

$$M = 4\cdot\frac{m}{4}\cdot\frac{a}{\sqrt{2}} = \frac{ma}{\sqrt{2}}, \qquad I = 4\cdot\frac{m}{4}\cdot\left(\frac{a}{\sqrt{2}}\right)^2 = \frac{ma^2}{2}.$$

5. $$m_r(0) = \frac{1}{p+q}\left[p\cdot(1-0)^r + q\cdot(0-0)^r\right] = \frac{p}{p+q}, \quad r = 1,2,3,\ldots,$$

$$m_r(1) = \frac{1}{p+q}\left[p\cdot(1-1)^r + q\cdot(0-1)^r\right] = \begin{cases} \dfrac{-q}{p+q}, & r = 1,3,5,7,\ldots, \\ \dfrac{q}{p+q}, & r = 2,4,6,8,\ldots. \end{cases}$$

6. Imamo $A = (p\cdot 1 + q\cdot 0)/(p+q) = p/(p+q)$, pa iz relacije (14) slijedi

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - A^2 = \frac{1}{p+q}\cdot p - \left(\frac{p}{p+q}\right)^2 = \frac{pq}{(p+q)^2} \quad \Rightarrow \quad \sigma = \frac{\sqrt{pq}}{p+q}.$$

7. Za zadani c imamo $A - c = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} c = \frac{1}{n}\sum_{i=1}^{n}(x_i - c) = m_1$, pa je

$$\begin{aligned} \mu_2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - c) + (c - A)\right]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 + 2(c-A)\frac{1}{n}\sum_{i=1}^{n}(x_i - c) + \frac{1}{n}\sum_{i=1}^{n}(c - A)^2 \\ &= m_2 + 2(-m_1)m_1 + (c-A)^2 = m_2 - 2m_1^2 + (-m_1)^2 = m_2 - m_1^2. \end{aligned}$$

Na sličan način dobivamo

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^3 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - c) + (c - A)\right]^3$$
$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^3 + 3(c-A)\frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 + 3(c-A)^2\frac{1}{n}\sum_{i=1}^{n}(x_i - c) + \frac{1}{n}\sum_{i=1}^{n}(c - A)^3$$
$$= m_3 + 3(-m_1)m_2 + 3(-m_1)^2 m_1 + (-m_1)^3 = m_3 - 3m_1 m_2 + 2m_1^3.$$

Analogno se mogu izvesti formule za μ_4, μ_5 itd.

8. Koristiti stvarne podatke. Teorijska normalna distribucija za zbrojeve 3,4,5,…,17,18 je

$$1, 3, 6, 10, 15, 21, 25, 27, 27, 25, 21, 15, 10, 6, 3, 1$$

od ukupno $6^3 = 216$ mogućnosti (zbroj 3 možemo dobiti na 1 način 1+1+1, zbroj 4 na tri načina 1+1+2, 1+2+1, 2+1+1 itd.). Za ovu distribuciju je

$$A = 10.5, \quad \sigma = \sqrt{\frac{1890}{216}} \approx 2.958, \quad \alpha_3 = 0, \quad \alpha_4 = \frac{\frac{42619.5}{216}}{\sigma^4} \approx 2.577.$$

9. Koristite konkretne podatke vaših procjena i stvarne težine predmeta.
10. Imamo: $A = (x_1 + x_2)/2$, $\sigma = |x_1 - x_2|/2$, $\mu_3 = 0$, $\mu_4 = (x_1 - x_2)^4/16$, iz čega slijedi $\alpha_3 = 0$, $\alpha_4 = 1$.
11. (a) Negativno asimetrična sa $56 < Me < 62$. (b) Pozitivno asimetrična sa $A > 9$.
12. Ako prevladavaju slabije ocjene pozitivno asimetrična. Ako prevladavaju bolje ocjene negativno asimetrična.

MJERE KONCENTRACIJE

U praksi je vrlo često važno poznavati kao je promatrana pojava raspoređena po pojedinim jedinicama obilježja, npr. kako su raspoređena prirodna bogatstva, kapital, industrija, poljoprivredne površine, gdje je najveća naseljenost, promet, zagađenost itd. Upravo su mjere koncentracije pokazatelji (ne)ravnomjernosti spomenutog rasporeda. Ako je veći dio pojave raspoređen na malo jedinica kažemo da je koncentracija velika i mjere koncentracije su blizu svojih maksimalnih vrijednosti.

Tako na primjer u 2015. godini 85 najbogatijih ljudi svijeta posjedovalo je toliko bogatstva kao i 3.5 milijardi najsiromašnijih ljudi svijeta (pola svjetske populacije). Ako je raspored pojave približno ravnomjeran mjere koncentracije su blizu svojih minimalnih vrijednosti, npr. svaki učenik u razredu koristi za sjedenje jednu stolicu. Upoznat ćemo najčešće korištene mjere: *koncentracijski omjer* i *Ginijev koeficijent koncentracije.*

DEFINICIJA 8. ZA NUMERIČKI NIZ $x_1 \geq x_2 \geq \ldots \geq x_{n-1} \geq x_n \geq 0, \quad x_1 \neq 0$ JE

$$C_r = \frac{\sum_{i=1}^{r} x_i}{\sum_{i=1}^{n} x_i} = \frac{1}{n}\sum_{i=1}^{r}\frac{x_i}{A}, \quad r = 1, 2, \ldots, n \qquad (r-\text{ti koncentracijski omjer}),$$

GDJE JE A ARITMETIČKA SREDINA NIZA.

Vidimo da se koncentracijski omjer izračunava iz monotono padajućeg niza i pokazuje udio prvih r najvećih članova niza u cjelokupnom totalu (zbroju numeričkih vrijednosti niza). Iz definicije slijedi da je $\frac{r}{n} \leq C_r \leq 1$. Što je C_r bliže $\frac{r}{n}$ raspored je ravnomjerniji, bliže je rasporedu $x_1 = x_2 = \ldots = x_n$. Vrijednost $\tilde{C}_{n-r} = 1 - C_r$ pokazuje udio $n-r$ najmanjih članova niza u totalu (odnosno $\tilde{C}_r = 1 - C_{n-r}$ udio r najmanjih članova).

Ako je niz zadan kao distribucija frekvencija,

$$(x_1, f_1), (x_2, f_2), \ldots, (x_k, f_k), \quad x_1 \geq x_2 \geq \ldots \geq x_{k-1} \geq x_k \geq 0, \ x_1 \neq 0, \quad f_1 + f_2 + \ldots f_k = n,$$

tada, uz navedeni izraz koji vrijedi za svaki $r = 1, 2, \ldots, n$, izraz za $s-$ti koncentracijski omjer prvih s najvećih grupiranih vrijednosti C_s' sada poprima oblik

$$C_s' = C_{f_1 + f_2 + \ldots + f_s} = \frac{\sum_{i=1}^{s} f_i x_i}{\sum_{i=1}^{k} f_i x_i} = \frac{1}{n}\sum_{i=1}^{s}\frac{f_i x_i}{A}, \qquad s = 1, 2, \ldots, k.$$

Ako su podaci grupirani u razrede za x_i uzimamo razredne sredine.

Primjer 12. Prema podacima za 2015. godinu 1% najbogatijeg svjetskog stanovništva posjedovao je 43% svjetskog bogatstva, slijedećih 19% stanovništva 51% bogatstva dok je preostalih 80% posjedovalo samo 6% bogatstva. Analizirajmo ovu raspodjelu svjetskog bogatstva mjerama koncentracije.

Ako je C_r koncentracija svjetskog bogatstva na $r\%$ najbogatijeg stanovništva, imamo

$$C_1 = \frac{43}{100} = 0.43, \quad C_{20} = \frac{43+51}{100} = 0.94, \quad C_{100} = \frac{43+51+6}{100} = 1.$$

Primijetimo da bi ravnomjerni raspored bio za $C_1 = 0.01$ i $C_{20} = 0.2$. Vrijednost $\tilde{C}_r = 1 - C_{n-r}$ pokazuje koncentraciju svjetskog bogatstva na $r\%$ najsiromašnijeg stanovništva, pa je

$$\tilde{C}_{80} = 1 - C_{20} = 0.06, \qquad \tilde{C}_{99} = 1 - C_1 = 0.57.$$

Definiciju Ginijevog koeficijenta koncentracije za monotono rastući niz objasnit ćemo pomoću grafičkog prikaza (Lorenzove krivulje). Promatramo općeniti slučaj distribucije frekvencija,

$$(x_1, f_1), (x_2, f_2), \dots, (x_k, f_k), \quad 0 \le x_1 \le x_2 \le \dots \le x_{k-1} \le x_k, \ x_k \ne 0, \quad f_1 + f_2 + \dots f_k = n.$$

Pojedinačni negrupirani podaci su samo specijalni slučaj ovog niza gdje je $f_i = 1, \ i = 1, 2, \dots, k$ pa je $k = n$. U koordinatnom sustavu prikazujemo točke $(P_i, X_i), \ i = 1, 2, \dots, k$ gdje su P_i relativne kumulativne frekvencije podataka (os apscisa) a X_i relativna kumulativna vrijednost niza – relativni udjeli podtotala u totalu (os ordinata),

$$P_i = \frac{1}{n}\sum_{j=1}^{i} f_j = \sum_{j=1}^{i} p_j, \qquad X_i = \frac{\sum_{j=1}^{i} f_j x_j}{\sum_{j=1}^{k} f_j x_j} = \frac{\frac{1}{n}\sum_{j=1}^{i} f_j x_j}{\frac{1}{n}\sum_{j=1}^{k} f_j x_j} = \frac{\sum_{j=1}^{i} p_j x_j}{A}, \qquad i = 1, 2, \dots, k,$$

gdje je A aritmetička sredina polaznog niza. Dakle, postupak je slijedeći. Iz zadanog niza frekvencija $f_1, f_2, \dots, f_k$ formiramo niz relativnih frekvencija $p_1 = \frac{f_1}{n}, p_2 = \frac{f_2}{n}, \dots, p_k = \frac{f_k}{n}$ te zatim njegovu kumulantu

$$P_1 = p_1, \ P_2 = p_1 + p_2, \ P_3 = p_1 + p_2 + p_3, \ \dots., \ P_k = p_1 + p_2 + \dots + p_k = 1.$$

Slično formiramo niz agregata $f_1 x_1, f_2 x_2, \dots, f_k x_k$, odredimo total (zbrojimo niz agregata) $T = \sum_{j=1}^{k} f_j x_j$ te formiramo niz relativnih agregata $\frac{f_1 x_1}{T}, \frac{f_2 x_2}{T}, \dots, \frac{f_k x_k}{T}$ i njegovu kumulantu

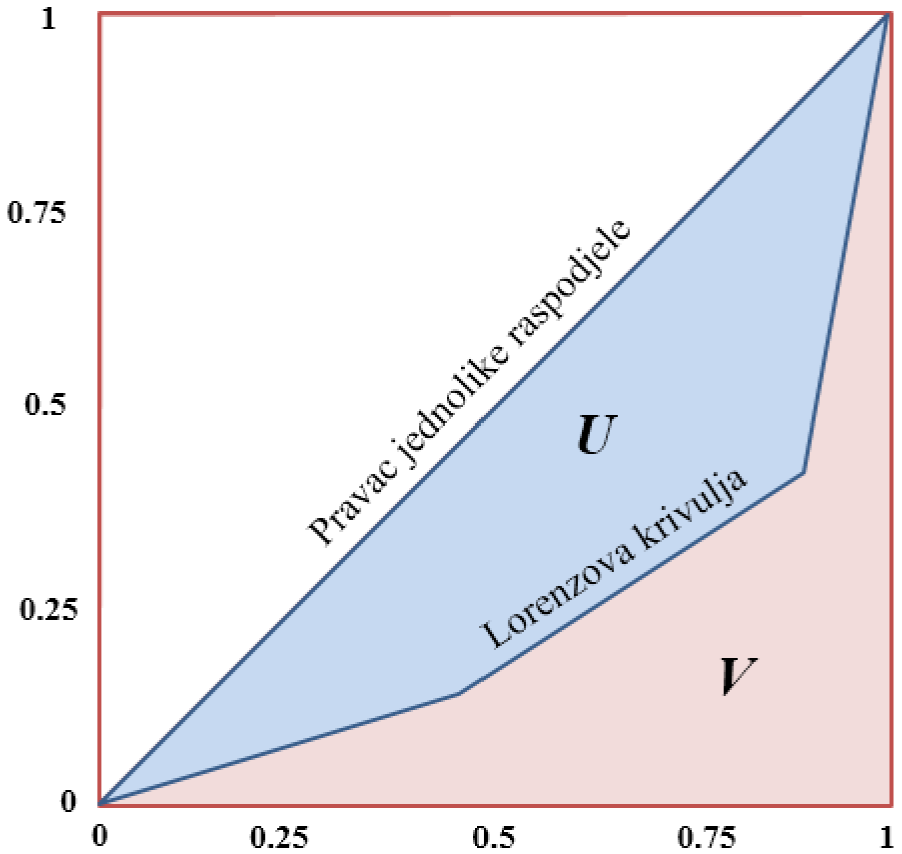
$$X_1 = \frac{f_1 x_1}{T}, \ X_2 = \frac{f_1 x_1}{T} + \frac{f_2 x_2}{T}, \ X_3 = \frac{f_1 x_1}{T} + \frac{f_2 x_2}{T} + \frac{f_3 x_3}{T}, \ \dots, \ X_k = \frac{f_1 x_1}{T} + \frac{f_2 x_2}{T} + \dots + \frac{f_k x_k}{T} = 1.$$

U slučaju niza negrupiranih pojedinačnih podataka u sve gore navedene izraze uvrstimo $f_1 = f_2 = \dots = f_k = 1$ i $k = n$, pa je tako na primjer niz relativnih frekvencija $\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}$, a kumulanta $P_1 = \frac{1}{n}, \ P_2 = \frac{2}{n}, \ P_3 = \frac{3}{n}, \ \dots, \ P_k = P_n = \frac{n}{n} = 1$.

Tako dobivene točke prikazujemo u koordinatnom sustavu i spajamo poligonalnom linijom

$$(P_0, X_0) = (0,0), \ (P_1, X_1), \ (P_2, X_2), \ \dots, \ (P_{k-1}, X_{k-1}), \ (P_k, X_k) = (1,1),$$

koju nazivamo Lorenzova krivulja (slika 29). Ona uvijek počinje u točki (0,0) a završava u (1,1). Ako je raspored podataka ravnomjeran (svi x_j su međusobno jednaki) krivulja se poklopi sa dužinom koja spaja (0,0) i (1,1) (pravac jednolike raspodjele).



Slika 29. Lorenzova krivulja

U protivnom, kako su podaci dani u monotono rastućem redoslijedu, Lorenzova krivulja je uvijek ispod te dužine (konveksna krivulja) i pripada pravokutnom trokutu (0,0), (1,0), (1,1). Što je krivulja dalje od pravca jednolike raspodjele (hipotenuze trokuta), tj. što je površina između njih veća, veća je neravnomjernost raspodjele. Ginijev koeficijent koncentracije G definira se kao dio koji ta površina (U) zauzima u navedenom trokutu (omjer te površine i površine trokuta). Kako je površina trokuta $U+V=0.5$, imamo

$$G=\frac{U}{0.5}=\frac{0.5-V}{0.5}=1-2V.$$

Površinu V možemo izraziti pomoću prikazanih točaka zbrajanjem površina trapeza sa vrhovima $(P_{i-1},0)$, $(P_i,0)$, (P_i,X_i), (P_{i-1},X_{i-1}), $i=1,2,...,k$. Kako je visina navedenog trapeza P_i-P_{i-1} a osnovice su X_i i X_{i-1}, njegova površina je $(X_i+X_{i-1})(P_i-P_{i-1})/2$ pa imamo

$$G=1-\sum_{i=1}^{k}(X_i+X_{i-1})(P_i-P_{i-1})=1-\sum_{i=1}^{k}p_i(X_i+X_{i-1}),\qquad X_0=P_0=0, \tag{16}$$

jer je $P_i-P_{i-1}=\frac{f_i}{n}=p_i$. Pri tome je $X_i+X_{i-1}=2X_{i-1}+\frac{f_ix_i}{T}=2X_i-\frac{f_ix_i}{T}$. Formula (16) može se i dalje transformirati (vidjeti zadatke 4 i 5).

Prema definiciji G je uvijek između 0 (jednolika raspodjela) i 1 (ekstremno nejednolika teorijska raspodjela – cjelokupna pojava je koncentrirana u jednoj točki) pa se rezultat može izraziti i u postocima. Ako su podaci grupirani u razrede, za vrijednosti x_i, kao i do sada, uzimamo razredne sredine.

Ponekad se u praksi koristi i *normirani Ginijev koeficijent koncentracije,*

$$G' = \frac{G}{G_{\max}},$$

gdje je $G_{\max}$ najveći mogući koeficijent koncentracije za promatranu distribuciju. On može biti determiniran praktičnim ograničenjima pojave koju promatramo ili se uzima teorijska vrijednost (vidjeti zadatke 1 i 3).

U primjeru 12 su već zadane relativne vrijednosti pa samo treba urediti niz u rastućem poretku (od najsiromašnijih). Kako je niz relativnih frekvencija $0.8, 0.19, 0.01$ a niz relativnih totala $0.06, 0.51, 0.43$, kumulativni nizovi su $0.8, 0.99, 1$ odnosno $0.06, 0.57, 1$, pa imamo: $(P_0, X_0) = (0,0)$, $(P_1, X_1) = (0.8, 0.06)$, $(P_2, X_2) = (0.99, 0.57)$ i $(P_3, X_3) = (1,1)$. Dakle,

$$G = 1 - \left[0.8 \cdot (0.06 + 0) + 0.19 \cdot (0.57 + 0.06) + 0.01 \cdot (1 + 0.57)\right] = 0.8166 = 81.66\%.$$

Vidimo izrazito značajnu nejednolikost u rasporedu svjetskog bogatstva (visoku koncentraciju bogatstva u pojedinim jedinicama – centrima financijske moći).

Kako se Lorenzova krivulja vrlo često koristi za analizu raspodjele novčanih sredstava (dohotka, fondova, proračuna i sl.) navodimo i slijedeći primjer.

Primjer 13. U jednoj firmi 6 radnika prima prosječnu plaću od 3600 kuna, 30 radnika 5500 kuna, 60 radnika 7800 kuna, 18 radnika 11200 kuna i 6 radnika 17300 kuna. Prikažimo ovu raspodjelu Lorenzovom krivuljom i odredimo koeficijent koncentracije.

U skladu sa navedenim postupkom formiramo tabelu.

i	x_i	f_i	p_i	P_i	$f_i x_i$
1	3600	6	0.05	0.05	21600
2	5500	30	0.25	0.3	165000
3	7800	60	0.5	0.8	468000
4	11200	18	0.15	0.95	201600
5	17300	6	0.05	1	103800
Σ		n=120	1		T=960000

i	$f_i x_i / T$	X_i	$X_i + X_{i-1}$	$p_i(X_i + X_{i-1})$
1	0.0225	0.0225	0.0225	0.001125
2	0.171875	0.194375	0.216875	0.05421875
3	0.4875	0.681875	0.87625	0.438125
4	0.21	0.891875	1.57375	0.2360625
5	0.108125	1	1.891875	0.09459375
Σ	1			0.824125

Tabela 11. Podaci za Lorenzovu krivulju i indeks koncentracije

Koristeći relaciju (16) dobivamo $G = 1 - 0.824125 = 0.175875 \approx 17.6\%$ što pokazuje primjeren stupanj ravnomjernosti rasporeda plaća u toj firmi. Lorenzovu krivulju konstruiramo spajanjem točaka (P_i, X_i), $i = 0,1,2,3,4,5$.

ZADACI

1. Ako se na nekom natjecanju dodjeljuje jedna nagrada a ima n natjecatelja, koliki je koncentracijski omjer i koeficijent koncentracije? Što se događa za $n \to \infty$?
2. Ako se na nekom natjecanju dodjeljuju tri nagrade: prva 5000, druga 3000 i treća 2000 kuna, a ima n natjecatelja, odredite koeficijent koncentracije. Konkretizirajte rezultat za $n = 10, 100, 1000$.
3. Odredite koeficijent koncentracije podjele jednog iznosa na dva dijela u omjeru $a:b$, $a \le b$. U kojim granicama se nalazi rezultat ovisno o vrijednostima a i b? Ako vaš brat dobije duplo veći džeparac od vas, koliki je normirani koeficijent koncentracije?
4. Dokažite da se izraz (16) može također pisati u obliku

$$G = \sum_{i=2}^{k} (P_{i-1} X_i - P_i X_{i-1}) \,.$$

5. Ako je $f_1 = f_2 = \ldots = f_k = 1$ dokažite da se izraz (16) može pisati u obliku

$$G = \frac{2}{nT} \sum_{i=1}^{n} (i \cdot x_i) - \frac{1}{n} - 1, \qquad T = \sum_{i=1}^{n} x_i \,.$$

   Provjerite ovu formulu na zadacima 1, 2 i 3.
6. U razredu formirajte grupe 3-6 učenika slučajnim odabirom. Zamislite da je svaka grupa otišla na izlet u prirodu. Članovi grupe su se razdvojili i jedan od njih je pronašao 10000 EUR a da ostali to nisu znali. Ako ste vi taj član, da li bi i kako bi podijelili iznos sa ostalim članovima grupe? Odredite koeficijent koncentracije vaše podjele te ga usporedite sa koeficijentom podjele ostalih članova grupe.
7. Za raspodjelu $10 = 1 + 2 + 3 + 4$ odredite koncentracijske omjere, koeficijent koncentracije te nacrtajte Lorenzovu krivulju.

8. Na jednom međunarodnom sportskom natjecanju dodijeljeno je 30 medalja. Ako je od 10 zemalja koje su sudjelovale na natjecanju jednoj zemlji pripala jedna medalja, dvije zemlje dobile su po dvije medalje, tri zemlje po tri medalje i četiri zemlje po četiri medalje, nacrtajte Lorenzovu krivulju i odredite koeficijent koncentracije za ovo natjecanje.
9. Ako su tri osobe zajedno dobile 30000 kuna a raspodjela je opisana Lorenzovom krivuljom kroz točke (0,0), (1/3,1/6), (2/3,1/2), (1,1), koji iznos je dobila svaka osoba? Koliki je koeficijent koncentracije?
10. Ispitajte koliki je mjesečni kućni budžet vaše obitelji te koliki dio potroši svaki član (zajedničke troškove – režije možete razdijeliti ravnomjerno). Nacrtajte Lorenzovu krivulju i odredite koeficijent koncentracije.
11. U jednu liječničku ordinaciju tokom prvog kvartala 2016. godine primljeno je 200 osoba oboljelih od gripe. Među njima je bilo 62 djece do 10 godina starosti, 39 mladih 10-20 godina, 31 odraslih 20-40 godina i 68 osoba 40-80 godina. Da li postoji značajnija koncentracija oboljelih po godinama starosti ili je gripa zahvatila sve dobi ravnomjerno?
12. Koliki je prvi koncentracijski omjer za prvih n prirodnih brojeva? Interpretirajte rezultat za $n = 99$. Poopćite rezultat na nenegativan aritmetički niz.
13. Koliki je prvi koncentracijski omjer za beskonačan niz 1, 1/2, 1/4, 1/8, 1/16, …? Poopćite rezultat na nenegativan konačan i beskonačan geometrijski niz.
14. U trgovini je na ponudi jedan proizvod u tri različite kvalitete, klasa I, II i III. Klasa II je za 25% skuplja od klase III a klasa I za 50% od klase III. Ako je prodano najmanje klase I a u odnosu na nju duplo više klase II i tri puta više klase III, odredite pripadne koncentracijske omjere prometa od navedene prodaje.
15. Odredite koeficijent koncentracije za prodaju u prethodnom zadatku.
16. Na tržište su plasirana četiri nova modela mobitela po cijenama od 1000, 2000, 4000 i 8000 kuna. Ako je prvog tjedna prodana jednaka količina svakog modela, odredite pripadne koncentracijske omjere i koeficijent koncentracije vrijednosti ove prodaje.

RJEŠENJA

1. Za računanje koncentracijskih omjera formiramo niz $x, 0, 0, ..., 0$, gdje je x vrijednost nagrade, pa je $C_r = 1$ za sve $r = 1, 2, ..., n$, što pokazuje potpunu koncentraciju pojave na samo jednu jedinicu obilježja. Za koeficijent koncentracije uzimamo niz $0, 0, ..., 0, x$ pa je $f_i = 1,\ p_i = 1/n,\ \ i = 1, 2, ..., n$, iz čega slijedi $G = 1 - \left(0 + 0 + ... + \frac{1}{n} \cdot \frac{x}{x}\right) = 1 - \frac{1}{n}$.

   Ovo je ujedno najveća teorijska koncentracija ($G_{\max}$) distribucije jednog iznosa na n dijelova. Za $n \to \infty$ imamo $G \to 1$ (teorijska koncentracija u jednoj točki).
2. Imamo niz $x_1 = x_2 = ... = x_{n-3} = 0$, $x_{n-2} = 2000$, $x_{n-1} = 3000$, $x_n = 5000$ i frekvencije $f_1 = f_2 = ... = f_n = 1$, pa je $X_1 = X_2 = ... = X_{n-3} = 0$, $X_{n-2} = 0.2$, $X_{n-1} = 0.2 + 0.3 = 0.5$, $X_n = 0.5 + 0.5 = 1$. Dakle,

$$G = 1 - \left[0 + 0 + ... + 0 + \frac{1}{n} \cdot 0.2 + \frac{1}{n} \cdot (0.5 + 0.2) + \frac{1}{n} \cdot (1 + 0.5)\right] = 1 - \frac{2.4}{n}.$$

Za $n = 10$ je $G = 0.76$, za $n = 100$ je $G = 0.976$ a za $n = 1000$ imamo $G = 0.9976$.

3. Ako je S iznos koji dijelimo, imamo: $x_1 = \frac{a}{a+b} S$, $x_2 = \frac{b}{a+b} S$, $f_1 = f_2 = 1 \Rightarrow$ $p_1 = p_2 = \frac{1}{2}$. Nadalje, $X_1 = \frac{a}{a+b}$, $X_2 = 1$, pa je

$$G = 1 - \left[\frac{1}{2} \cdot \frac{a}{a+b} + \frac{1}{2} \cdot \left(1 + \frac{a}{a+b}\right)\right] = \frac{b-a}{2(b+a)} .$$

Vidimo da je $G \geq G_{\min} = 0$ (za $a = b$) i $G \leq G_{\max} = 0.5$ (za $a = 0$) pa je normirani koeficijent

$$G' = \frac{G}{0.5} = \frac{b-a}{b+a} .$$

Za slučaj raspodjele džeparca je $a = 1$, $b = 2 \Rightarrow G' = 1/3 \approx 33\%$.

4. Imamo

$$\begin{aligned}
G &= 1 - \sum_{i=1}^{k} (X_i + X_{i-1})(P_i - P_{i-1}) = 1 - \sum_{i=1}^{k} (P_i X_i + P_i X_{i-1} - P_{i-1} X_i - P_{i-1} X_{i-1}) \\
&= 1 - \sum_{i=1}^{k} P_i X_i + \sum_{i=1}^{k} P_{i-1} X_{i-1} + \sum_{i=1}^{k} (P_{i-1} X_i - P_i X_{i-1}) = 1 - P_k X_k + \sum_{i=1}^{k} (P_{i-1} X_i - P_i X_{i-1}) \\
&= \sum_{i=1}^{k} (P_{i-1} X_i - P_i X_{i-1}) = \sum_{i=2}^{k} (P_{i-1} X_i - P_i X_{i-1}),
\end{aligned}$$

jer je $P_k = X_k = 1$ i $P_0 X_1 - P_1 X_0 = 0 \cdot X_1 - P_1 \cdot 0 = 0$.

5. Dokaz je složeniji pa je namijenjen boljim učenicima. Kako je $k = n$ te $p_i = 1/n$, $X_i + X_{i-1} = 2X_i - x_i / T$ za sve $i = 1, 2, ..., n$, imamo

$$G = 1 - \frac{1}{n} \sum_{i=1}^{n} (X_i + X_{i-1}) = 1 - \frac{1}{n}\left(2\sum_{i=1}^{n} X_i - \frac{1}{T}\sum_{i=1}^{n} x_i\right) = 1 - \frac{1}{n}\left(2\sum_{i=1}^{n} X_i - 1\right),$$

što uz $\sum_{i=1}^{n} X_i = \frac{1}{T}\left[n x_1 + (n-1)x_2 + (n-2)x_3 + ... + 1 x_n\right] = \frac{1}{T}\left[(n+1)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} i x_i\right]$, daje

$$\begin{aligned}
G &= 1 - \frac{1}{n}\left[\frac{2(n+1)}{T}\sum_{i=1}^{n} x_i - \frac{2}{T}\sum_{i=1}^{n} i x_i - 1\right] = 1 - \frac{1}{n}\left[\frac{2(n+1)}{T} \cdot T - \frac{2}{T}\sum_{i=1}^{n} i x_i - 1\right] \\
&= 1 - 2 - \frac{2}{n} + \frac{2}{nT}\sum_{i=1}^{n} i x_i + \frac{1}{n} = \frac{2}{nT}\sum_{i=1}^{n} i x_i - \frac{1}{n} - 1.
\end{aligned}$$

6. Koristite svoje podatke.

7. Za niz 4,3,2,1 je $C_1 = 0.4$, $C_2 = 0.7$, $C_3 = 0.9$ i $C_4 = 1$. Za niz 1,2,3,4 je $k = n = 4$ te $f_i = 1$, $p_i = 0.25$ za $i = 1,2,3,4$. Točke za Lorenzovu krivulju su: (0, 0), (0.25, 0.1), (0.5, 0.3), (0.75, 0.6), (1,1). Imamo $G = 1 - 0.25(0.1 + 0.4 + 0.9 + 1.6) = 0.25$.
8. Tu je x_i broj medalja a f_i broj zemalja koje su dobile x_i medalja. Imamo $x_i = f_i = i$ za $i = 1,2,3,4$. Računamo kao u tabeli 11. Za Lorenzovu krivulju dobijemo točke:

$$(0,0),\ \left(\frac{1}{10},\frac{1}{30}\right),\ \left(\frac{3}{10},\frac{5}{30}\right),\ \left(\frac{6}{10},\frac{14}{30}\right),\ (1,1),$$

pa je $G = 1 - \left(\frac{1}{10}\cdot\frac{1}{30} + \frac{2}{10}\cdot\frac{6}{30} + \frac{3}{10}\cdot\frac{19}{30} + \frac{4}{10}\cdot\frac{44}{30}\right) = 0.18$.
9. Prva osoba 1/6 iznosa (5000 kuna), druga $1/2 - 1/6 = 1/3$ iznosa (10000 kuna) i treća $1 - 1/2 = 1/2$ iznosa (15000 kuna). $G = 1 - \left(\frac{1}{3}\cdot\frac{1}{6} + \frac{1}{3}\cdot\frac{2}{3} + \frac{1}{3}\cdot\frac{3}{2}\right) = \frac{2}{9}$.
10. Koristite stvarne podatke.
11. Imamo razredne sredine: $x_1 = 5$, $x_2 = 15$, $x_3 = 30$ i $x_4 = 60$ te frekvencije $f_1 = 62$, $f_2 = 39$, $f_3 = 31$ i $f_4 = 68$. Dalje postupamo kao u tabeli 11. Dobijemo $n = 200$, $T = 5905$, pa je

$$G = 1 - \left(0.31\cdot\frac{310}{T} + 0.195\cdot\frac{1205}{T} + 0.155\cdot\frac{2720}{T} + 0.34\cdot\frac{7730}{T}\right) = \frac{2524.125}{5905} = 0.427...,$$

što pokazuje značajnu koncentraciju (neravnomjernost) oboljenja po starosnoj dobi.
12. Imamo niz $n, n-1, n-2, ..., 2, 1$. Kako je zbroj niza $S = n(n+1)/2$, bit će $C_1 = n/S = 2/(n+1)$. Za $n = 99$ je $C_1 = 0.02$ što znači da prvi, najveći član niza (99) čini 2% ukupnog zbroja svih članova niza (4950). Kako općenito zbroj aritmetičkog niza $a_1 \geq a_2 \geq ... \geq a_n \geq 0,\ a_1 \neq 0$ iznosi $S = \frac{n(a_1 + a_n)}{2}$, imamo $C_1 = \frac{2a_1}{n(a_1 + a_n)}$.
13. Za geometrijski niz $a_1 \geq a_2 \geq ... \geq a_n \geq 0,\ a_1 \neq 0$ kvocijent niza je $0 \leq q \leq 1$ (za $q = 1$ niz je konstantan) a zbroj $S = a_1\frac{1-q^n}{1-q}$, pa je $C_1 = \frac{a_1}{S} = \frac{1-q}{1-q^n}$ (za $q = 1$ je $C_1 = \frac{1}{n}$). Za beskonačni niz vrijedi $n \to \infty \Rightarrow q^n \to 0$ pa je $C_1 = 1 - q$. Za zadani niz je $a_1 = 1,\ q = 0.5$, pa je $C_1 = 0.5$ što znači da prvi član niza (1) čini 50% ukupnog zbroja svih članova niza (2).
14. Ako je x cijena proizvoda klase III tada je $1.25x$ cijena za klasu II i $1.5x$ za klasu I. Nadalje, ako je f količina prodanih proizvoda klase I tada je $2f$ količina prodane klase II a $3f$ klase III. Imamo $\sum_{i=1}^{k} f_i x_i = f\cdot 1.5x + 2f\cdot 1.25x + 3f\cdot x = 7fx$, pa je

$$C_1' = \frac{1.5\,fx}{7\,fx} = \frac{3}{14} \approx 21.4\%, \quad C_2' = \frac{1.5\,fx + 2.5\,fx}{7\,fx} = \frac{4}{7} \approx 57.1\%, \quad C_3' = 1 = 100\%.$$

15. Poredamo podatke za klase III, II i I te slijedimo postupak iz tabele 11. Dobijemo $p_1 = 1/2$, $p_2 = 1/3$, $p_3 = 1/6$, $X_1 = 3/7$, $X_2 = 5.5/7$, $X_3 = 1$, pa je

$$G = 1 - \left[\frac{1}{2}\cdot\frac{3}{7} + \frac{1}{3}\cdot\left(\frac{5.5}{7} + \frac{3}{7}\right) + \frac{1}{6}\cdot\left(1 + \frac{5.5}{7}\right)\right] = 1 - \frac{77}{84} = \frac{7}{84} \approx 8.3\%,$$

a što ukazuje na vrlo ravnomjernu vrijednost prometa po klasama proizvoda.

16. Imamo $f_1 = f_2 = f_3 = f_4 = f$. Za koncentracijske omjere koristimo redoslijed podataka $x_1 = 8000$, $x_2 = 4000$, $x_3 = 2000$, $x_4 = 1000$, pa je

$$C_1' = \frac{8000\,f}{15000\,f} = \frac{8}{15}, \quad C_2' = \frac{12000\,f}{15000\,f} = \frac{4}{5}, \quad C_3' = \frac{14000\,f}{15000\,f} = \frac{14}{15}, \quad C_4' = 1.$$

Za koeficijent koncentracije uzimamo $x_1 = 1000$, $x_2 = 2000$, $x_3 = 4000$, $x_4 = 8000$, pa je $p_1 = p_2 = p_3 = p_4 = 1/4$, $X_1 = 1/15$, $X_2 = 3/15$, $X_3 = 7/15$, $X_4 = 1$. Imamo

$$G = 1 - \left[\frac{1}{4}\cdot\frac{1}{15} + \frac{1}{4}\cdot\left(\frac{3}{15} + \frac{1}{15}\right) + \frac{1}{4}\cdot\left(\frac{7}{15} + \frac{3}{15}\right) + \frac{1}{4}\cdot\left(1 + \frac{7}{15}\right)\right] = 1 - \frac{37}{60} = \frac{23}{60} \approx 38.3\%.$$

VREMENSKI NIZOVI

Gotovo svaka pojava u prirodi i društvu protokom vremena se mijenja. Da bi uopće mogli registrirati bilo kakvu promjenu potrebno je pratiti pojavu kroz vrijeme. Tako na primjer pratimo cijenu nafte na svjetskom i domaćem tržištu, dionice na burzama, ponudu i potražnju, cijene u trgovinama, devizni tečaj, vremenske uvjete, temperaturu, tjelesnu težinu itd. Kronološkim bilježenjem i uređivanjem podataka za neko obilježje kroz vrijeme dobivamo *vremenski niz* koji može biti *trenutačni* (stock time) ili *intervalni* (flow time). Vrijednosti trenutačnog niza prikazuju stanje pojave u određenim vremenskim točkama i općenito nemaju svojstvo kumulativnosti, nema ih smisla zbrajati, npr. temperatura, stanje na računu, trenutna cijena, broj učenika u razredu i sl. Intervalni niz dobije se zbrajanjem vrijednosti pojave u čitavom intervalu promatranja pa prirodno ima svojstvo kumulativnosti, npr. broj turista po mjesecima, dnevni promet, mjesečna plaća, tjedna proizvodnja itd. Intervali promatranja najčešće su međusobno jednaki ali mogu biti i različiti. Vidimo da vremenski nizovi nisu ništa novo. Radi se samo o specijalnom slučaju statističkih nizova koji su kronološki uređeni. Sve metode za analizu koje smo do sada upoznali možemo primijeniti i na vremenske nizove (neke smo već i ranije analizirali – primjer 11). Da bi rezultati statističke analize vremenskih nizova bili što pouzdaniji potrebno je obraditi podatke dobivene sistematskim praćenjem pojave kroz duže vrijeme. Zbog njihove važnosti i zastupljenosti upoznajemo ih još detaljnije.

Vremenski niz se uređuje kronološki, redovito u pozitivnom smjeru po vremenskoj osi a vrijednosti niza su najčešće frekvencije pojedinih oblika obilježja. Vrijednosti vremenskih nizova mogu sadržavati više vremenskih komponenti: *trend, cikličku, sezonsku,* i *slučajnu* komponentu. Trend komponenta odražava osnovni tok pojave u vremenu, tj. način kako pojava zavisi o vremenskom toku (linearni, kvadratni, obrnuto proporcionalni, eksponencijalni i sl. trend). Ciklička komponenta očituje se u ponavljanju istih ili vrlo sličnih efekata u vremenskom tijeku pojave sa jednakim periodom (ciklusi). Ciklusi mogu imati više razina pa tako unutar glavnih ciklusa možemo imati manje podcikluse, unutar njih nove itd. Ako se slični efekti pojavljuju u istom dijelu ciklusa (istom dijelu dana ili istom dobu godine) govorimo o sezonskoj komponenti. Ove tri komponente u pravilu se mogu izraziti i opisati nekom funkcijom vremena pa spadaju u *sistematske* ili *determinističke* vremenske komponente. Slučajna komponenta odražava nepravilna odstupanja koja nisu predvidiva a nastaju zbog iznenadnih slučajnih ili nepoznatih faktora (rezidualna odstupanja). Ova komponenta spada u *nesistematske* ili *stohastičke* komponente. Vremenska pojava može sadržavati jednu, dvije ili više navedenih komponenti.

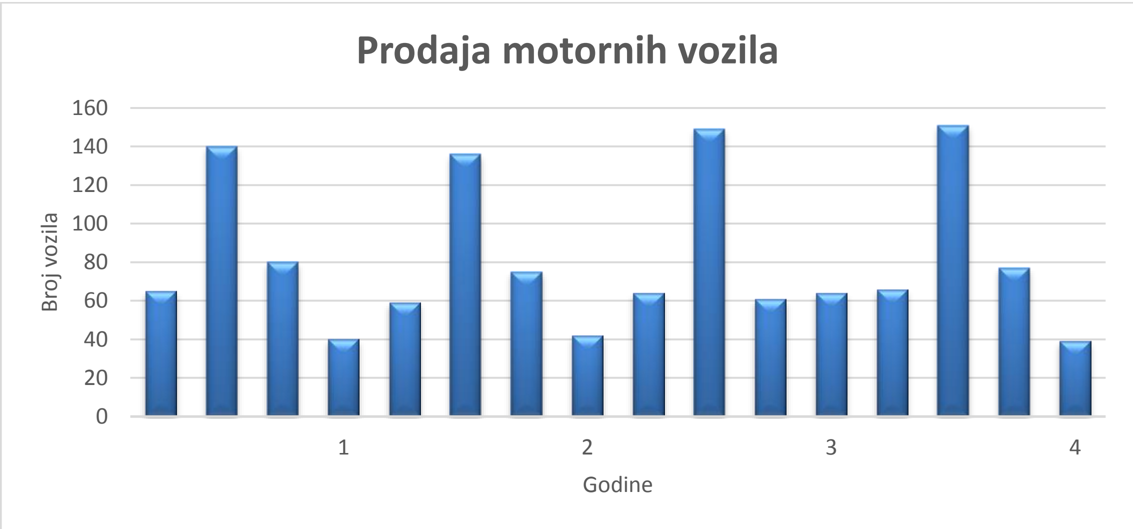
Uređeni vremenski nizovi se prikazuju popisivanjem, tablično i grafički. Za grafički prikaz intervalnih nizova možemo koristiti linijske i površinske dijagrame (histograme) a za trenutačne samo linijske dijagrame. Pri tome, ako za intervalne nizove koristimo linijske dijagrame, vrijednosti se nanose na početku, na kraju ili negdje unutar intervala (npr. u sredini) ovisno o zadanim podacima i prirodi pojave koju prikazujemo. Isto tako, ako su duljine vremenskih intervala međusobno različite, frekvencije kod intervalnih nizova treba korigirati na fiksnu duljinu intervala jer se one odnose na čitav interval, dok kod trenutačnih nizova ne treba budući da one pokazuju stanje pojave u jednom vremenskom trenutku. Sve ove dijagrame već smo ranije upoznali. Za grafički prikaz pojava sa izraženom sezonskom ili cikličkom komponentom prikladni su i *polarni dijagrami* koje ćemo upoznati kroz primjere.

Primjer 14. Prodaja motornih vozila na promatranom tržištu tokom posljednje četiri godine po kvartalima navedena je u tabeli 12. Analizirajmo prodaju grafički. Koje komponente prodaje uočavamo?

Godina	I	II	III	IV	Ukupno
2012.	65	140	80	40	325
2013.	59	136	75	42	312
2014.	64	149	61	64	338
2015.	66	151	77	39	333

Tabela 12. Pregled prodaje motornih vozila po godinama i kvartalima

Prodaju motornih vozila prikazali smo površinskim dijagramom na slici 30 gdje smo pojednostavnili numeraciju godina (1., 2., 3. i 4. godina promatranja pojave). Na slici možemo uočiti cikličku komponentu prodaje po godinama te sezonsku po pojedinim kvartalima. Izuzetak je druga polovica treće (2014.) godine gdje je sezonski trend narušen pa imamo slučajnu komponentu. Tako u trećem kvartalu te godine vidimo znatno smanjenu prodaju u odnosu na sezonski trend iz prijašnjih godina (poremećaji na tržištu, možda rast cijena, pad standarda ili sl.) a u četvrtom nagli porast. Općenito se statističkom analizom detektiraju takve situacije koje se onda mogu pobliže analizirati u svrhu otklanjanja neželjenih i intenziviranja poželjnih efekata.

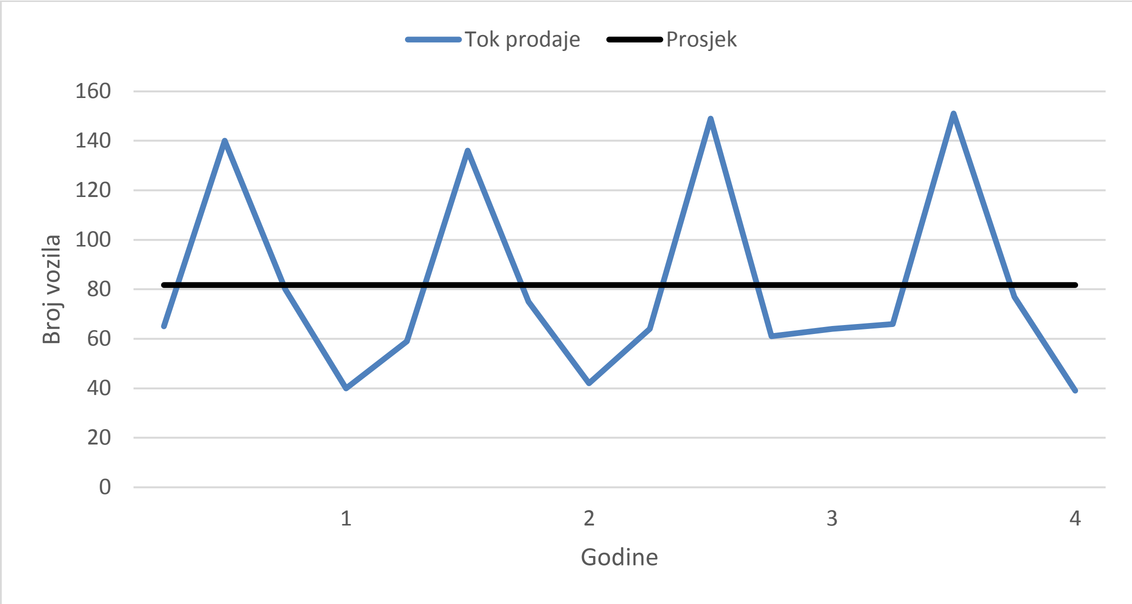


Slika 30. Površinski dijagram prodaje motornih vozila

Na slici 31 prikazan je linijski dijagram promatrane prodaje na kojem su te komponente još uočljivije. Prosječan broj prodanih vozila na godišnjoj razini (aritmetička sredina) iznosi A=(325+312+338+333)/4=327 (na kvartalnoj 327/4=81.75) što je također prikazano na slici 31. Analiziramo li odstupanja prodaje po godinama mjerama raspršenosti, dobivamo

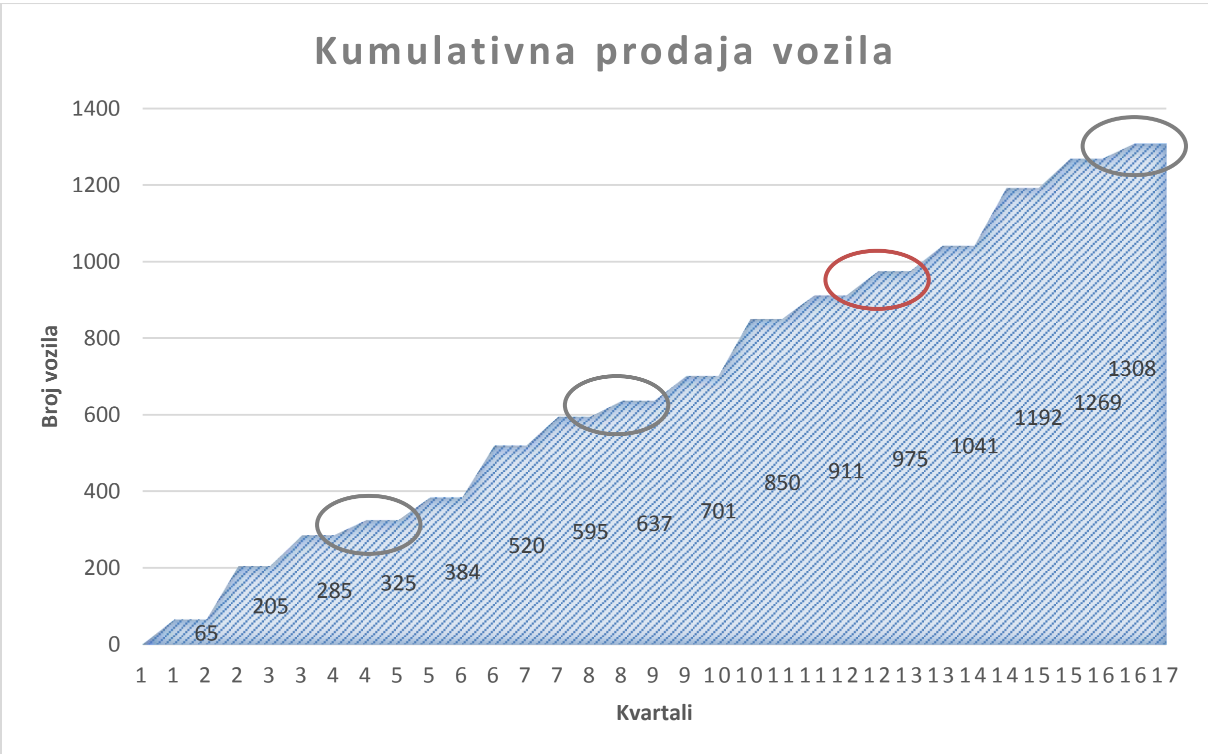
$$\sigma=\sqrt{\frac{(-2)^2+(-15)^2+11^2+6^2}{4}}=9.82344..., \qquad V=\frac{\sigma}{A}=0.030041...\approx 3\%.$$

Vidimo da je koeficijent varijacije izuzetno malen pa aritmetička sredina vrlo dobro reprezentira prodaju na godišnjoj razini i predstavlja konstantnu trend komponentu prodaje.



Slika 31. Linijski dijagram prodaje motornih vozila

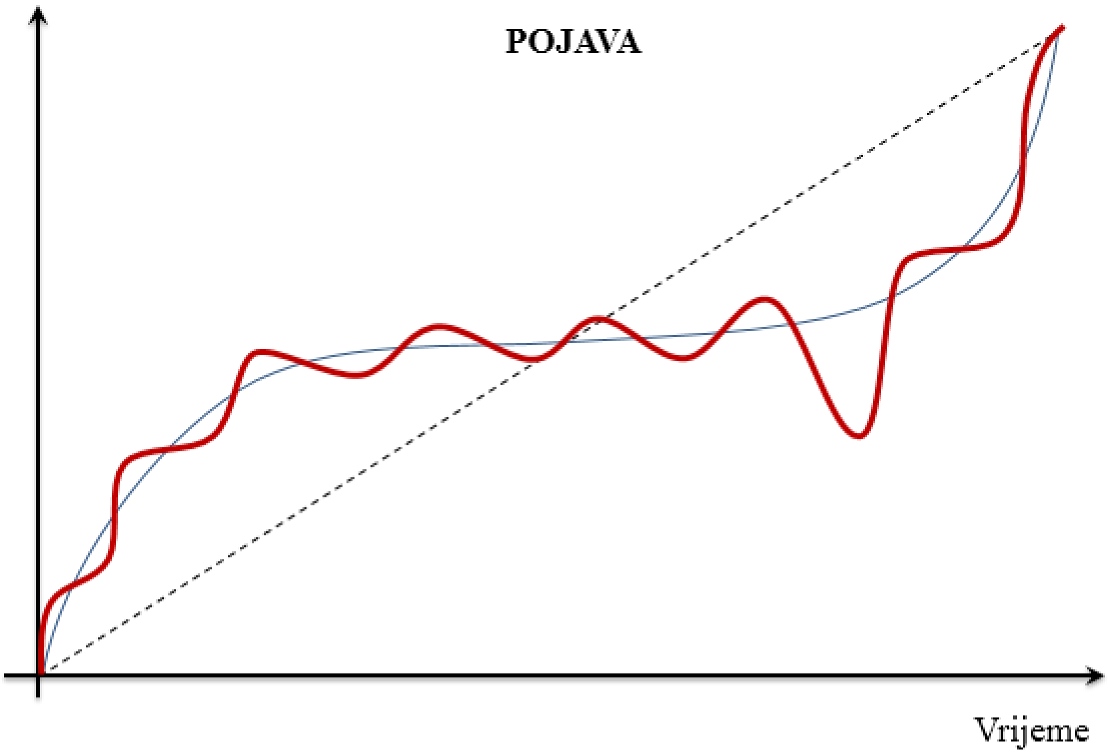
Ciklička i sezonska komponenta dobro su uočljive i na prikazu kumulativne prodaje na slici 32. Slučajna komponenta (crveni oval) vidljiva je ali je slabije izražena zbog relativno male vrijednosti te komponente u odnosu na kumulativnu vrijednost. Ako u nekoj situaciji imamo zadani kumulativni prikaz (sliku 32), iz njega uvijek možemo dobiti dijagram frekvencija (sliku 31) na kome su pojedine komponente jasnije izražene.



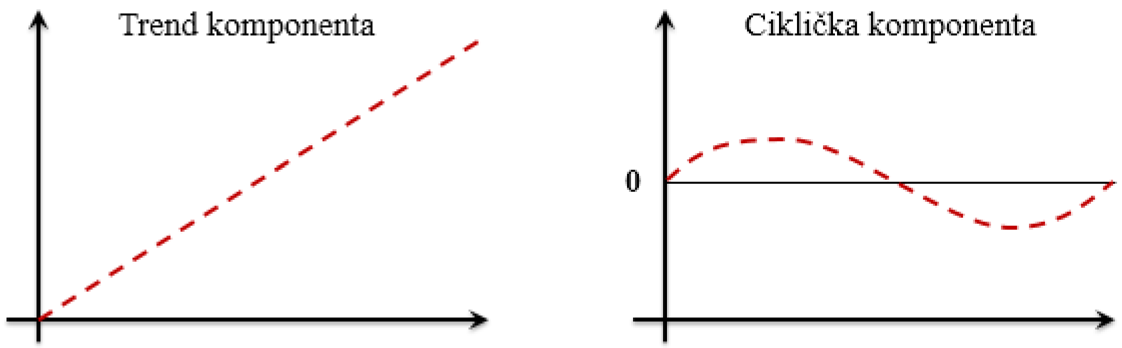
Slika 32. Kumulativni dijagram prodaje motornih vozila po kvartalima

Na slici 33 imamo ilustraciju pojave (crveni graf) sa sve četiri komponente. Na slikama 34 i 35 komponente su izdvojene, svaka zasebno. Zbroj svih komponenti na ta četiri grafikona daje graf pojave.

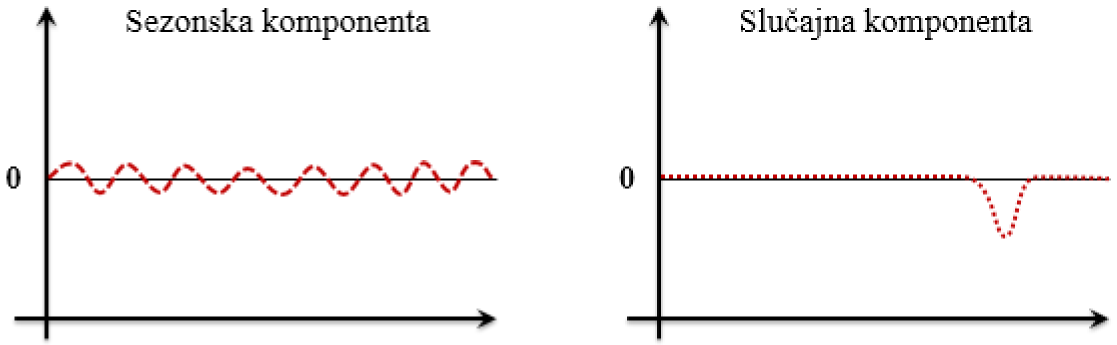
Uočavanje i izdvajanje pojedinih komponenti iz promatrane pojave u praksi je vrlo važno pa su razvijene mnoge metode za taj postupak. Veliku primjenu nalazimo u kriminalistici (izdvajanje slučajnih i nebitnih elemenata iz tragova i otisaka), telekomunikaciji (uklanjanje šumova i smetnji iz signala), astronomiji i geologiji (pročišćavanje valova – razdvajanje cikličke i stohastičke komponente kod elektromagnetskih, svjetlosnih, seizmičkih i drugih valova). Veliku primjenu nalazimo i u praćenju lokalnih i globalnih ekonomskih trendova (cikličko pojavljivanje perioda ekonomske krize, izmjena perioda rasta i pada cijena), zatim u praćenju vremenskih uvjeta (periodičko kretanje ciklona i anticiklona, morskih struja i sl.), u razumijevanju izmjene životnih ciklusa kod živih jedinki (biljke, životinje i ljudi) kao i kod čitavih populacija.



Slika 33. Pojava sa trend, cikličkom, sezonskom i slučajnom komponentom



Slika 34. Trend i ciklička komponenta pojave

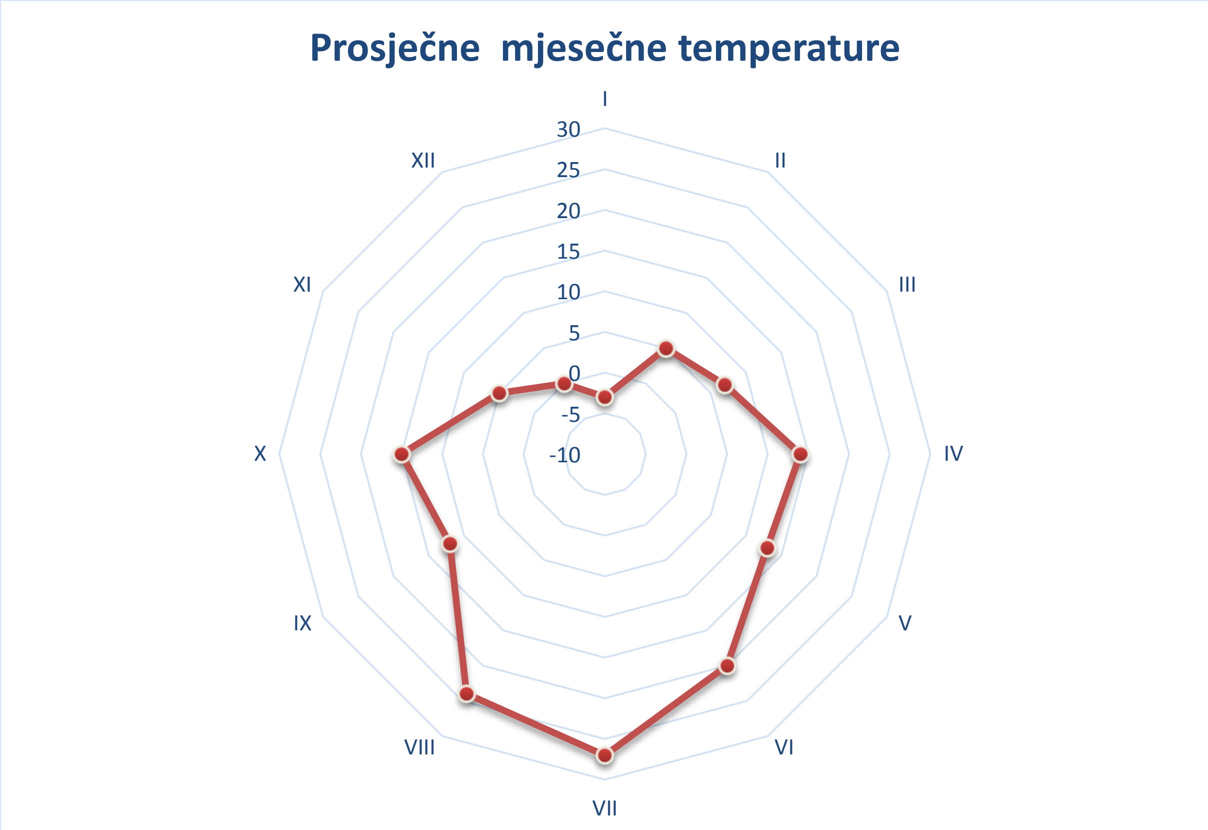


Slika 35. Sezonska i slučajna komponenta pojave

Za grafički prikaz cikličkih pojava sa sezonskim komponentama prikladan je i polarni dijagram (radar). Možemo ga koristiti i za ostale pojave a ilustriramo ga slijedećim primjerom.

Primjer 15. Na jednoj meteorološkoj mjernoj postaji tokom kalendarske godine zabilježene su slijedeće prosječne mjesečne temperature: –3, 5, 7, 14, 13, 20, 27, 24, 12, 15, 5, 0 °C. Prikažimo ove podatke polarnim dijagramom.

Koncentrične kružnice (ili poligonalne linije) predstavljaju iste razine vrijednosti varijable (poput izohipsi na zemljopisnim kartama). Zrake iz središta dijagrama predstavljaju vremenske točke (ukupno vrijeme promatranja pojave reprezentirano je punim krugom – kutom od 360°).



Slika 36. Prikaz prosječnih mjesečnih temperatura polarnim dijagramom

Na istom dijagramu možemo prikazati dva pa i više ciklusa (ovdje godina) radi usporedbe.

ZADACI

1. Odredite analitički oblik (formulu) trenda slijedećih vremenskih nizova:
   (a) 7,7,7,7,7,7,7. (b) 3,6,9,12,15,18,21. (c) 50,48,46,44,42,40. (d) 1,2,4,8,16,32,64.
   (e) 420,240,80,105,84,70. (f) 99,96,91,84,75,64,51. (g) 1/2,2/3,3/4,4/5,5/6,6/7,7/8.
2. Odredite trend i slučajnu komponentu niza 1,2,3,4,5,6,6,9,9,10,11,12.
3. Odredite trend i cikličku komponentu niza 52,98,152,198,252,298,352,398,452,498.
4. Razdvojite trend, cikličku i slučajnu komponentu niza 11,21,30,41,51,60,71,81,90, 102,111,121,131,141,150.
5. Napravite grafičke prikaze za neke od prethodnih zadataka. Zadajte ili pronađite sami slične nizove, razdvojite komponente i napravite grafičke prikaze.
6. Grafički prikažite tok kontinuirane vremenske pojave $x(t)$ za vrijeme $t \geq 0$ zadane sa:
   (a) $x(t) = 3t + 1$, (b) $x(t) = t^2$, (c) $x(t) = 420/t$, (d) $x(t) = 2^t$, (e) $x(t) = \sin t$.
7. Razdvojite pojedine komponente kontinuiranog signala $x(t) = A\sin(\omega t + c) + Bt^{\alpha}$, $t \geq 0$, u ovisnosti o vrijednosti parametra α ako su A, B, ω, c zadane konstante. Napravite grafički prikaz signala i njegovih komponenti za slučajeve $\alpha = 0$, $\alpha = 1$ i $\alpha = -1$ ako je $A = B = \omega = 1,\ c = 0$.

8. Što možete reći o strukturi kontinuiranog signala $x(t) = At^{\alpha} \sin(\omega t + c)$ u ovisnosti o vrijednosti parametra α ako su A, ω, c zadane konstante? Napravite grafički prikaz signala za slučajeve $\alpha = 0$, $\alpha = 1$ i $\alpha = -1$ ako je $A = \omega = 1,\ c = 0$.
9. Strukturnim stupcima prikažite vaš tjedni raspored sati. Napomena: svaki stupac predstavlja dnevni raspored sati i strukturiran je po broju sati za pojedini predmet tog dana, pri čemu svaki predmet možete prikazati drugom bojom.
10. Linijskim dijagramom prikažite dnevnu potrošnju svog džeparca tokom mjeseca.
11. Dvostrukim stupcima prikažite vrijeme buđenja i vrijeme odlaska na spavanje po danima tokom perioda od barem dva tjedna. Napomena: svakom danu pripadaju dva stupca različitih boja čije visine predstavljaju dnevno vrijeme buđenja i odlaska na spavanje. Postoji li u vašem dijagramu izražena ciklička komponenta?
12. Trostrukim ili višestrukim stupcima prikažite vrijeme doručka, ručka i večere (te međuobroka ako ga imate) slično kao u prethodnom zadatku. Postoji li ovdje ciklička komponenta ili prevladava slučajna?
13. Ocijenite vrijeme tokom dana ocjenom od 0 do 5. Pri tome neka broj 5 predstavlja vedar i sunčan dan, broj 4 djelomično sunčan a djelomično oblačan dan bez padalina, broj 3 djelomično sunčan i djelomično oblačan sa povremenim padalinama, broj 2 potpuno oblačan bez padalina, broj 1 potpuno oblačan sa povremenim padalinama te broj 0 cjelodnevne padaline. Pratite vrijeme kroz kalendarski mjesec te ga prikažite linijskim dijagramom na opisani način. Koje komponente uočavate u prikazu?
14. Slikovnim dijagramom ili dijagramom sa točkama prikažite tjedni pregled broja učenika u vašem razredu izostalih sa nastave po danima. Ako je neki učenik izostao samo pojedine sate u jednom danu prikažite ga dijelom sličice ili drugačijom sličicom ili sličicom druge boje.
15. Polarnim dijagramom prikažite broj slobodnih dana (dana kada nemate nastave) po mjesecima u godini.

RJEŠENJA

1. U vremenskom nizu $x_1, x_2, ..., x_n$ indeksi $t = 1, 2, ..., n$ su vremenski intervali ili točke pa treba pronaći analitičku vezu vrijednosti x_t i vremena t (funkciju $x_t(t)$). Imamo: (a) 7, (b) $3t$, (c) $52 - 2t$, (d) 2^{t-1}, (e) $420/t$, (f) $100 - t^2$, (g) $t/(t+1)$.
2. Trend komponenta je 1,2,3,4,5,6,7,8 ,9,10,11,12 a slučajna 0,0,0,0,0,0,–1,1,0,0,0,0.
3. Trend komponenta je 50,100,150,200,250,300,350,400,450,500, ciklička komponenta je 2,–2,2,–2,2,–2,2,–2,2,–2.
4. Trend je 11,21,31,41,51,61,71,81,91,101,111,121,131,141,151, ciklička komponenta 0,0,–1,0,0,–1,0,0,–1,0,0,–1,0,0,–1 a slučajna 0,0,0,0,0,0,0,0,0,1,0,1,0,0,0.
5. Vlastiti izbor.
6. Grafovi elementarnih funkcija.
7. Za $\alpha < 0$ član $A\sin(\omega t + c)$ je glavna (trend) komponenta a Bt^{α} je smetnja ili šum (slučajna) komponenta koja se protokom vremena smanjuje (prigušuje). Primijetimo da je glavna komponenta periodična jer je takva priroda signala. Za $\alpha > 0$ član Bt^{α} postaje glavna komponenta dok $A\sin(\omega t + c)$ postaje smetnja (za velike t oscilacije

smetnji su relativno sve manje u odnosu na veličinu glavne komponente). Za $\alpha = 0$ glavnu komponentu čini čitav signal $A\sin(\omega t + c) + B$. U konkretnom slučaju grafički prikazujemo $x(t) = \sin t + t$ koji razdvajamo na $\sin t$ i t, zatim $x(t) = \sin t + 1/t$ koji razdvajamo na $\sin t$ i $1/t$ te na kraju $x(t) = \sin t + 1$ koji ne razdvajamo.

8. Za $\alpha > 0$ imamo sve veće pojačavanje a za $\alpha < 0$ sve veće prigušivanje signala kroz vrijeme. Za $\alpha = 0$ imamo periodični signal konstantne jačine (amplitude). Signal ne možemo razdvojiti na aditivne komponente. Za ilustraciju prikazujemo grafove $x(t) = \sin t$, $x(t) = t\sin t$ i $x(t) = (\sin t)/t$.
9. Koristite svoj aktualni raspored sati.
10. Vodite evidenciju dnevne potrošnje.
11. Vodite evidenciju buđenja i odlaska na spavanje.
12. Izražena ciklička komponenta u ovom i prethodnom zadatku ukazuje na zdrave životne navike. Praćenje i analiza vlastitih aktivnosti tokom duljeg perioda (kao u ovom i prethodna dva zadatka) razvija svijest o potrebi samokontrole i discipline te potiče usvajanje zdravih navika.
13. Ocjenjujte dnevno vrijeme tokom kalendarskog mjeseca.
14. Pratite izostanke po danima i satima. Možete definirati sličice ili boje za izostanak sa jednog sata, dva sata, tri sata itd. sve do cjelodnevnog izostanka.
15. Koristite stvarne podatke.

INDEKSI

Jedan od osnovnih ciljeva analize vremenskih nizova je utvrditi dinamiku kojom se pojava mijenja kroz vrijeme. U tu svrhu trebamo usporediti vrijednosti članova niza. Ako uzmemo neki proizvoljni član niza, možemo ga usporediti sa njemu susjednim ili sa jednim odabranim članom (baznom vrijednošću). Osnovne mjere za usporedbu su apsolutna i relativna udaljenost (vidjeti poglavlje *Odnos dvaju podataka*). Pomoću njih se definiraju *verižni ili lančani* i *bazni indeksi* a svi oni zajedno čine *pokazatelje dinamike* vremenskog niza (pojave).

DEFINICIJA 9. NEKA JE $x_1, x_2, \ldots, x_{n-1}, x_n$ VREMENSKI NIZ.

ZA $t = 2, 3, \ldots, n$ JE

$$\Delta_t = x_t - x_{t-1} \text{ (prva diferencija)}, \quad \delta_t = \frac{x_t - x_{t-1}}{x_{t-1}} \text{ (stopa promjene)}, \quad V_t = \frac{x_t}{x_{t-1}} \text{ (verižni indeks)}.$$

ZA $t = 1, 2, \ldots, n$ I ZA ZADANI $b \in \{1, 2, \ldots, n\}$ JE

$$\Delta_t^* = x_t - x_b \text{ (prva diferencija)}, \quad \delta_t^* = \frac{x_t - x_b}{x_b} \text{ (stopa promjene)}, \quad I_t = \frac{x_t}{x_b} \text{ (bazni indeks)}$$

U ODNOSU NA BAZNO RAZDOBLJE b.

Vidimo da prve tri veličine mjere odnos dvaju susjednih članova niza a preostale tri odnos proizvoljnog člana sa referentnim baznim članom niza. Dok su diferencije izražene u mjernim jedinicama pojave, stope promjene i indeksi su neimenovani brojevi (relativne vrijednosti) pa se često izražavaju u postocima. Napomenimo da općenito vrijednost x_b ne mora biti član niza već bilo koja referentna vrijednost u odnosu na koju želimo analizirati odstupanja. Iz definicije proizlaze slijedeće veze između pojedinih pokazatelja dinamike.

$$\delta_t = \frac{\Delta_t}{x_{t-1}} = V_t - 1, \quad V_t = 1 + \delta_t, \quad \delta_t^* = \frac{\Delta_t^*}{x_b} = I_t - 1, \quad I_t = 1 + \delta_t^*, \quad \Delta_t = \Delta_t^* - \Delta_{t-1}^*, \quad V_t = \frac{I_t}{I_{t-1}}.$$

Ako želimo odrediti bazne indekse pomoću verižnih tada za baznu veličinu možemo uzeti prvi član niza pa imamo $I_1 = 1 = 100\ (\%)$, $I_i = V_i \cdot I_{i-1}$, $i = 2, 3, \ldots, n$. Pomoću navedenih pokazatelja možemo pratiti razvoj jedne pojave u vremenu (vidjeti također i primjer 18) pa govorimo o *individualnim* pokazateljima dinamike.

Primjer 16. U jednom renoviranom hotelu sa 100 kreveta ostvareno je 2625 noćenja u srpnju i 2850 u kolovozu. Ako je očekivana prosječna popunjenost hotela u ljetnim mjesecima 90%, analizirajmo ostvareni promet pokazateljima dinamike.

Imamo $x_1 = 2625$, $x_2 = 2850$ ($n = 2$). Za baznu vrijednost uzimamo očekivanu popunjenost $x_b = 100 \cdot 31 \cdot 0.90 = 2790$ noćenja. Imamo jedan verižni i dva bazna indeksa (i njima pridružene diferencije i stope promjene),

$$\begin{aligned} &\Delta_2 = 225, \quad \delta_2 = 0.08571\ldots \approx 8.57\%, \quad V_2 = 1.08571\ldots && \text{(verižni)}, \\ &\Delta_1^* = -165, \quad \delta_1^* = -0.05913\ldots \approx -5.91\%, \quad I_1 = 0.94086\ldots && \text{(bazni)}, \\ &\Delta_2^* = 60, \quad \delta_2^* = 0.02150\ldots \approx 2.15\%, \quad I_2 = 1.02150\ldots && \text{(bazni)}. \end{aligned}$$

Vidimo 8.57% (225) više noćenja u kolovozu u odnosu na srpanj te 2.15% (60) više od očekivanog dok je u srpnju 5.91% (165) manje od očekivanog broja noćenja.

Za razliku od individualnih imamo i *skupne* pokazatelje dinamike kojima možemo istodobno pratiti razvoj dviju ili više pojava u vremenu. Za primjer navodimo važne i često korištene *skupne indekse cijena, količina* i *vrijednosti*.

Odaberimo n različitih roba (proizvoda, usluga) koje koristimo u svakodnevnom životu čije cijene i količine promatramo u dva vremenska razdoblja (prvom-baznom i drugom-tekućem). Neka su p_i (p_i') cijene u prvom (drugom) a q_i (q_i') količine u prvom (drugom) razdoblju, $i = 1, 2, \ldots, n$. Umnožak cijene i količine je vrijednost pojedine robe u nekom od promatranih razdoblja. Slijedeći skupni indeksi mjere relativnu promjenu cijena, količina i vrijednosti svih odabranih roba u drugom razdoblju u odnosu na prvo razdoblje.

$$I_p = \frac{\sum_{i=1}^{n} p_i' q_i}{\sum_{i=1}^{n} p_i q_i}, \qquad I_p' = \frac{\sum_{i=1}^{n} p_i' q_i'}{\sum_{i=1}^{n} p_i q_i'} \qquad \text{(indeksi cijena),}$$

$$I_q = \frac{\sum_{i=1}^{n} p_i q_i'}{\sum_{i=1}^{n} p_i q_i}, \qquad I_q' = \frac{\sum_{i=1}^{n} p_i' q_i'}{\sum_{i=1}^{n} p_i' q_i} \qquad \text{(indeksi količina),}$$

$$I_v = \frac{\sum_{i=1}^{n} p_i' q_i'}{\sum_{i=1}^{n} p_i q_i} = I_p \cdot I_q' = I_q \cdot I_p' \qquad \text{(indeks vrijednosti).}$$

Indeksi I_p (I_p') pokazuju relativnu promjenu svih cijena promatranih roba na temelju količina iz prvog (drugog) razdoblja. Slično, indeksi I_q (I_q') pokazuju relativnu promjenu svih količina promatranih roba na temelju cijena iz prvog (drugog) razdoblja, dok indeks I_v pokazuje relativnu promjenu vrijednosti tih roba.

Kako je cijena jedna od osnovnih gospodarskih kategorija, ovi indeksi se koriste za praćenje životnog standarda i općenito gospodarske situacije na nacionalnoj razini. Primjenjuju se za određivanje *indeksa troškova života* ili *indeksa potrošačkih cijena* (CPI – Consumer Price Indeks) na temelju skupa reprezentativnih proizvoda (potrošačke ili obiteljske košarice).

$$\mathrm{CPI} = I_p = \frac{\sum_{i=1}^{n} p_i' q_i}{\sum_{i=1}^{n} p_i q_i} = \frac{\sum_{i=1}^{n} \frac{p_i'}{p_i} p_i q_i}{\sum_{j=1}^{n} p_j q_j} = \sum_{i=1}^{n} \frac{p_i'}{p_i} \cdot w_i, \qquad w_i = \frac{p_i q_i}{\sum_{j=1}^{n} p_j q_j}.$$

Vidimo da je CPI u stvari ponderirana aritmetička sredina indeksa cijena pri čemu su ponderi udjeli vrijednosti pojedinih proizvoda u ukupnoj vrijednosti košarice.

U RH izbor proizvoda i usluga za košaricu (trenutno oko 740) određuje se iz *ankete o potrošnji kućanstva* koju Državni zavod za statistiku provodi na godišnjoj razini na uzorku od 6500 kućanstava. Cijene proizvoda i usluga iz košarice prate se mjesečno u 9 gradova: Zagreb, Slavonski Brod, Osijek, Sisak, Rijeka, Pula, Split, Dubrovnik i Varaždin.

CPI se redovito objavljuje i služi za praćenje standarda i skupne razine cijena na tržištu. Pomoću njega se određuje veličina realnih plaća (omjer nominalne plaće i CPI) te stopa inflacije. Još jedna značajna primjena ovakvih indeksa su *burzovni indeksi* kojima se prati kretanje cijena dionica na burzama (CROBEX za Zagrebačku burzu).

Primjer 17. U potrošačku košaricu uvršteno je pet vrsta proizvoda: kruh (A), mlijeko i mliječni proizvodi (B), meso i mesne prerađevine (C), povrće (D) i voće (E). Podaci o cijenama i količinama za bazni i tekući mjesec dani su u tabeli 13. Količine su dane u komadima (1 kom = 750 grama) za proizvod A, u litrima za B te u kilogramima za C, D i E a cijene u kunama po jedinici količine. Odredimo indekse cijena, količina i vrijednosti te CPI za ovaj izbor.

	Bazni mjesec		Tekući mjesec	
Proizvod	cijena	količina	cijena	količina
A	7.00	30	7.00	31
B	5.90	20	6.00	22
C	23.50	12	24.50	14
D	5.60	10	5.90	12
E	10.90	8	10.50	7

Tabela 13. Podaci o sastavu potrošačke košarice

Određujemo sve indekse redom. Indeks cijena na temelju količina iz baznog razdoblja,

$$I_p = \mathrm{CPI} = \frac{7\cdot 30+6\cdot 20+24.5\cdot 12+5.9\cdot 10+10.5\cdot 8}{7\cdot 30+5.9\cdot 20+23.5\cdot 12+5.6\cdot 10+10.9\cdot 8} = \frac{767}{753.2} = 1.0183218\ldots .$$

Indeks cijena na temelju količina iz tekućeg razdoblja,

$$I'_p = \frac{7\cdot 31+6\cdot 22+24.5\cdot 14+5.9\cdot 12+10.5\cdot 7}{7\cdot 31+5.9\cdot 22+23.5\cdot 14+5.6\cdot 12+10.9\cdot 7} = \frac{836.3}{819.3} = 1.0207494\ldots .$$

Indeks količina na temelju cijena iz baznog razdoblja,

$$I_q = \frac{7\cdot 31+5.9\cdot 22+23.5\cdot 14+5.6\cdot 12+10.9\cdot 7}{7\cdot 30+5.9\cdot 20+23.5\cdot 12+5.6\cdot 10+10.9\cdot 8} = \frac{819.3}{753.2} = 1.08775889\ldots .$$

Indeks količina na temelju cijena iz tekućeg razdoblja,

$$I'_q = \frac{7\cdot 31+6\cdot 22+24.5\cdot 14+5.9\cdot 12+10.5\cdot 7}{7\cdot 30+6\cdot 20+24.5\cdot 12+5.9\cdot 10+10.5\cdot 8} = \frac{836.3}{767} = 1.09035202\ldots .$$

Indeks vrijednosti,

$$I_v = \frac{7\cdot 31+6\cdot 22+24.5\cdot 14+5.9\cdot 12+10.5\cdot 7}{7\cdot 30+5.9\cdot 20+23.5\cdot 12+5.6\cdot 10+10.9\cdot 8} = \frac{836.3}{753.2} = 1.11032926\ldots .$$

Vidimo da svi dobiveni indeksi upućuju na porast troškova života (svi su veći od 1) u tekućem mjesecu u odnosu na bazni mjesec. Indeksi cijena (količina) pokazuju skupno povećanje cijena (količina) dok indeks vrijednosti pokazuje da su ukupna izdvajanja za košaricu (njena vrijednost) veća za oko 11%.

ZADACI

1. Odredite verižne indekse vremenskog niza 40, 45, 51, 48, 60, 54, 52 te zatim bazne indekse u odnosu na medijalno razdoblje po veličini (ne po kronološkom redoslijedu).
2. Proizvodnja pšenice na jednom poljoprivrednom kompleksu tokom zadnjih pet godina dana je verižnim indeksima: – , 95, 110, 102, 90. Ako je u zadnjoj godini proizvodnja bila 95931 tonu, odredite proizvodnju u prve četiri godine.
3. Odredite verižne indekse za mjesece u kalendarskoj godini po broju dana.
4. U prvih šest mjeseci tekuće godine kretanje potražnje jedne vrste robe na određenom tržištu opisano je baznim indeksima: 90, 85, 100, 108, 104, 110. Opišite dinamiku potražnje stopama promjene te interpretirajte dobivene rezultate.
5. Rekonstruirajte stanje računa na početku svake od proteklih pet godina ako je na početku ove godine stanje bilo 12347.35 EUR a godišnje stope promjene (kamatne stope) su bile redom (u %): 4.1, 3.45, 3.8, 4.23 i 4.05.
6. Poznati su bazni indeksi: 120, 100, 110, 95, 105. Odredite bazne stope promjene, verižne indekse i verižne stope promjene.
7. Poznati su bazni indeksi broja učenika od prvih do osmih razreda jedne osnovne škole: 95, 90, 100, 107.5, 112.5, 92.5, 110, 85. Ako u školi ima ukupno 634 učenika, koliko ih pohađa prvi, drugi, treći, …, osmi razred?
8. Poznati su verižni indeksi broja učenika od prvih do četvrtih razreda jedne srednje škole: –, 90, 120, 87.5. Ako u školi ima ukupno 785 učenika, koliko ih pohađa prvi, drugi, treći i četvrti razred?
9. (a) Odredite verižne indekse za niz: 5, 6, 0, 0, 0, 4, 3.
   (b) Napišite općeniti niz koji ima verižne indekse: – , 150, 0, ∞, 200, 0, 100.
10. Od nastavnika zatražite podatke o broju upisanih učenika u prve razrede u vašoj školi tokom zadnjih 10 godina te ih prikažite: (a) dijagramom prvih diferencija, (b) dijagramom stopa promjena u odnosu na aritmetičku sredinu kao baznu veličinu.
11. U jednom proizvodnom pogonu je u baznom razdoblju utrošeno 420 kg sirovine A po cijeni 35 kn/kg, 300 l sirovine B po cijeni 75 kn/l i 520 komada sirovine C po cijeni 60 kn/kom. U tekućem razdoblju potrošeno je 450 kg A po 33 kn/kg, 280 l B po 78 kn/l i 550 kom C po 59 kn/kom. Odredite indekse cijena i količina u odnosu na bazno razdoblje te indeks vrijednosti.
12. Odaberite nekoliko proizvoda i odredite njihove količine za potrošačku košaricu. Prikupite podatke o njihovim cijenama u različitim trgovačkim centrima (Konzum, Billa, Plodine, KTC i sl.) te odredite skupne indekse cijena jednog centra u odnosu na drugi (u parovima).

RJEŠENJA

1. Verižni indeksi (u postocima, zaokruženi na dvije decimale): 112.50, 113.33, 94.12, 125.00, 90.00, 96.30. Medijalno razdoblje po veličini je treće pa je $x_b = 51$ (jer su u tri razdoblja vrijednosti manje od 51 a u tri veće). Bazni indeksi: 78.43, 88.24, 100.00, 94.12, 117.65, 105.88, 101.96.

2. Označimo proizvodnje sa $P_1, P_2, P_3, P_4.P_5$. Imamo $P_5 = 95931 = 0.9P_4 \Rightarrow P_4 = 106590$, $P_4 = 1.02P_3 \Rightarrow P_3 = 104500$, $P_3 = 1.1P_2 \Rightarrow P_2 = 95000$, $P_2 = 0.95P_1 \Rightarrow P_1 = 100000$ tona.
3. Zaokruženi na dvije decimale indeksi su: – , 90.32 (93.55), 110.74 (106.90), 96.77, 103.33, 96.77, 103.33, 100, 96.77, 103.33, 96.77, 103.33. U zagradama su indeksi za veljaču i ožujak u prijestupnoj godini.
4. Imamo $\delta_t = V_t - 1 = I_t / I_{t-1} - 1,\ i = 2,3,4,5,6$. Dakle, $\delta_2 = 85/90 - 1 = -0.04444...$, $\delta_3 = 100/85 - 1 = 0.17647...$, $\delta_4 = 108/100 - 1 = 0.08$, $\delta_5 = 104/108 - 1 = -0.03703...$, $\delta_6 = 110/104 - 1 = 0.05769...$. Interpretacija: u veljači je potražnja bila za oko 4.4% manja nego u siječnju, u ožujku za oko 17.6% veća nego u veljači itd.
5. Iz $\delta_i = (x_i - x_{i-1})/x_{i-1}$ slijedi $x_i = (1+\delta_i)x_{i-1}$, odnosno $x_{i-1} = x_i/(1+\delta_i)$, pa su stanja bila: 10184.98, 10602.57, 10968.36, 11385.15 i 11866.75 EUR.
6. Oduzmemo li 100 od baznih indeksa dobivamo bazne stope promjene (u %): 20, 0, 10, –5, 5. Kvocijent uzastopnih baznih indeksa daje verižne: –, 100/120=0.8333…= 83.33…%, 110/100=1.1=110%, 95/110=0.863636…=86.36…%, 105/95=1.10526…= 110.526…%. Oduzimanjem 100 od verižnih indeksa dobivamo verižne stope promjene (u %): –, –16.666…, 10, –13.636…, 10.526… . Uočimo da prvi verižni indeks i stopa promjene nisu definirani.
7. Zbroj baznih indeksa je 792.5% = 634 učenika, pa je 1% = 0.8 učenika. Množeći bazne indekse sa 0.8 dobivamo broj učenika: 76, 72, 80, 86, 90, 74, 88, 68.
8. Izračunamo bazne indekse. Vidjeli smo da iz $V_t = I_t / I_{t-1}$ slijedi $I_t = V_t \cdot I_{t-1}$, pa imamo $I_1 = 100\%$, $I_2 = 0.9 \cdot 100 = 90\%$, $I_3 = 1.2 \cdot 90 = 108\%$, $I_4 = 0.875 \cdot 108 = 94.5\%$. Dalje postupamo kao u prethodnom zadatku. Zbroj baznih indeksa je 392.5% = 785 učenika, pa je 1% = 2, odnosno, po razredima: 200, 180, 216, 189.
9. (a) $-$, 120, 0, 100, 100, ∞, 75. (b) $a,\ 1.5a,\ 0,\ b,\ 2b,\ 0,\ 0$, pri čemu je $a, b \neq 0$.
10. Koristite stvarne podatke za vašu školu.
11. Imamo $I_p = (33\cdot 420 + 78\cdot 300 + 59\cdot 520) : (35\cdot 420 + 75\cdot 300 + 60\cdot 520) = 0.99327...$, zatim $I_q = (35\cdot 450 + 75\cdot 280 + 60\cdot 550) : (35\cdot 420 + 75\cdot 300 + 60\cdot 520) = 1.0197368...$, te $I_v = (33\cdot 450 + 78\cdot 280 + 59\cdot 550) : (35\cdot 420 + 75\cdot 300 + 60\cdot 520) = 1.0108187...$.
12. Koristite stvarne podatke.

SREDNJE VRIJEDNOSTI

Do sada smo razvoj pojave u vremenu pratili pomoću uređenog vremenskog niza, njegovog grafičkog prikaza i pokazatelja dinamike. Ako želimo vremenski niz okarakterizirati samo jednim brojčanim pokazateljem koristimo srednje vrijednosti koje smo već ranije upoznali.

Ako imamo uređeni intervalni niz, to je samo specijalni slučaj niza kvantitativnih podataka, pa se srednje vrijednosti i mjere odstupanja određuju na ranije opisani način (vidjeti poglavlja *Srednje vrijednosti* i *Mjere raspršenosti*). Pri tome treba paziti da su intervali međusobno jednaki. Ako nisu podaci se korigiraju na fiksnu duljinu intervala. Srednje vrijednosti su prikladne za opisivanje nizova sa pretežno nesistematskim komponentama. Za sistematske komponente prikladniji su modeli vremenskih trendova a srednje vrijednosti daju uvid u prosječne vrijednosti pojave.

Pogledajmo sada pokazatelje dinamike. Kako je prva diferencija aditivna veličina (dobivena je operacijama zbrajanja i oduzimanja, $\Delta_t = x_t - x_{t-1}$, $x_t = x_{t-1} + \Delta_t$) za određivanja prosjeka koristimo aritmetičku sredinu. Imamo *prosječnu prvu diferenciju* susjednih članova (Δ), u odnosu na raspon varijacije (Δ_R) i prosječnu apsolutnu prvu diferenciju (Δ_{abs}),

$$\Delta = \frac{\Delta_2 + \Delta_3 + \ldots + \Delta_n}{n-1} = \frac{x_n - x_1}{n-1}, \qquad \Delta_R = \frac{\max\{x_1, x_2, \ldots, x_n\} - \min\{x_1, x_2, \ldots, x_n\}}{n-1},$$
$$\Delta_{\text{abs}} = \frac{|\Delta_2| + |\Delta_3| + \ldots + |\Delta_n|}{n-1}.$$

Slično tome imamo prosječnu prvu (Δ^*) i prosječnu apsolutnu prvu (Δ^*_{abs}) diferenciju u odnosu na bazno razdoblje,

$$\Delta^* = \frac{\Delta_1^* + \Delta_2^* + \ldots + \Delta_n^*}{n} = \frac{\sum_{t=1}^{n}(x_t - x_b)}{n} = \frac{nA - nx_b}{n} = A - x_b, \qquad \Delta^*_{\text{abs}} = \frac{|\Delta_1^*| + |\Delta_2^*| + \ldots + |\Delta_n^*|}{n},$$

gdje je A aritmetička sredina niza. Primijetimo da je u prve tri relacije nazivnik $n-1$, jer za n članova imamo $n-1$ susjednu diferenciju, dok je u preostalima nazivnik n, jer za n članova imamo isto toliko diferencija sa baznom vrijednošću.

Napomenimo da se ponekad, za potrebe analize računaju momenti (vidjeti definiciju 6) pa se mjere raspršenosti (varijanca, standardna devijacija i koeficijent varijacije) uzimaju u odnosu na zadanu baznu vrijednost x_b umjesto u odnosu na aritmetičku sredinu A (u definiciji 5 zamijenimo A sa x_b). Time dobivamo pokazatelje raspršenosti pojave oko te bazne vrijednosti.

Ako pogledamo indekse, oni su multiplikativne veličine (dobiveni su operacijama množenja i dijeljenja, $V_t = x_t / x_{t-1}$, $x_t = V_t \cdot x_{t-1}$) pa za njihov prosjek koristimo geometrijsku sredinu. Tako imamo *prosječni verižni indeks* (V) i *prosječni bazni indeks* (I) te pripadne *prosječne stope promjene* (δ i δ^*),

$$V = \sqrt[n-1]{V_2 V_3 \cdot \ldots \cdot V_n} = \sqrt[n-1]{\frac{x_n}{x_1}}, \quad \delta = V - 1, \qquad I = \sqrt[n]{I_1 I_2 \cdot \ldots \cdot I_n} = \sqrt[n]{\frac{x_1 x_2 \cdot \ldots \cdot x_n}{x_b^n}} = \frac{G}{x_b}, \quad \delta^* = I - 1,$$

pri čemu je G geometrijska sredina niza. U analizi pojave se također mogu koristiti minimalna i maksimalna vrijednost verižnih i baznih indeksa ($V_{\min}, V_{\max}$ i $I_{\min}, I_{\max}$).

Primjer 18. Jedan trgovački centar je tokom tjedna ostvario slijedeći promet: ponedjeljak, utorak i srijedu zajedno 168000 kuna, četvrtak i petak zajedno 126000 kuna, subotu 84000 kuna i nedjelju 42000 kuna. Analizirajmo količinu prometa pomoću pojedinih pokazatelja dinamike.

Kako vremenski intervali nisu iste duljine podatke ćemo korigirati na fiksnu duljinu od jednog dana. Razredni omjeri su 3,2,1,1 pa su korigirani podaci: 168000:3=56000, 126000:2=63000, 84000 i 42000. Dakle, uređeni niz je $(56000,3),\ (63000,2),\ (84000,1),\ (42000,1)$. Aritmetička sredina niza je

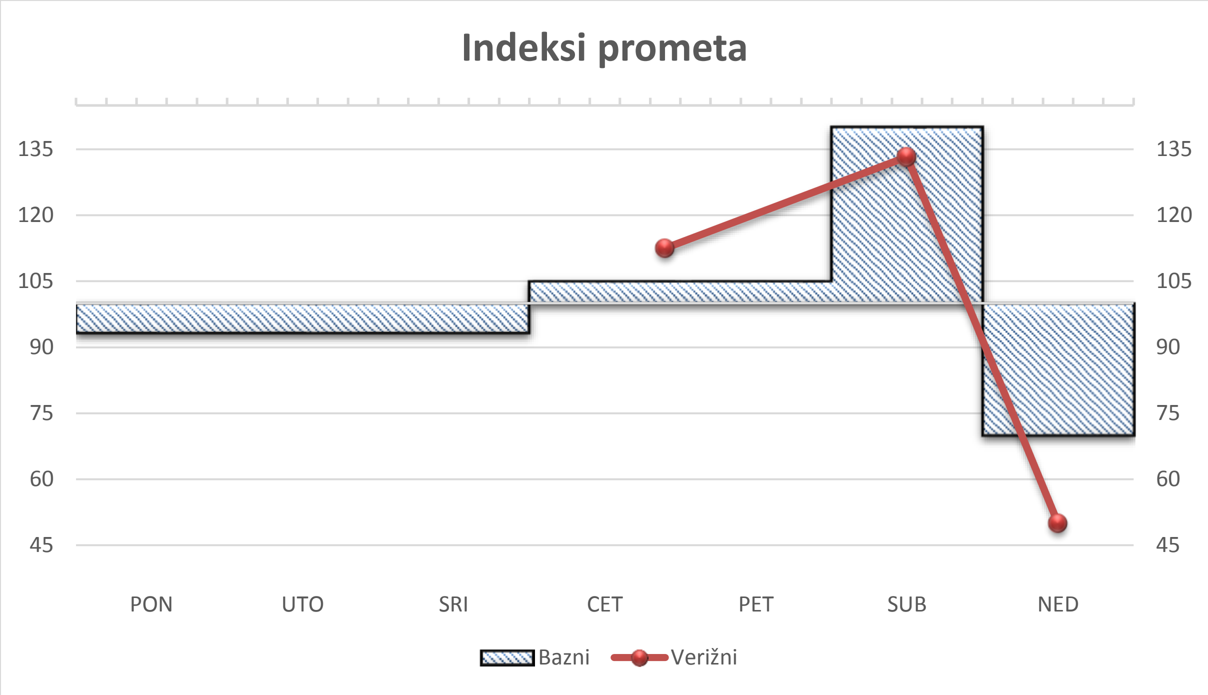
$$A = \frac{3\cdot 56000 + 2\cdot 63000 + 1\cdot 84000 + 1\cdot 42000}{3+2+1+1} = 60000,$$

što je prosječni dnevni promet trgovačkog centra. Pojedini pokazatelji dinamike prometa dani su u tabeli 14 (zaokruženi na tri decimale) i na slici 37. Za baznu vrijednost uzeli smo aritmetičku sredinu, $x_b = A = 60000$.

t	Δ_t	V_t	δ_t	Δ_t^*	I_t	δ_t^*
1	–	–	–	–4000	0.933	–0.067
2	7000	1.125	0.125	3000	1.050	0.050
3	21000	1.333	0.333	24000	1.400	0.400
4	–42000	0.500	–0.500	–18000	0.700	–0.300

Tabela 14. Pokazatelji dinamike prometa trgovačkog centra

Podaci iz tabele 14 pokazuju dinamiku promjene dnevnog prometa iz razdoblja u razdoblje. Tako na primjer 21000 pokazuje da je dnevni promet u trećem razdoblju bio za 21000 kuna veći od dnevnog prometa u drugom razdoblju, 1.125 pokazuje da je dnevni promet u drugom razdoblju bio 1.125 puta veći (iznosio je 112.5%) od dnevnog prometa u prvom razdoblju itd. Interpretirajte i ostale podatke iz tabele 14.



Slika 37. Linijski dijagram verižnih indeksa i površinski dijagram baznih indeksa prometa trgovačkog centra

Za prosječne pokazatelje dinamike imamo

$$\Delta = \frac{42000-56000}{4-1} = -4666.67, \quad \Delta_R = \frac{84000-42000}{4-1} = 14000, \quad \Delta_{\text{abs}} = \frac{70000}{4-1} = 23333.33.$$

Tako vidimo da su postojale velike prosječne oscilacije dnevnog prometa tokom tjedna u oba smjera (povećanje i smanjenje) jer se Δ_R i Δ_{abs} znatno razlikuju, a prosječna tendencija tokom tjedna bila je smanjenje prometa (Δ je negativan). To je vidljivo i iz prosječnih vrijednosti indeksa,

$$V = \sqrt[3]{\frac{42000}{56000}} = 0.90856\ldots < 1, \qquad I = \frac{\sqrt[4]{56 \cdot 63 \cdot 84 \cdot 42}}{60} = 0.989949\ldots < 1.$$

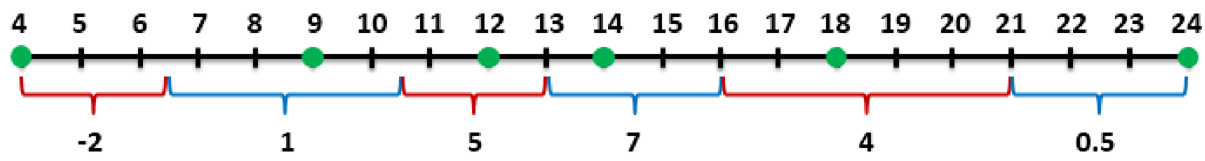
Kod trenutačnih nizova frekvencije nema smisla zbrajati jer pokazuju stanje pojave u jednom trenutku. Za takve nizove računamo ponderirane srednje vrijednosti. Ako ponderi nisu definirani samom prirodom promatrane pojave, određujemo ih uz pretpostavku da *razina pojave u nekoj vremenskoj točki vrijedi za pola razdoblja ispred i pola razdoblja iza te točke* (za prvu vrijednost samo pola iza a za zadnju samo pola ispred nje). Duljina razdoblja za koje razina vrijedi je ponder za tu razinu (ili je njoj proporcionalna). Postupak ilustriramo na slijedećem primjeru.

Primjer 19. Tokom jednog dana imamo slijedeća očitanja temperature na mjernoj postaji: u 4 sata ujutro –2°C, u 9 sati 1°C, u 12 sati 5°C, u 14 sati 7°C, u 18 sati 4°C i u 24 sata 0.5°C. Odredite prosječnu temperaturu tog dana.

Promatranje pojave (mjerenje temperature) počinje u 4 a završava u 24 sata. Prvo mjerenje je u 4 sata a drugo u 9. Razmak je 5 sati. Za prvu polovicu tog razmaka (2.5 sati) vrijedi temperatura –2°C a za drugu (2.5 sati) 1°C. Slijedeći razmak od 9 do 12 sati je 3 sata. Za prvu polovicu tog razmaka (1.5 sati) vrijedi temperatura 1°C a za drugu (1.5 sati) 5°C. Nastavljamo na isti način. Od 12 do 14 sati je razmak od 2 sata. Za prvu polovicu tog razmaka (1 sat) vrijedi temperatura 5°C a za drugu (1 sat) 7°C. Od 14 do 18 sati razmak je 4 sata. Za prvu polovicu tog razmaka (2 sata) vrijedi temperatura 7°C a za drugu (2 sata) 4°C. Na kraju od 18 do 24 sata je razmak 6 sati. Za prvu polovicu tog razmaka (3 sata) vrijedi temperatura 4°C a za drugu (3 sata) 0.5°C i to je kraj mjerenja. Time smo vremenski tijek pojave (od 4 do 24 sata = 20 sati) podijelili na intervale pridružene svakoj pojedinoj temperaturi. Duljina tih intervala su ponderi za određivanje srednje vrijednosti, pa imamo

$$A = \frac{-2\cdot 2.5 + 1\cdot(2.5+1.5) + 5\cdot(1.5+1) + 7\cdot(1+2) + 4\cdot(2+3) + 0.5\cdot 3}{20} = 2.7^{\circ}\mathrm{C}.$$

Za bolji uvid u ovu razdiobu možemo koristiti prikaz na vremenskoj osi (slika 38).



Slika 38. Vremenski ponderi za temperaturna očitanja

ZADACI

1. Potrošnja vode iz javnog vodovoda u mjestu X tokom jednog tjedna bila je:

Dan	*PON*	*UTO*	*SRI*	*ČET*	*PET*	*SUB*	*NED*
Potrošnja (m^3)	920	881	895	925	1050	1160	805

   Odredite svih pet prosječnih prvih diferencija. Za baznu veličinu uzmite prosječnu dnevnu potrošnju vode.
2. Odredite varijancu, standardnu devijaciju i koeficijent varijacije u prethodnom zadatku: (a) u odnosu na prosječnu potrošnju, (b) u odnosu na potrošnju na početku tjedna (ponedjeljak).
3. Odredite svih pet prosječnih prvih diferencija za niz x_1, x_2. Za baznu veličinu uzmite aritmetičku sredinu niza.
4. Kakva je veza između Δ, Δ_R, Δ_{abs} ako je niz $x_1, x_2, ..., x_n$: (a) monotono rastući, (b) monotono padajući ?

5. U vremenskom nizu x_1, x_2, x_3, x_4, x_5 poznate su prve diferencije $\Delta_2 = \Delta_5 = 2$, $\Delta_3 = 1$, $\Delta_4 = -4$. Ako je $x_A = 31$ napišite niz.
6. Odredite niz x_1, x_2, x_3 za koji je $x_1 = 50$, $\Delta = 3$ i $\Delta_{\text{abs}} = 5$.
7. Ako putujete u školu javnim prijevozom, pratite točnost dolaska prijevoznog sredstva (barem 10 opažanja) u odnosu na vrijeme po voznom redu (bazna veličina). Odredite prosječne diferencije Δ i Δ_{abs}. U koliko % slučajeva se radilo o točnom dolasku, ranijem dolasku i zakašnjenju?
8. Pratite svoj omiljeni sportski klub (ekipu) tokom natjecateljske sezone te u svakom kolu bilježite neki od slijedećih podataka: (a) mjesto na prvenstvenoj ljestvici, (b) broj postignutih golova (pogodaka), (c) gol razliku, (d) broj osvojenih bodova. Prikažite tako prikupljene podatke linijskim dijagramom, odredite prve diferencije i svih pet prosječnih diferencija. Na temelju dobivenih pokazatelja interpretirajte dinamiku uspješnosti kluba iz kola u kolo te prosječnu uspješnost tokom sezone.
9. Broj prevezenih putnika od strane jednog autobusnog prijevoznika po kvartalima prošle kalendarske godine bio je: 8453 (I), 9215 (II), 10320 (III) i 8855 (IV). Odredite verižne indekse, bazne indekse u odnosu na IV kvartal, te za njih najveće, najmanje i prosječne indekse kao i prosječne stope promjene.
10. Na izlaznom vodu trafostanice za opskrbu jedne gradske četvrti u razdoblju od 15 do 22 sata izmjerena je jakost isporučene električne energije (u amperima).

Vrijeme	***15.20***	***16.10***	***17.30***	***18.06***	***18.40***	***19.15***	***19.55***	***20.20***	***20.58***	***21.30***
Jakost (A)	45	40	60	68	76	94	108	96	100	80

Na temelju ovih mjerenja odredite prosječnu jakost isporučene električne energije u navedenom periodu te procijenite ukupnu potrošnju uz približno konstantni napon gradske mreže od 220 V.

11. Tokom obilnih kiša dnevni vodostaj jedne rijeke imao je slijedeće vrijednosti:

Vrijeme (h)	***3***	***8***	***11***	***17***	***19***
Vodostaj (cm)	120	145	180	235	250

Odredite prosječnu vrijednost vodostaja tokom ovih 16 sati te prosječan porast vodostaja po satu. Ako se prosječan porast nastavi istim tempom, kada možemo očekivati dostizanje kritične vrijednosti od 350 cm za uvođenje izvanrednih mjera obrane od poplave?

12. Neka je $x_1, x_2, \ldots, x_n$ trenutačni vremenski niz sa jednakim vremenskim razmacima između članova niza. Odredite aritmetičku sredinu ovog niza ako navedeni podaci pripadaju: (a) sredini svakog razdoblja u kome je pojava promatrana, (b) prvi početku prvog a zadnji kraju zadnjeg razdoblja u kome je pojava promatrana.

RJEŠENJA

1. Imamo: $\Delta = (805 - 920):(7-1) \approx -19.167$, $\Delta_R = (1160 - 805):(7-1) \approx 59.167$, $\Delta_{\text{abs}} = (39 + 14 + 30 + 125 + 110 + 355):6 \approx 112.167$. Prosječna potrošnja je $A = 948$, pa imamo $\Delta^* = 948 - 948 = 0$, $\Delta^*_{\text{abs}} = (28 + 67 + 53 + 23 + 102 + 212 + 143):7 \approx 89.714$.

2. (a) $\sigma^2 = (28^2 + 67^2 + 53^2 + 23^2 + 102^2 + 212^2 + 143^2) : 7 = 12058.2857...$, $\sigma = 109.81...$, $V = 0.11583... \approx 11.6\%$. (b) U definiciji 5 umjesto A uzimamo 920 kao baznu veličinu pa imamo $\sigma^2 = (0^2 + 39^2 + 25^2 + 5^2 + 130^2 + 240^2 + 115^2) : 7 = 12842.2857...$, $\sigma = 113.3238...$, $V = 0.123178... \approx 12.3\%$.
3. Imamo: $\Delta = x_2 - x_1$, $\Delta_R = |x_2 - x_1| = |\Delta| = \Delta_{\text{abs}}$. Kako je $A = (x_1 + x_2)/2 = x_b$, slijedi $\Delta^* = 0$, $\Delta^*_{\text{abs}} = \left(|x_1 - A| + |x_2 - A|\right)/2 = |x_2 - x_1|/2 = \Delta_{\text{abs}}/2$.
4. (a) $\Delta_{\text{abs}} = \Delta_R = \Delta$, (b) $\Delta_{\text{abs}} = \Delta_R = -\Delta$.
5. Na temelju poznatih prvih diferencija dobivamo $x_2 = x_1 + 2$, $x_3 = x_2 + 1 = x_1 + 3$, $x_4 = x_3 - 4 = x_1 - 1$, $x_5 = x_4 + 2 = x_1 + 1$. Kako je $x_1 + x_2 + x_3 + x_4 + x_5 = 31 \cdot 5$, imamo $5x_1 + 5 = 155 \Rightarrow x_1 = 30$, pa je niz 30, 32, 33, 29, 31.
6. Iz $\Delta = 3$ i $\Delta_{\text{abs}} = 5$ dobivamo sustav $\Delta_2 + \Delta_3 = 3 \cdot 2$, $|\Delta_2| + |\Delta_3| = 5 \cdot 2$. Iz prve jednadžbe je $\Delta_3 = 6 - \Delta_2$ što uvršteno u drugu daje $|\Delta_2| + |6 - \Delta_2| = 10$. Za $\Delta_2 < 0$ jednadžba glasi $-\Delta_2 + 6 - \Delta_2 = 10$ iz čega slijedi $\Delta_2 = -2$ pa je $\Delta_3 = 8$. Za $0 \le \Delta_2 \le 6$ je $\Delta_2 + 6 - \Delta_2 = 10$ što nema rješenja a za $\Delta_2 > 6$ je $\Delta_2 - 6 + \Delta_2 = 10$, odnosno $\Delta_2 = 8$ i $\Delta_3 = -2$. Dakle, imamo dva takva niza: 50, 48, 56 i 50, 58, 56.
7. Koristite stvarne podatke. Rezultati ukazuju na (ne)realnost voznog reda.
8. Koristite stvarne podatke. Primijetite da se rezultati analize temelje na sistematskom promatranju pojave tokom dužeg vremenskog perioda.
9. Verižni indeksi su $V_2 = 109.015$, $V_3 = 111.991$, $V_4 = 85.804$ a bazni $I_1 = 95.460$, $I_2 = 104.065$, $I_3 = 116.544$, $I_4 = 100$, pa su $V_{\min} = V_4$, $V_{\max} = V_2$, $I_{\min} = I_1$, $I_{\max} = I_3$, $V = 101.561$, $\delta = 1.561$, $I = 103.730$, $\delta^* = 3.730$. Svi rezultati navedeni su u postocima zaokruženim na tri decimale.
10. Polovica vremenskog razmaka između dva uzastopna mjerenja ulazi u ponder svakog od njih, za prvo mjerenje čitav razmak prije a za zadnje čitav razmak poslije njega. Za jasniju predodžbu napravite sliku (poput slike 38). Prosječnu jakost dobijemo da zbroj 45(20+25) + 40(25+40) + 60(40+18) + 68(18+17) + 76(17+17.5) + 94(17.5+20) + 108(20+12.5) + 96(12.5+19) + 100(19+16) + 80(16+30) podijelimo sa 420 (7 sati izraženo u minutama) što daje 72.25238095… A. Snaga isporučene energije ja umnožak prosječne jakosti i napona, 15895.52381…W = 15.89552… kW, a potrošnja (rad) je umnožak snage i vremena (7 sati), 111.26866… kWh.
11. Slično kao u prethodnom zadatku dobivamo

$$A = \frac{120 \cdot 2.5 + 145 \cdot (2.5 + 1.5) + 180 \cdot (1.5 + 3) + 235 \cdot (3 + 1) + 250 \cdot 1}{16} = 180 \text{ cm.}$$

Prosječan porast vodostaja dobivamo iz raspona varijacije, $(250 - 120) : 16 = 8.125$ cm po satu. Uz ovaj tempo porasta, za $(350 - 250) : 8.125 \approx 12.3$ sati možemo očekivati dostizanje navedene kritične razine (idućeg dana između 7 i 8 sati ujutro).

12. (a) $A=(x_1+x_2+...+x_n)/n$, (b) $A=(0.5x_1+x_2+x_3+...+x_{n-1}+0.5x_n)/(n-1)$, jer je ponder za x_1 samo prva polovica prvog razdoblja a za x_n samo druga polovica zadnjeg razdoblja dok je za ostale x_i ponder cijelo razdoblje (pola prije i pola poslije), pa je zbroj svih pondera $n-1$ (nazivnik).

VREMENSKI TRENDOVI

Osim analize vremenskih pojava koje smo do sada upoznali, često je potrebno predvidjeti kretanje pojave u budućnosti. Takvo predviđanje najčešće se može provesti približno, sa određenom vjerojatnošću (kretanje cijena na burzama, potražnje na tržištu, prognoza vremena itd.). U tu svrhu koristimo ranije spomenutu dekompoziciju vremenske varijable (x) na trend (T), cikličku (C), sezonsku (S) i slučajnu (s) komponentu u vremenskom trenutku t,

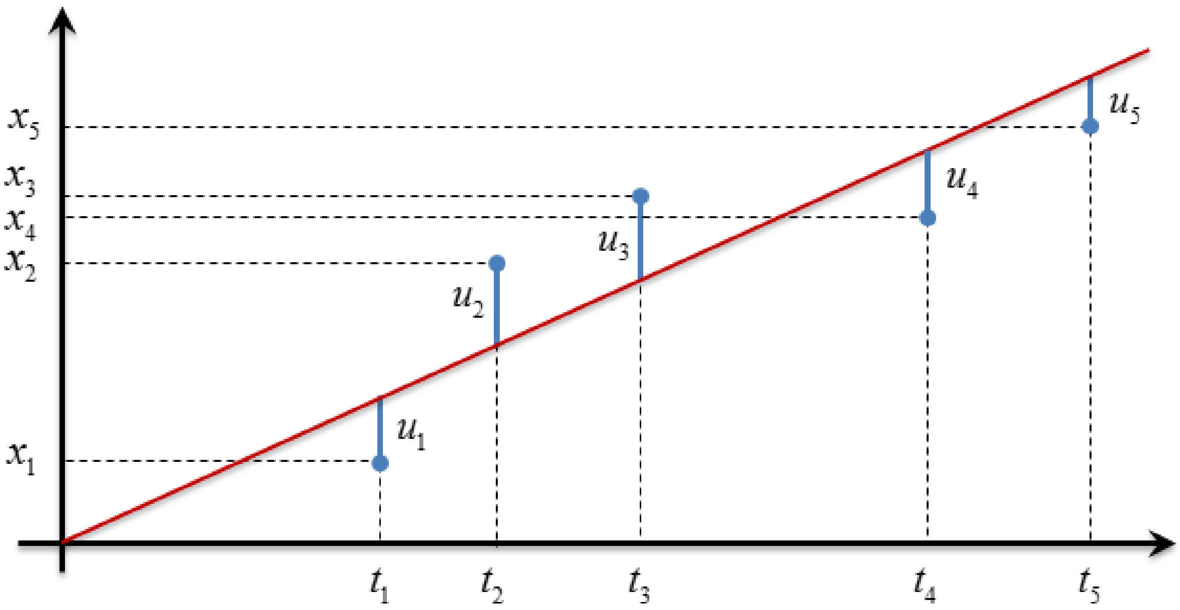
$$x(t)=T(t)+C(t)+S(t)+s(t).$$

Najčešće se koristi ovakav aditivni model (gdje se komponente zbrajaju) ali postoje i drugi modeli (multiplikativni, kombinirani i sl.). U navedenom izrazu $x(t)$ je diskretna ili neprekidna funkcija vremena t. U diskretnom slučaju vremenske intervale (ili točke) općenito označavamo sa t_i, $i=1,2,...$ (ili samo sa $t=1,2,...$ ako su numerirani prirodnim brojevima – prvo, drugo itd. razdoblje ili trenutak) a pripadne vrijednosti funkcija sa x_i, T_i, C_i, S_i, s_i kao i do sada. Da bi mogli predvidjeti vrijednost funkcije x u nekoj točki $t=t_*$, trebamo naći analitički oblik (formulu) za funkciju $x(t)$ ili barem za njenu glavnu komponentu – trend $T(t)$. Ovisno o tipu funkcije (linearna, kvadratna, eksponencijalna i sl.) govorimo o linearnom, kvadratnom, eksponencijalnom i sličnim vremenskim trendovima promatrane pojave. Tako je na primjer put kod jednolikog gibanja po pravcu opisan linearnom funkcijom $x(t)=vt+x_0$ (v je konstantna brzina a x_0 početni položaj), jednoliko ubrzano gibanje kvadratnom funkcijom $x(t)=0.5at^2+x_0$ (a je konstantna akceleracija) itd.

U nastavku ćemo pobliže upoznati *model linearnog trenda* za zadani vremenski niz $x_1,x_2,...,x_n$ čije vrijednosti pripadaju vremenskim intervalima ili trenucima $t_1,t_2,...,t_n$. Dakle, tražimo funkciju $x(t)$ za koju je $x(t_i)=x_i$, $i=1,2,...,n$, a pri tome je njezin trend linearan, $T(t)=at+b$. Za definiranje modela treba odrediti parametre a i b. Imamo

$$x(t)=at+b+u(t), \quad u(t)=C(t)+S(t)+s(t),$$

gdje je $u(t)$ *rezidualno odstupanje* pojave $x(t)$ od njenog linearnog trenda (slika 39).



Slika 39. Rezidualna odstupanja

Za zadani vremenski niz je

$$x_i = at_i + b + u_i \quad \Rightarrow \quad u_i = x_i - at_i - b, \quad i = 1, 2, ..., n.$$

Za određivanje parametara a i b trebamo dva uvjeta. Linearni trend će se prilagoditi podacima to bolje što su ukupna odstupanja za čitav niz manja. Zbog toga postavljamo uvjete da je zbroj svih odstupanja jednak nuli a zbroj kvadrata odstupanja minimalan,

$$\sum_{i=1}^{n} u_i = 0, \qquad \sum_{i=1}^{n} u_i^2 \to \min. \tag{17}$$

Kako za aritmetičke sredine x_A i t_A nizova $x_1, x_2, ..., x_n$ i $t_1, t_2, ..., t_n$ vrijedi $\sum_{i=1}^{n} x_i = nx_A$ i $\sum_{i=1}^{n} t_i = nt_A$, iz prvog uvjeta slijedi

$$\sum_{i=1}^{n} u_i = 0 \;\Rightarrow\; \sum_{i=1}^{n} x_i - a\sum_{i=1}^{n} t_i - \sum_{i=1}^{n} b = 0 \;\Rightarrow\; nx_A - ant_A - nb = 0 \;\Rightarrow\; b = x_A - at_A,$$

Dobiveni izraz za b uvrštavamo u drugi uvjet,

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (x_i - at_i - b)^2 = \sum_{i=1}^{n} \left[(x_i - x_A) - a(t_i - t_A) \right]^2$$
$$= a^2 \sum_{i=1}^{n} (t_i - t_A)^2 - 2a \sum_{i=1}^{n} (x_i - x_A)(t_i - t_A) + \sum_{i=1}^{n} (x_i - x_A)^2 \quad \to \quad \min.$$

Dobili smo kvadratnu funkciju oblika $\alpha a^2 + \beta a + \gamma$ (razlikujte slovo a od $\alpha =$ alfa) sa vodećim koeficijentom $\alpha > 0$, koja ima minimum u tjemenu parabole, $a = -\beta / (2\alpha)$, pa je

$$a = \frac{2\sum_{i=1}^{n} (x_i - x_A)(t_i - t_A)}{2\sum_{i=1}^{n} (t_i - t_A)^2} = \frac{\sum_{i=1}^{n} x_i t_i - t_A \sum_{i=1}^{n} x_i - x_A \sum_{i=1}^{n} t_i + \sum_{i=1}^{n} x_A t_A}{\sum_{i=1}^{n} t_i^2 - 2t_A \sum_{i=1}^{n} t_i + \sum_{i=1}^{n} t_A^2}$$
$$= \frac{\sum_{i=1}^{n} x_i t_i - t_A \cdot n x_A - x_A \cdot n t_A + n x_A t_A}{\sum_{i=1}^{n} t_i^2 - 2t_A \cdot n t_A + n t_A^2} = \frac{\sum_{i=1}^{n} x_i t_i - n x_A t_A}{\sum_{i=1}^{n} t_i^2 - n t_A^2}.$$

Navedeni postupak za određivanje parametara a i b naziva se *metoda najmanjih kvadrata*.

Dakle, za parametre optimalnog linearnog trenda $x = at + b$ koji najbolje aproksimira podatke

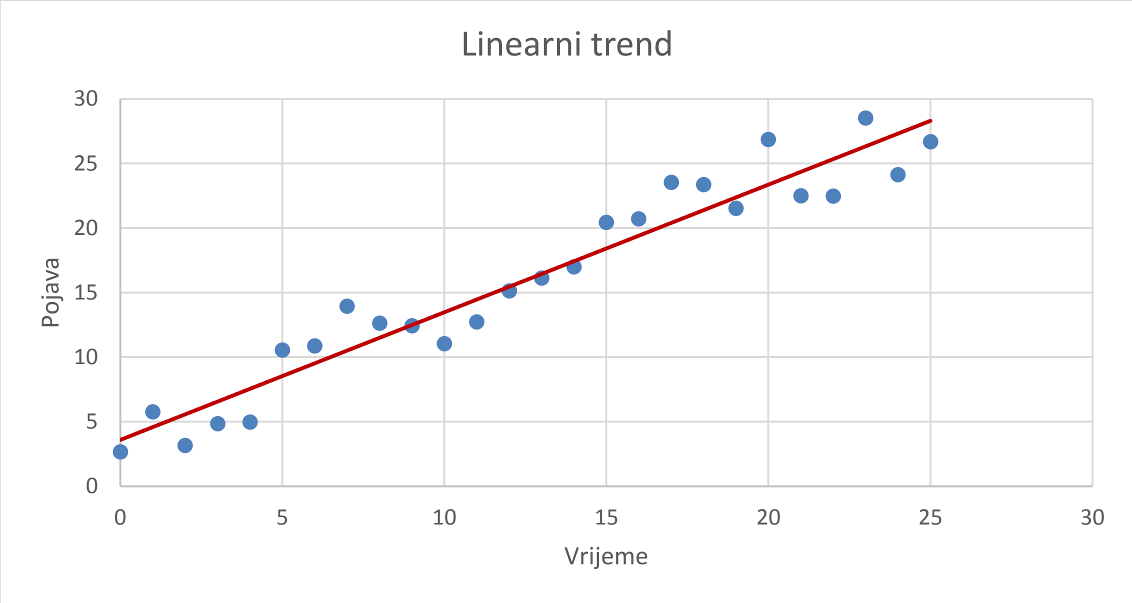
$$(t_1, x_1), \; (t_2, x_2), \; \ldots \; (t_{n-1}, x_{n-1}), \; (t_n, x_n),$$

vrijedi

$$a = \frac{\sum_{i=1}^{n} (x_i - x_A)(t_i - t_A)}{\sum_{i=1}^{n} (t_i - t_A)^2} = \frac{\sum_{i=1}^{n} x_i t_i - n x_A t_A}{\sum_{i=1}^{n} t_i^2 - n t_A^2}, \quad b = x_A - a t_A, \quad x_A = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad t_A = \frac{1}{n} \sum_{i=1}^{n} t_i. \qquad (18)$$

Tako dobiveni trend (pravac) općenito ne prolaziti točkama niza ali im se najbolje prilagođava u smislu da su odstupanja $u_i = x_i - (at_i + b)$, $i = 1, 2, \ldots, n$ najmanja moguća.

Ovakav pristup posebno je prikladan za velik broj podataka koji u osnovi imaju linearnu tendenciju promjene (slika 40). Koristeći dobiveni analitički oblik funkcije možemo odrediti (predvidjeti) kretanje pojave u vremenu koje nije obuhvaćeno zadanim vremenskim nizom.



Slika 40. Linearni trend vremenskog niza

Na analogni način se određuju parametri trenda bilo kojeg tipa (kvadratnog $x = at^2 + bt + c$, eksponencijalnog $x = a \cdot b^t + c$ ili sl.) samo što su izvodi složeniji.

Primjer 20. Nakon sustavnog uvođenja poticajnih mjera zapošljavanja na razini jedne županije, ukupan broj nezaposlenih osoba 31.12. je bio: 5725 (2011.), 5358 (2012.), 5020 (2013.), 4505 (2014.) i 4215 (2015.). Koja su očekivanja broja nezaposlenih u naredne dvije godine? Ako je 2015. godine stopa nezaposlenosti bila 15%, kada možemo očekivati njezin pad ispod 10%?

Broj nezaposlenih osoba je vremenski niz $x_i,\ i = 1,2,3,4,5$. Niz je trenutačni jer je svaki podatak očitan u jednoj vremenskoj točki, na zadnji dan godine, a odnosi se na tu godinu. Te točke možemo numerirati na proizvoljan način (budući da priroda problema ne nameće fiksnu numeraciju), npr. $t_1 = 11,\ t_2 = 12,\ t_3 = 13,\ t_4 = 14,\ t_5 = 15$. Koristimo izraze (18),

$$t_A = 13, \quad x_A = 4964.6, \quad a = \frac{318826 - 5 \cdot 4964.6 \cdot 13}{855 - 5 \cdot 13^2} = -387.3, \quad b = 4964.6 + 387.3 \cdot 13 = 9999.5.$$

Dobili smo linearni trend nezaposlenosti $x(t) = -387.3t + 9999.5$. Koeficijent $a = -387.3$ pokazuje prosječni godišnji trend pada broja nezaposlenih. Očekivanja broja nezaposlenih u neredne dvije godine su: $x(16) = 3802.7 \approx 3803$, $x(17) = 3415.4 \approx 3415$. Nadalje, ako 15% nezaposlenih iznosi 4215 tada je 1%=281 odnosno 10%=2810. Tražimo t za koji će biti $x(t) \leq 2810$. Dobivamo $t \geq 18.563...$ što znači da će, ako se ovakav trend nastavi, nezaposlenost pasti ispod 10% tijekom 2019. godine.

ZADACI

1. Odredite zbroj i zbroj kvadrata rezidualnih odstupanja podataka (0,0), (1,2), (2,2), (3,4) od linearnog trenda $x(t)=t+1$. Da li je to optimalni linearni trend? Zašto? Ako nije odredite optimalni trend i navedena odstupanja.
2. Znamo da kroz dvije točke prolazi samo jedan pravac. Kroz tri točke, koje nisu na istom pravcu, takav pravac ne postoji ali postoji samo jedan pravac koji prolazi najbliže tim točkama. Odredite taj pravac za točke (0,0), (1,1), (2,3). Koliko je ukupno odstupanje pravca od tih točaka?
3. U jednom turističkom mjestu već petu godinu za redom održava se ljetni festival. Zbog atraktivnog programa iz godine u godinu bilježi se porast broja posjetitelja festivala: u prvoj godini 34000, u drugoj 36000, u trećoj 39000, u četvrtoj 42000 te u petoj 44000. Ako se ovakav trend nastavi koliko posjetitelja možemo očekivati u slijedeće tri godine?
4. Tokom dana evidentiran je broj korisnika internet aplikacija jedne poslovne banke:

Vrijeme (h)	**0-3**	**3-6**	**6-9**	**9-12**	**12-15**	**15-18**	**18-21**	**21-24**
Broj korisnika	50	400	2200	1500	2100	1100	300	150

   Napravite prikaz linijskim dijagramom. Odredite jednadžbu linearnog trenda i prikažite ga na istom dijagramu. Da li on dobro reprezentira pojavu?
5. Učenik treba pročitati knjigu od 400 stranica. Odlučio je čitati svaki dan. Ako je prvi dan pročitao 15 stranica, drugi 10, treći 15, četvrti 20, peti 10 a šesti 20, kada će, uz ovakav trend, pročitati knjigu?
6. U jednom proizvodnom pogonu utrošak električne energije tokom tjedna bio je: u ponedjeljak, 530 kWh, u srijedu 600 kWh, u petak 640 kWh. Kako je u utorak i četvrtak brojilo bilo u kvaru, treba napraviti procjenu potrošnje za ta dva dana.
7. Tokom osam tjedana ($t_i = 1,2,3,...,8$) bilježite broj izostanaka u vašem razredu (x_i) i odredite jednadžbu linearnog trenda. Napravite procjenu za slijedeća četiri tjedna te usporedite sa stvarnim stanjem. Da li linearni model adekvatno reprezentira pojavu?
8. Sakupite podatke o cijeni jedne ili više vrsta goriva (benzin, plin, dizel i sl.) tokom nekoliko (5 do 10) uzastopnih tjedana te odredite jednadžbu linearnog trenda. Koristeći dobivenu jednadžbu procijenite cijenu za iduća 2 do 3 tjedna nakon čega ćete moći utvrditi realnost procjene.
9. Za koje vrijednosti parametara a, b izraz

$$(a+b-1)^2+(2a+b-1)^2+(3a+b-2)^2+(4a+b-2)^2$$

   poprima najmanju moguću vrijednost i koliko ona iznosi?
10. Jednoliko ubrzano gibanje $x(t)=0.5\alpha t^2$, $t \geq 0$, gdje je α konstantno ubrzanje (akceleracija), aproksimirajte jednolikim gibanjem po pravcu u prve četiri jedinice vremena, $t \in \{0,1,2,3\}$. Kolika je prosječna brzina u tom razdoblju?

11. Put čestice u elektromagnetskom polju opisan je jednadžbom $x(t)=3600/(t+1)^2$, gdje je $t\geq 0$ vrijeme u mikrosekundama (μs) a x put u metrima. Aproksimirajte put u trećoj, četvrtoj i petoj μs jednolikim gibanjem po pravcu. Kolika je prosječna brzina čestice u tom razdoblju? Napomena: svako gibanje se u dovoljno kratkim vremenskim intervalima može dobro aproksimirati jednolikim gibanjem po pravcu.

RJEŠENJA

1. Imamo $u_1 = x_i-(t_i+1)$, $i=1,2,3,4$. Dakle,

$$\sum_{i=1}^{4} u_i = (-1)+0+(-1)+0=-2, \qquad \sum_{i=1}^{4} u_i^2 = 2.$$

Ne radi se o optimalnom trendu jer zbroj odstupanja nije 0 (relacija (17)). Koristeći (18) imamo $t_A=1.5$, $x_A=2$, pa je $a=0.75$, $b=0.875$. Dakle, $x(t)=0.75t+0.875$ je jednadžba optimalnog trenda. Odstupanja su

$$\sum_{i=1}^{4} u_i = (-0.875)+0.375+(-0.375)+0.875=0, \qquad \sum_{i=1}^{4} u_i^2 = 1.8125.$$

2. Imamo $t_1=x_1=0$, $t_2=x_2=1$, $t_3=2$, $x_3=3$ $(n=3)$. Koristeći izraze (18) dobivamo

$$t_A=1,\quad x_A=\frac{4}{3},\quad a=\frac{(0\cdot 0+1\cdot 1+2\cdot 3)-3\cdot\frac{4}{3}\cdot 1}{(0^2+1^2+2^2)-3\cdot 1^2}=\frac{3}{2},\quad b=\frac{4}{3}-\frac{3}{2}\cdot 1=-\frac{1}{6}.$$

Jednadžba traženog pravca je $x=\frac{3}{2}t-\frac{1}{6}$. Odstupanja su

$$u_1=x_1-x(0)=0-\left(\frac{3}{2}\cdot 0-\frac{1}{6}\right)=\frac{1}{6},\qquad u_2=x_2-x(1)=1-\left(\frac{3}{2}\cdot 1-\frac{1}{6}\right)=-\frac{1}{3},$$
$$u_3=x_3-x(2)=3-\left(\frac{3}{2}\cdot 2-\frac{1}{6}\right)=\frac{1}{6}\quad\Rightarrow\quad u_1+u_2+u_3=0,\quad u_1^2+u_2^2+u_3^2=\frac{1}{6}.$$

Dobiveno ukupno odstupanje je najmanje moguće. Svaki drugi pravac ima od zadanih točaka veće ukupno odstupanje od ovog.

3. Određujemo jednadžbu linearnog trenda. Imamo $t_A=3$, $x_A=39000$, pa je $a=2600$, $b=31200$. Iz jednadžbe trenda $x(t)=2600t+31200$, slijede procjene: $x(6)=46800$, $x(7)=49400$ i $x(8)=52000$.

4. Za vrijednosti vremenske koordinate uzimamo sredine razreda: 1.5, 4.5, 7.5 itd., koje koristimo za prikaz dijagramom i za jednadžbu. Imamo $t_A=12$, $x_A=975$, pa je

$$x=-\frac{625}{63}t+\frac{22975}{21}\approx -9.920635t+1094.04762.$$

Iz dijagrama je odmah vidljivo da pojava nema linearni trend pa dobiveni model ne reprezentira pojavu (vidjeti također naredno poglavlje *Korelacijska analiza*).

5. Kako očigledno tempo čitanja ne predstavlja linearni trend, napravit ćemo kumulativni niz ukupno pročitanih stranica po danima: (1,15), (2,25), (3,40), (4,60), (5,70), (6,90). Za ovaj niz određujemo linearni trend. Imamo $t_A = 3.5$, $x_A = 50$, pa je $a = 106/7$, $b = -3$. Iz jednadžbe trenda $x(t) = 15.142857t - 3$ imamo $x(t) = 400 \Rightarrow t = 26.6132...$, što znači da će čitanje završiti 27. dan. Napomena: zadatak smo mogli riješiti i jednostavnije, naime, budući da je učenik prosječno pročitao $90:6 = 15$ stranica dnevno, čitanje će trajati $400:15 = 26.666... \approx 27$ dana.
6. Tražimo linearni trend za točke (1,530), (3,600), (5,640). Imamo $t_A = 3$, $x_A = 590$, pa je $a = 27.5$, $b = 507.5$. Iz jednadžbe trenda $x(t) = 27.5t + 507.5$ slijede procjene za utorak, $x(2) = 562.5$ kWh i za četvrtak, $x(4) = 617.5$ kWh.
7. Koristite stvarne podatke.
8. Koristite prikupljene podatke. Cijena energenata u praksi sadrži i cikličku i slučajnu komponentu (osim trenda) koje, ako se pojave u periodu koji procjenjujemo trendom, mogu uzrokovati nepreciznost procjene.
9. Iz relacije (17) slijedi da trebamo naći parametre linearnog trenda za točke (1,1), (2,1), (3,2) i (4,2). Imamo $t_A = 2.5$, $x_A = 1.5$, pa je $a = 0.4$, $b = 0.5$. Za ove vrijednosti parametara navedeni izraz ima minimalnu vrijednost koja iznosi 0.2.
10. Imamo točke $(0,\,0),\ (1,\,\alpha/2),\ (2,\,2\alpha),\ (3,\,9\alpha/2)$, pa je $t_A = 1.5$, $x_A = 1.75\alpha$, iz čega slijedi $a = 1.5\alpha,\ b = -0.5\alpha$. Dakle, linearna aproksimacija je $x(t) = 1.5\alpha \cdot t - 0.5\alpha$, pa je prosječna brzina parametar a, tj. 1.5α (razlikujte slovo a od $\alpha =$ alfa).
11. Imamo točke (3,225), (4,144) i (5,100), pa je $t_A = 4$, $x_A = 469/3$, iz čega dobivamo $a = -62.5,\ b = 1219/3$, odnosno $x(t) = -62.5t + 406.33333...$, pa je prosječna brzina $-62.5\,\frac{\text{m}}{\mu\text{s}} = -62.5\,\frac{\text{m}}{10^{-6}\text{s}} = -62.5 \cdot 10^{6}\,\frac{\text{m}}{\text{s}} = -62500\,\frac{\text{km}}{\text{s}}$, pri čemu negativni predznak označava približavanje čestice.

ODNOSI MEĐU POJAVAMA

Svijet u kome živimo jedna je cjelina. Pojave koje uočavamo i doživljavamo međusobno su povezane na direktan ili indirektan način. Budući da ih često promatramo izdvojeno, povezanost nije uvijek očigledna. Utvrđivanje i razumijevanje veza među pojavama važan je čimbenik u svim ljudskim djelatnostima kao i u svakodnevnom životu. Ako dovoljno dobro poznajemo prirodu njihove povezanosti, možemo utjecati i kontrolirati razvoj pojedinih situacija u željenom smjeru. Ispitivanje odnosa među pojavama je predmet korelacijske i regresijske analize. U korelacijskoj analizi ustanovljuje se postojanje (ili nepostojanje) povezanosti među pojavama te njen oblik, jačina i smjer ne ulazeći u uzročno-posljedičnu vezu (koja od pojava je uzrok – nezavisna varijabla a koja posljedica – zavisna varijabla). Regresijskom analizom utvrđujemo analitički oblik (uzročno-posljedičnu vezu) između zavisne i jedne ili više nezavisnih varijabli.

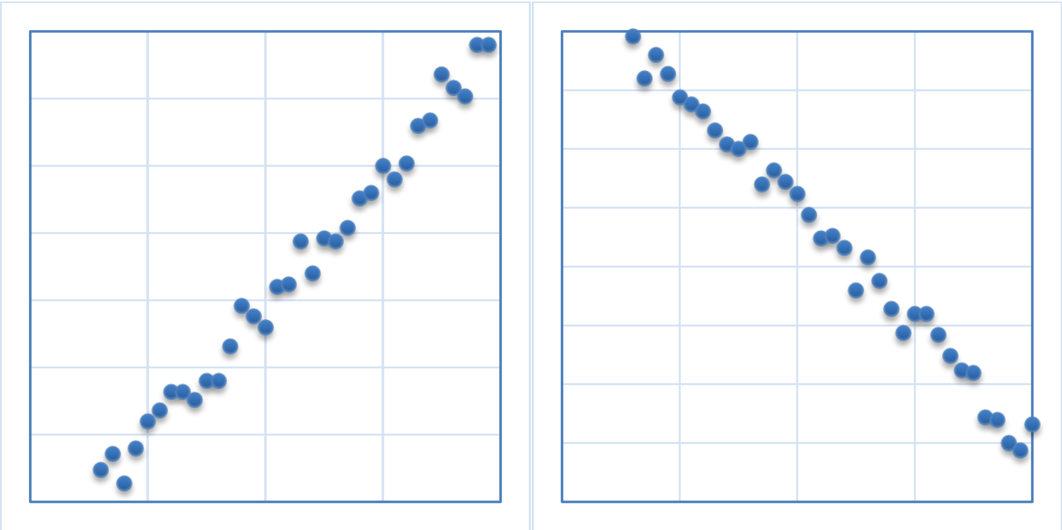
KORELACIJSKA ANALIZA

Veza (korelacija) među pojavama opisuje se mjerama *kovarijacije*. Uspoređujući tijek pojava ustanovljujemo da li postoji veza između njih, te ako postoji koje je jačine, oblika i smjera. Prema jačini veza može biti *funkcionalna* (jaka veza koja se može opisati analitički – formulom) ili *statistička* (slabija veza sa slučajnim utjecajima). Prema obliku može biti *linearna* i *nelinearna* te *jednostavna* (između dviju pojava) i *višestruka* (istovremena povezanost više pojava). Prema smjeru veza može biti *pozitivna* (pojave se mijenjaju u istom smjeru) i *negativna* (promjene su suprotnog smjera). Navedena svojstva mjerimo različitim statističkim pokazateljima (mjerama kovarijacije) od kojih upoznajemo dijagram rasipanja, kovarijancu i koeficijente korelacije.

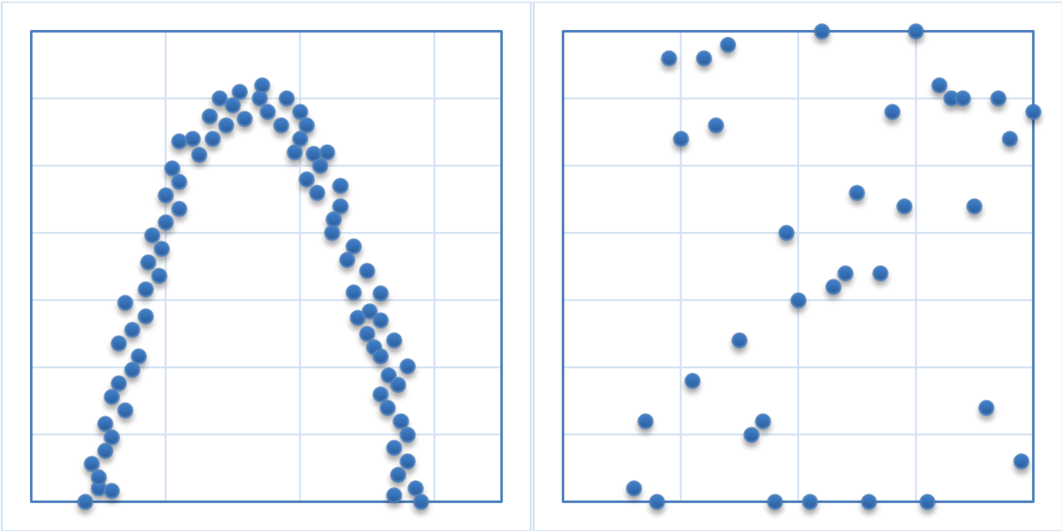
Prvi korak u ispitivanju međusobnog odnosa dviju pojava X i Y može biti grafički prikaz. Parove vrijednosti

$$(x_1, y_1),\ (x_2, y_2),\ \dots,\ (x_{n-1}, y_{n-1}),\ (x_n, y_n)$$

čiji se odnos ispituje prikazujemo u pravokutnom koordinatnom sustavu. Dobiveni prikaz je *dijagram rasipanja*. Pri tome je važno prikazati što veći broj podataka jer se time povećava pouzdanost dobivenih zaključaka. Tipične izglede dijagrama rasipanja vidimo na slikama 41 i 42.



Slika 41. Dijagram rasipanja – linearna povezanost



Slika 42. Dijagram rasipanja – nelinearna povezanost i nepovezanost

Na slici 41 vrijednosti se, manje ili više, grupiraju oko zamišljenog pravca što ukazuje na linearni oblik veze. Nagib pravca ukazuje na smjer veze (pozitivna na slici lijevo a negativna na slici desno) dok raspršenost točaka oko pravca ukazuje na jačinu veze (veća raspršenost – slabija veza i obratno). Na slici 42 lijevo vidimo grupiranje točaka oko neke zamišljene krivulje što upućuje na postojanje nelinearne (krivo linijske) veze dok na slici desno ne uočavamo povezanost među pojavama.

Vezu među pojavama izražavamo i brojčanim statističkim pokazateljima od kojih navodimo najvažnije.

DEFINICIJA 10. ZA NIZ PAROVA NUMERIČKIH VRIJEDNOSTI DVIJU POJAVA $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ JE

$$\mu_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_A)(y_i - y_A) \qquad \text{(kovarijanca),}$$

$$r = \frac{\mu_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - x_A)(y_i - y_A)}{\sqrt{\left(\sum_{i=1}^{n}(x_i - x_A)^2\right)\left(\sum_{i=1}^{n}(y_i - y_A)^2\right)}} \qquad \text{(koeficijent linearne korelacije),}$$

GDJE SU x_A I y_A ARITMETIČKE SREDINE A σ_x I σ_y STANDARDNE DEVIJACIJE NIZOVA $x_1, x_2, ..., x_n$ I $y_1, y_2, ..., y_n$.

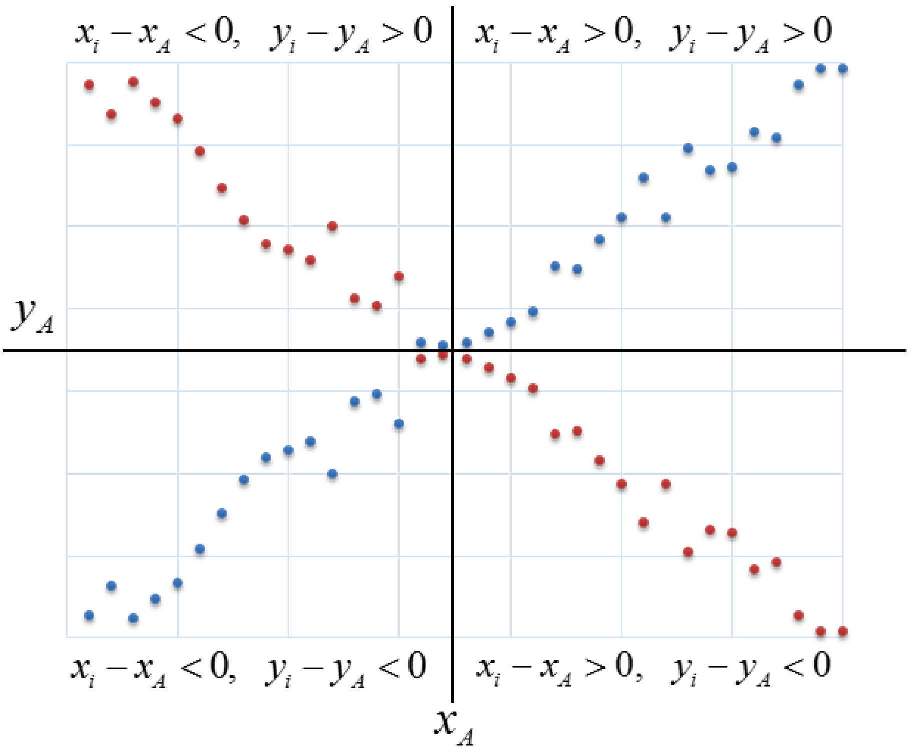
Primijetimo analogiju ove definicije sa pokazateljima za jednu varijablu (definicije 5, 6 i 7). Naime za $x_i = y_i,\ i = 1, 2, ..., n$ kovarijanca postaje varijanca a $r = \alpha_2 = 1$. Kovarijanca je izražena produktom mjernih jedinica promatranih pojava dok je koeficijent korelacije relativni pokazatelj (neimenovani broj). Daljim razvojem izraza za navedene pokazatelje dobivamo

$$\mu_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - x_A \cdot \frac{1}{n}\sum_{i=1}^{n} y_i - y_A \cdot \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} x_A y_A = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - x_A y_A - y_A x_A + x_A y_A,$$

što zajedno sa relacijom (14), primijenjenom na σ_x i σ_y, daje

$$\mu_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - x_A y_A, \qquad r = \frac{\sum_{i=1}^{n} x_i y_i - n x_A y_A}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n x_A^2\right) \cdot \left(\sum_{i=1}^{n} y_i^2 - n y_A^2\right)}}. \tag{19}$$

Predznak ovih pokazatelja izražava smjer veze među pojavama. Naime, isti predznaci faktora $x_i - x_A$ i $y_i - y_A$ daju pozitivan produkt a različiti negativan (slika 43).



Slika 43. Pozitivna i negativna kovarijacija

Koeficijent korelacije r (Pearsonov koeficijent) uvijek pripada intervalu $[-1, 1]$ a njegova vrijednost pokazuje jakost međusobne veze dviju pojava. Tako za $|r| = 1$ imamo najjaču (funkcionalnu) vezu, za $0.8 < |r| < 1$ imamo jaku vezu, za $0.5 < |r| \leq 0.8$ veza je srednje jačine, za $0 < |r| \leq 0.5$ veza je slaba a za $r = 0$ ne postoji veza (beznačajna je). U svim navedenim slučajevima za $r > 0$ veza je pozitivnog a za $r < 0$ negativnog smjera.

Primjer 21. Uz osnovnu cijenu novog šampona od 40.49 kuna po komadu prosječna tjedna prodaja u jednom trgovačkom centru bila je 401 komad. U svrhu promoviranja proizvoda u slijedeća četiri tjedna provedena je akcijska prodaja uz postupno smanjenje cijene. Rezultati prodaje dani su u tabeli 15. Ispitajmo korelaciju između cijene i potražnje.

Tjedan	1	2	3	4	5
Cijena	40.49	35.99	31.49	29.99	27.99
Potražnja	401	425	442	460	462

Tabela 15. Podaci o prodaji novog šampona

Neka su x_i cijene a y_i prodane količine, $i=1,2,3,4,5$ (označiti možemo i obrnuto jer se za ove pokazatelje ne pravi razlika između zavisne i nezavisne varijable). Imamo

$$x_A=\frac{165.95}{5}=33.19,\ \sigma_x=\sqrt{\frac{101.3}{5}}=4.50111...,\ y_A=\frac{2190}{5}=438,\ \sigma_y=\sqrt{\frac{2614}{5}}=22.86482...,$$

pa je

$$\mu_{xy}=\frac{-508.5}{5}=-101.7, \qquad r=\frac{\mu_{xy}}{\sigma_x\sigma_y}=-0.988174... .$$

Vidimo vrlo jaku (skoro funkcionalnu) negativnu linearnu vezu cijene i potražnje (negativna znači da porast jedne veličine uzrokuje smanjenje druge i obratno).

Ako vrijednosti svake pojave rangiramo brojevima $1,2,...,n$ od najmanje vrijednosti prema najvećoj ili obrnuto, tada njihovu povezanost možemo izraziti i *koeficijentom korelacije ranga* (Spearmanov koeficijent),

$$r_S=1-\frac{6}{n^3-n}\sum_{i=1}^{n}\left[\text{rang}(x_i)-\text{rang}(y_i)\right]^2 .$$

Nazivnik n^3-n je izabran tako da koeficijent uvijek pripada intervalu $[-1,1]$. Naime ako su svi parovi vrijednosti (x_i,y_i) istog ranga, tada je $\text{rang}(x_i)-\text{rang}(y_i)=0$ pa je $r_S=1$ što pokazuje potpunu pozitivnu korelaciju ranga. U suprotnom slučaju, kad je prvi član po rangu (1) u paru sa zadnjim članom (n), drugi (2) sa predzadnjim ($n-1$), treći (3) sa pred-predzadnjim ($n-2$), …, $k-$ti sa $(n-k+1)-$im, …, predzadnji ($n-1$) sa drugim (2) i zadnji (n) sa prvim (1), suma u izrazu iznosi

$$\sum_{i=1}^{n}\left[\text{rang}(x_i)-\text{rang}(y_i)\right]^2=2\cdot\left[(n-1)^2+(n-3)^2+(n-5)^2+...\right]=2\cdot\frac{n^3-n}{6}=\frac{n^3-n}{3},$$

pa je $r_S = 1-2 = -1$ što pokazuje potpunu negativnu korelaciju ranga. Za $r_S = 0$ korelacija ranga ne postoji (beznačajna je). Primijetimo da kod ovog pokazatelja nisu bitne vrijednosti varijabli već samo njihov rang (redni broj u redoslijednom – rastućem ili padajućem poretku).

U primjeru 21 (tabela 15) navedene cijene imaju rang 1,2,3,4,5 (od najveće do najmanje) a pridružene količine 5,4,3,2,1, pa je

$$r_S = 1-\frac{6}{5^3-5}\cdot\left[(1-5)^2+(2-4)^2+(3-3)^2+(4-2)^2+(5-1)^2\right] = 1-\frac{6}{120}\cdot 40 = -1,$$

što pokazuje potpunu negativnu korelaciju između cijena i prodanih količina.

Primjer 22. Na ispitima iz matematike i hrvatskog jezika na državnoj maturi postotni broj bodova koji su učenici ostvarili naveden je u tabeli 16. Ispitajmo postoji li korelacija ranga između ta dva predmeta.

Učenik	A	B	C	D	E	F	G	H	K	M
Matematika	60	50	85	75	60	55	80	70	40	45
Hrvatski	70	55	90	85	70	60	70	80	50	65

Tabela 16. Postotni bodovi na državnoj maturi

Bodovima iz svakog predmeta pridružimo njihov rang, npr. počevši od najmanjeg.

Učenik	A	B	C	D	E	F	G	H	K	M
Matematika	5.5	3	10	8	5.5	4	9	7	1	2
Hrvatski	6	2	10	9	6	3	6	8	1	4

Tabela 17. Rang postotnih bodova na državnoj maturi

Učenici A i E iz matematike imaju rang 5 i 6 sa istim brojem bodova pa im se pridružuje aritmetička sredina $(5+6)/2$. Slično tome, učenici A, E i G iz hrvatskog imaju rang 5, 6 i 7 sa istim brojem bodova pa im se pridružuje rang $(5+6+7)/3$. Sada je

$$\begin{aligned} r_S &= 1-\frac{6}{10^3-10}\cdot\left[(5.5-6)^2+(3-2)^2+(10-10)^2+(8-9)^2+(5.5-6)^2+(4-3)^2+(9-6)^2\right.\\ &\qquad \left.+(7-8)^2+(1-1)^2+(2-4)^2\right] = 1-\frac{6}{990}\cdot 17.5 = \frac{59}{66} = 0.893939\ldots . \end{aligned}$$

Vidimo jaku pozitivnu korelaciju ranga što znači da učenici sa boljim rezultatom iz jednog predmeta imaju ujedno i bolji rezultat iz drugog predmeta i obratno.

ZADACI

1. Dokažite da ako postoji linearna funkcionalna veza $y_i = ax_i + b$, $i = 1, 2, ..., n$ tada je $r = 1$ za $a > 0$ odnosno $r = -1$ za $a < 0$.
2. Odredite kovarijancu i koeficijent linearne korelacije za niz $(x_1, y_1), (x_2, y_2)$.
3. Da li niz točaka (1,3), (2,1), (3,3), (4,2), (5,4) ima izraženiju rastuću ili padajuću tendenciju?
4. Odredite koeficijent linearne korelacije za drugi i četvrti zadatak iz prethodnog poglavlja *Vremenski trendovi*.
5. Pratite kroz neko vrijeme (bar 10 dana) sa koliko osoba ste tokom dana (ne računajući vrijeme provedeno u školi) kontaktirali mobitelom (razgovarali ili izmjenjivali poruke) a sa koliko ste kontaktirali direktno (razgovarali u živo). Prikažite podatke za te dvije veličine dijagramom rasipanja i izračunajte koeficijent linearne korelacije.
6. Tokom zimskog perioda u jednom bolničkom centru praćeni su pacijenti koji su imali simptome sezonske gripe. Pacijentima je kod prijema mjerena tjelesna temperatura T. Dobiveni rezultati mjerenja dani su u tabeli.

Broj pacijenata	9	16	19	26	27	20	14	5
$T \geq 38^o C$	3	7	9	14	15	9	5	0
$37^o C < T < 38^o C$	2	5	7	9	8	7	6	2
$T \leq 37^o C$	4	4	3	3	4	4	3	3

Koeficijentom korelacije ispitajte povezanost simptoma gripe sa tjelesnom temperaturom (odredite sva tri koeficijenta) te interpretirajte rezultate.

7. Ispitajte korelaciju između broja prodanih kozmetičkih proizvoda nove linije i troškova marketinga u zadnjih šest mjeseci.

Mjesec	**1**	**2**	**3**	**4**	**5**	**6**
Troškovi marketinga	12000	13000	14000	15000	16000	17000
Broj prodanih proizvoda	1925	1920	1933	2122	2020	2188

8. Da li postoji povezanost broja domaćih (x_i) i stranih (y_i) gostiju jednog hotela na temelju podataka iz pet nasumce odabranih razdoblja u godini?

Razdoblje	**1**	**2**	**3**	**4**	**5**
x_i	175	502	40	885	310
y_i	90	269	25	505	180

9. U svrhu organiziranja specijalnih aktivnosti u jednom zabavnom parku, praćen je broj posjetitelja parka tokom tjedna, od ponedjeljka (1) do nedjelje (7).

Dan	1	2	3	4	5	6	7
Broj posjetitelja	318	299	310	281	332	590	509

Odredite prosječan broj posjetitelja po danu te standardnu devijaciju i koeficijent varijacije. Koeficijentom linearne korelacije utvrdite da li postoji povezanost broja posjetitelja sa danima u tjednu.

10. Sakupite podatke o cijeni nafte na inozemnom (npr. mediteranskom, evropskom ili svjetskom) i domaćem tržištu tokom nekoliko uzastopnih tjedana ili mjeseci te odredite i interpretirajte stupanj njihove međusobne korelacije.
11. Postoji li povezanost između ukupnog iznosa štednje građana i izdanih kredita jedne poslovne banke u zadnjih 6 godina? Iznosi su navedeni u milijunima kuna. Povezanost ispitajte koeficijentom linearne korelacije i koeficijentom korelacije ranga.

Godina	1	2	3	4	5	6
Štednja	1175	1302	1400	1250	980	1051
Krediti	905	820	800	960	1041	880

12. Na testiranje vještina i sposobnosti za novo radno mjesto prijavilo se 10 kandidata različite starosne dobi. Rezultati testa bili su slijedeći:

Kandidat	A	B	C	D	E	F	G	H	K	L
Bodovi	75	58	85	80	90	60	52	45	40	75
Dob (god.)	30	29	38	25	33	24	24	30	25	28

Postoji li značajna korelacija ranga između broja bodova i godina starosti?

RJEŠENJA

1. Kako je $y_i - y_A = (ax_i + b) - (ax_A + b) = a(x_i - x_A)$, iz definicije 10 slijedi

$$r = \frac{\sum_{i=1}^{n}(x_i - x_A)\cdot a(x_i - x_A)}{\sqrt{\left(\sum_{i=1}^{n}(x_i - x_A)^2\right)\left(\sum_{i=1}^{n}a^2(x_i - x_A)^2\right)}} = \frac{a\sum_{i=1}^{n}(x_i - x_A)^2}{\sqrt{a^2\left(\sum_{i=1}^{n}(x_i - x_A)^2\right)^2}} = \frac{a}{\sqrt{a^2}} = \frac{a}{|a|},$$

što daje $+1$ za $a > 0$ odnosno -1 za $a < 0$.

2. Kako je $x_A = (x_1 + x_2)/2$, $y_A = (y_1 + y_2)/2$, koristeći definiciju 10 dobivamo

$$\mu_{xy} = \frac{1}{2}\left(\frac{x_1 - x_2}{2}\cdot\frac{y_1 - y_2}{2} + \frac{x_2 - x_1}{2}\cdot\frac{y_2 - y_1}{2}\right) = \frac{(x_1 - x_2)(y_1 - y_2)}{4},$$

$$r = \frac{(x_1 - x_2)(y_1 - y_2)}{\sqrt{(x_1 - x_2)^2(y_1 - y_2)^2}} = \frac{(x_1 - x_2)(y_1 - y_2)}{|x_1 - x_2|\cdot|y_1 - y_2|} = \begin{cases} +1 & \text{za} \quad (x_1 - x_2)(y_1 - y_2) > 0, \\ -1 & \text{za} \quad (x_1 - x_2)(y_1 - y_2) < 0. \end{cases}$$

Primijetimo da r možemo dobiti i iz prethodnog zadatka iz jednadžbe pravca kroz dvije točke, gdje je koeficijent smjera $a = (y_1 - y_2)/(x_1 - x_2) = (y_2 - y_1)/(x_2 - x_1)$.

3. Trebamo samo odrediti predznak kovarijance. Kako je $x_A = 3$, $y_A = 2.6$, koristeći definiciju 10 dobivamo $\mu_{xy} = +0.6$, pa je rastuća tendencija izraženija.
4. U 2. zadatku je $r = 0.98198...$ što je vrlo visoka korelacija a u 4. je $r = -0.083156...$ što pokazuje da linearna veza među pojavama gotovo ne postoji a što je bilo odmah vidljivo iz grafičkog prikaza.
5. Koristite svoje podatke. Rezultati vas mogu iznenaditi.
6. Jedna varijabla (X) je broj pacijenata a Y je svaka od tri kategorije prema izmjerenoj temperaturi (38+, 37-38 i 37–). Imamo $x_A = 17$.

   Za kategoriju 38+ je $y_A = 7.75$, $r = \dfrac{275}{\sqrt{412 \cdot 185.5}} = 0.9947458...$.

   Za kategoriju 37–38 je $y_A = 5.75$, $r = \dfrac{133}{\sqrt{412 \cdot 47.5}} = 0.9507278...$.

   Za kategoriju 37– je $y_A = 3.5$, $r = \dfrac{4}{\sqrt{412 \cdot 2}} = 0.1393466...$.

   Vidimo vrlo visoku korelaciju broja pacijenata sa prvom i drugom kategorijom što pokazuje da je povišena temperatura (37-38) a pogotovo značajnije povišena temperatura (preko 38) jedan od sigurnih simptoma sezonske gripe.
7. Imamo $x_A = 14500$ i $y_A = 2018$, pa iz definicije 10 slijedi

$$r = \frac{902\,000}{\sqrt{17\,500\,000 \cdot 65198}} = 0.844443...,$$

   a što pokazuje značajan utjecaj marketinga na broj prodanih proizvoda.
8. Imamo $x_A = 382.4$ i $y_A = 213.8$, pa koristeći relaciju (19) dobivamo

$$r = \frac{654513 - 408785.6}{\sqrt{(1163554 - 731148.8) \cdot (368511 - 228552.2)}} = 0.9988678...,$$

   što pokazuje vrlo visoku korelaciju broja domaćih i stranih gostiju. To znači da u periodima kad je više (manje) gostiju tada je više (manje) i domaćih i stranih.
9. Prosječan broj posjetitelja je $y_A = 377$ pa je $\sigma_y = \sqrt{12584} = 112.178429... \approx 112$ te $V_y = 0.297555... \approx 29.8\%$ a što pokazuje znatne oscilacije broja posjetitelja tokom tjedna. Nadalje, $r = 1177 / \sqrt{2466464} = 0.74944...$ pa je korelacija srednje jačine (prema kraju tjedna porast broja posjetitelja).
10. Koristite stvarne podatke.
11. Imamo $x_A = 1193$ (štednja) i $y_A = 901$ (krediti), pa je $r = -0.754245...$ što pokazuje umjerenu negativnu vezu (veća štednja a manje kredita i obratno). Rangovi u rastućem redoslijedu su 3,5,6,4,1,2 za štednju te 4,2,1,5,6,3 za kredite, pa je

$$r_S = 1 - \frac{6}{6^3 - 6}\left[(3-4)^2 + (5-2)^2 + (6-1)^2 + (4-5)^2 + (1-6)^2 + (2-3)^2\right] = -\frac{27}{35}.$$

Dakle, $r_S = -0.77142857...$, pa imamo isti zaključak

12. Rangovi su 4.5, 7, 2, 3, 1, 6, 8, 9, 10, 4.5 za bodove te 3.5, 5, 1, 7.5, 2, 9.5, 9.5, 3.5, 7.5, 6 za godine starosti, iz čega slijedi

$$r_S = 1 - \frac{6}{10^3 - 10}\left[1^2 + 2^2 + 1^2 + (-4.5)^2 + (-1)^2 + (-3.5)^2 + (-1.5)^2 + 5.5^2 + 2.5^2 + (-1.5)^2\right]$$

a što iznosi $r_S = 0.51212121...$, pa korelacija nije značajna (osrednja je).

REGRESIJSKA ANALIZA

Nakon što smo korelacijskom analizom utvrdili postojanje veze među pojavama, cilj regresijske analize je naći analitički izraz (formulu) koji tu vezu opisuje. Analiza počiva na utvrđivanju zavisne i nezavisne (jedne ili više) varijable te izboru modela kojim vezu opisujemo. Model se sastoji od jedne ili više jednadžbi određenog tipa (linearne, kvadratne, eksponencijalne i sl.) i postupaka kojim se procjenjuju njihovi parametri (koeficijenti). Kao što smo već vidjeli najvažnije mjesto u statističkoj analizi ima *linearni regresijski model*. U tom modelu odnos dviju (ili više) pojava opisuje se linearnom regresijskom jednadžbom. Ako su dvije pojave X i Y dane statističkim nizom uređenih parova numeričkih vrijednosti

$$(x_1, y_1),\ (x_2, y_2),\ ...,\ (x_{n-1}, y_{n-1}),\ (x_n, y_n),$$

pri čemu je X nezavisna a Y zavisna varijabla, tada je linearna regresijska jednadžba

$$y(x) = ax + b, \qquad a = ?,\ b = ?.$$

Jasno je da se općenito, za bilo koje vrijednosti parametara a i b, vrijednosti y_i i $y(x_i)$ razlikuju, pa imamo

$$y_i = y(x_i) + u_i \ \Rightarrow\ u_i = y_i - y(x_i) = y_i - ax_i - b, \quad i = 1, 2, ..., n,$$

gdje su u_i rezidualna odstupanja zadanih podataka (točaka) od regresijskog pravca. Parametre a i b u regresijskoj jednadžbi određujemo uz uvjet da je ukupni zbroj rezidualnih odstupanja jednak nuli a zbroj njihovih kvadrata minimalan (metoda najmanjih kvadrata). Ovi uvjeti su dani relacijom (17) u poglavlju *Vremenski trendovi* gdje imamo cjelovit izvod metode. Naime, vremenski trend je specijalni slučaj veze dviju pojava: promatrane pojave (zavisna varijabla) i vremena (nezavisna varijabla). Traženi parametri dani su relacijom (18) koja u terminima ovih općenitih oznaka (x postaje y a t postaje x) sada glasi

$$a = \frac{\sum_{i=1}^{n}(y_i - y_A)(x_i - x_A)}{\sum_{i=1}^{n}(x_i - x_A)^2} = \frac{\sum_{i=1}^{n} x_i y_i - n x_A y_A}{\sum_{i=1}^{n} x_i^2 - n x_A^2}, \quad b = y_A - a x_A, \quad x_A = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad y_A = \frac{1}{n}\sum_{i=1}^{n} y_i. \qquad (20)$$

Parametar a (koeficijent smjera regresijskog pravca) predstavlja prosječan iznos za koji se promijeni zavisna varijabla y ako se nezavisna varijabla x poveća za jedinicu. Parametar b jednak je vrijednosti $y(0)$ i, ovisno o situaciji, ne mora imati prikladno praktično značenje.

Primjer 23. Odredimo linearnu regresijsku jednadžbu za potražnju iz primjera 21.

Imamo $x_A = 33.19$ i $y_A = 438$, pa je

$$a = \frac{72177.6 - 5\cdot 33.19\cdot 438}{5609.1805 - 5\cdot 33.19^2} = -5.0197433..., \qquad b = 438 - 33.19\cdot a = 604.6052813....$$

Dakle, linearna regresijska jednadžba funkcije potražnje (y) u ovisnosti o cijeni (x) glasi

$$y(x) \;=\; -5.02x + 604.605,$$

gdje smo parametre zaokružili na tri decimale. Dobivena jednadžba nam omogućava procjenu potražnje uz zadanu cijenu, npr. $y(50) \approx 353$, $y(25) \approx 479$ i sl., kao i procjenu cijene uz koju bi imali zadanu razinu potražnje, npr. $y(x) \geq 500 \Rightarrow x \leq 20.83$, $y(x) \leq 300 \Rightarrow x \geq 60.68$.

Vidjeli smo da zadane vrijednosti zavisne varijable y_i odstupaju od izračunatih regresijskih vrijednosti $y(x_i) = ax_i + b$ za iznos rezidualnog odstupanja $u_i = y_i - y(x_i) = y_i - ax_i - b$. Parametre a i b dobili smo iz uvjeta da je ukupni zbroj rezidualnih odstupanja jednak nuli a zbroj njihovih kvadrata minimalan. Upravo zbog toga je navedeni zbroj kvadrata mjera rasipanja podataka (stvarnih vrijednosti) oko regresije (izračunatih vrijednosti). Imamo

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} u_i \cdot u_i = \sum_{i=1}^{n} u_i \cdot (y_i - ax_i - b) = \sum_{i=1}^{n} u_i y_i - a\sum_{i=1}^{n} u_i x_i - b\sum_{i=1}^{n} u_i$$

Kako iz (20) slijedi $\sum_{i=1}^{n} x_i y_i - nx_A y_A = a\cdot\left(\sum_{i=1}^{n} x_i^2 - nx_A^2\right)$ i $b = y_A - ax_A$, dobivamo

$$\begin{aligned}\sum_{i=1}^{n} u_i x_i \;&=\; \sum_{i=1}^{n} (y_i - ax_i - b)x_i \;=\; \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i \;=\; \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - (y_A - ax_A)\cdot nx_A \\ &=\; \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - nx_A y_A + anx_A^2 \;=\; \sum_{i=1}^{n} x_i y_i - nx_A y_A - a\left(\sum_{i=1}^{n} x_i^2 - nx_A^2\right) \;=\; 0.\end{aligned}$$

Dakle, $\sum_{i=1}^{n} u_i x_i = 0$ što zajedno sa $\sum_{i=1}^{n} u_i = 0$ uvrštavamo u polaznu jednakost,

$$\sum_{i=1}^{n} u_i^2 \;=\; \sum_{i=1}^{n} u_i y_i - a\cdot 0 - b\cdot 0 \;=\; \sum_{i=1}^{n} (y_i - ax_i - b)y_i \;=\; \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} y_i\,.$$

Na temelju ovog izraza definiramo mjere raspršenosti podataka oko regresije.

DEFINICIJA 11. ZA NIZ PAROVA NUMERIČKIH VRIJEDNOSTI DVIJU POJAVA $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ I REGRESIJU $y(x) = ax + b$ DEFINIRANU RELACIJOM (20), JE

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - y(x_i)\right]^2 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} y_i\right) \quad \text{(varijanca regresije),}$$

$$\sigma_y = \sqrt{\sigma_y^2} \qquad \text{(standardna devijacija regresije),}$$

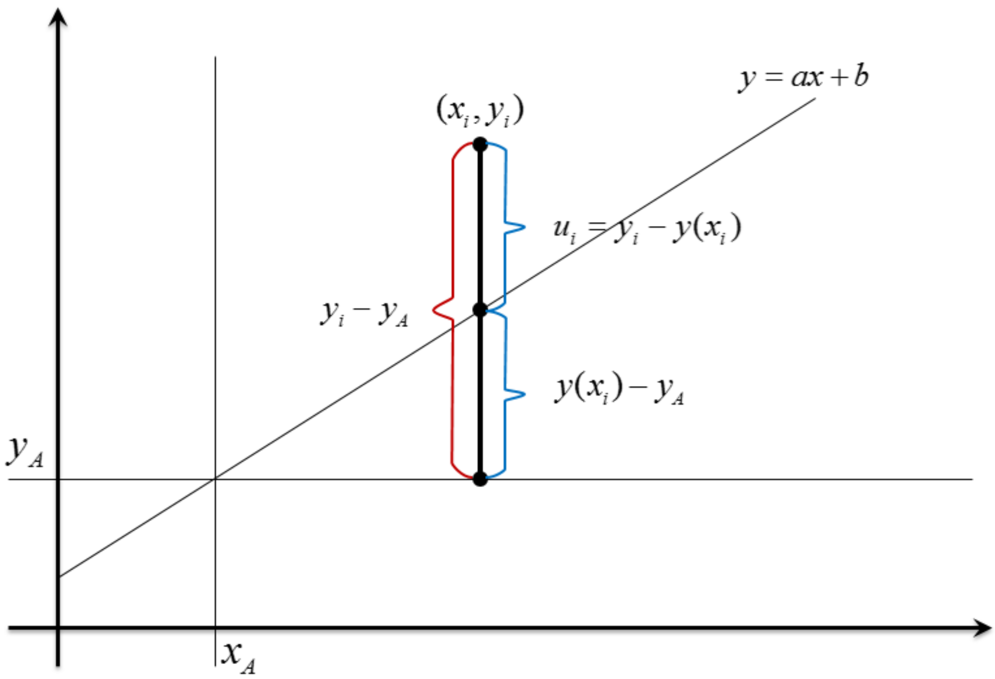
$$V_y = \frac{\sigma_y}{y_A} \qquad \text{(koeficijent varijacije regresije).}$$

Standardna devijacija je prosječno odstupanje stvarnih vrijednosti zavisne varijable od njenih regresijskih vrijednosti. Vidimo da je, kao i ranije, standardna devijacija izražena u mjernim jedinicama zavisne varijable dok je koeficijent varijacije neimenovani broj pa se može koristiti za usporedbu stupnja raspršenosti potpuno različitih pojava.

Da bi se u potpunosti procijenilo u kojoj mjeri je regresijski model reprezentativan tj. koliko dobro je varijabilnost zavisne varijable objašnjena dobivenom linearnom vezom sa nezavisnom varijablom, odstupanje pojedinog podatka od srednje vrijednosti rastavljamo na dvije komponente,

$$y_i - y_A = \left[y_i - y(x_i)\right] + \left[y(x_i) - y_A\right] = \left[y(x_i) - y_A\right] + u_i, \quad i = 1, 2, \ldots, n.$$

Komponenta $y(x_i) - y_A$ je odstupanje izračunate (regresijske) vrijednosti od sredine i to je *odstupanje protumačeno modelom* a druga komponenta je rezidualno odstupanje empirijske vrijednosti od regresijske vrijednosti koje *nije protumačeno modelom* (slika 44).



Slika 44. Protumačena i ne protumačena odstupanja u regresijskom modelu

Kvadriranjem i zbrajanjem pojedinačnih odstupanja dobivamo ukupna kvadratna odstupanja,

$$\sum_{i=1}^{n}(y_i - y_A)^2 = \sum_{i=1}^{n}\left[y(x_i) - y_A\right]^2 + 2\sum_{i=1}^{n}u_i\left[y(x_i) - y_A\right] + \sum_{i=1}^{n}u_i^2 .$$

Kako smo vidjeli da je $\sum_{i=1}^{n}u_i x_i = 0$ i $\sum_{i=1}^{n}u_i = 0$, imamo

$$\sum_{i=1}^{n}u_i\left[y(x_i) - y_A\right] = \sum_{i=1}^{n}u_i\left[ax_i + b - y_A\right] = a\sum_{i=1}^{n}u_i x_i + (b - y_A)\sum_{i=1}^{n}u_i = a\cdot 0 + (b - y_A)\cdot 0 = 0,$$

Iz čega slijedi

$$\underbrace{\sum_{i=1}^{n}(y_i - y_A)^2}_{ST} = \sum_{i=1}^{n}\left[y(x_i) - y_A\right]^2 + \sum_{i=1}^{n}u_i^2 = \underbrace{\sum_{i=1}^{n}\left[y(x_i) - y_A\right]^2}_{SP} + \underbrace{\sum_{i=1}^{n}\left[y_i - y(x_i)\right]^2}_{SR} .$$

Dakle, $ST = SP + SR$, gdje je:

ST – ukupan zbroj kvadrata odstupanja empirijskih vrijednosti zavisne varijable od prosjeka (mjera varijabilnosti empirijskih vrijednosti zavisne varijable od prosjeka),

SP – ukupan zbroj kvadrata odstupanja regresijskih vrijednosti od prosjeka (mjera varijabilnosti empirijskih vrijednosti zavisne varijable od prosjeka protumačenih modelom),

SR – ukupan zbroj kvadrata odstupanja empirijskih vrijednosti zavisne varijable od regresijskih vrijednosti (mjera varijabilnosti empirijskih vrijednosti zavisne varijable koji nisu protumačeni modelom).

Napišimo ove pokazatelje u razvijenom obliku. Izraz za SR slijedi direktno iz definicije 11, $SR = n\sigma_y^2$. Izraz za ST dobijemo kvadriranjem,

$$\sum_{i=1}^{n}(y_i - y_A)^2 = \sum_{i=1}^{n}y_i^2 - 2y_A\sum_{i=1}^{n}y_i + \sum_{i=1}^{n}y_A^2 = \sum_{i=1}^{n}y_i^2 - 2y_A\cdot ny_A + ny_A^2 = \sum_{i=1}^{n}y_i^2 - ny_A^2 ,$$

dok je $SP = ST - SR$. Dakle,

$$\begin{aligned}
ST &= \sum_{i=1}^{n}(y_i - y_A)^2 = \sum_{i=1}^{n}y_i^2 - ny_A^2 ,\\
SP &= \sum_{i=1}^{n}\left[y(x_i) - y_A\right]^2 = a\sum_{i=1}^{n}x_i y_i + b\sum_{i=1}^{n}y_i - ny_A^2 ,\\
SR &= \sum_{i=1}^{n}\left[y_i - y(x_i)\right]^2 = \sum_{i=1}^{n}y_i^2 - a\sum_{i=1}^{n}x_i y_i - b\sum_{i=1}^{n}y_i .
\end{aligned}$$

Osim pokazatelja navedenih u definiciji 11, kvaliteta regresijskog modela izražava se *koeficijentom determinacije*. To je omjer zbroja kvadrata odstupanja protumačenih modelom i ukupnog zbroja kvadrata odstupanja,

$$r^2 = \frac{SP}{ST} = 1 - \frac{SR}{ST} = \frac{\sum_{i=1}^{n}\left[y(x_i) - y_A\right]^2}{\sum_{i=1}^{n}(y_i - y_A)^2} = \frac{a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i - n y_A^2}{\sum_{i=1}^{n} y_i^2 - n y_A^2}, \quad 0 \le r^2 \le 1. \quad (21)$$

Ovaj koeficijent pokazuje u kojoj mjeri regresijski model određuje (determinira) promatranu pojavu. Što je r^2 veći to je model bolji (reprezentativniji). Koeficijent determinacije r^2 jednak kvadratu koeficijenta linearne korelacije r (vidjeti definiciju 10 i relaciju (19)), a što opravdava njegovu oznaku r^2. Za dokaz ove činjenice brojnik SP transformiramo koristeći izraz za b iz relacije (20),

$$SP = a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i - n y_A^2 = a\sum_{i=1}^{n} x_i y_i + (y_A - a x_A)\cdot n y_A - n y_A^2 = a\left(\sum_{i=1}^{n} x_i y_i - n x_A y_A\right),$$

te zatim izraz za a iz relacije (20) i relaciju (19),

$$r^2 = \frac{SP}{ST} = a\cdot\frac{\sum_{i=1}^{n} x_i y_i - n x_A y_A}{\sum_{i=1}^{n} y_i^2 - n y_A^2} = \frac{\sum_{i=1}^{n} x_i y_i - n x_A y_A}{\sum_{i=1}^{n} x_i^2 - n x_A^2}\cdot\frac{\sum_{i=1}^{n} x_i y_i - n x_A y_A}{\sum_{i=1}^{n} y_i^2 - n y_A^2} = (r)^2,$$

čime smo dokazali da je: $\text{koeficijent determinacije} = (\text{koeficijent linearne korelacije})^2$.

Napomenimo da svi navedeni pokazatelji mjere stupanj linearne povezanosti dviju pojava. Budući da se ona u praksi najviše pojavljuje, u empirijskim istraživanjima pretpostavka o linearnoj povezanosti najčešće se i koristi. Naime, većina nelinearnih oblika povezanosti mogu se transformirati ili vrlo dobro aproksimirati linearnim oblikom koji je najjednostavnije analizirati, interpretirati i primijeniti. Navodimo neke tipične primjere linearizacije.

Logaritamska ovisnost $y = a\log x + b$ zamjenom $\tilde{x} = \log x$, prelazi u linearnu $y = a\tilde{x} + b$.

Eksponencijalna ovisnost $y = b\cdot a^x$ logaritmiranjem prelazi u oblik $\log y = \log b + x\log a$, koji zamjenom $\tilde{y} = \log y$, $\tilde{b} = \log b$, $\tilde{a} = \log a$ postaje linearna ovisnost $\tilde{y} = \tilde{a}x + \tilde{b}$.

Nelinearna ovisnost $y = bx^a$ također logaritmiranjem postaje $\log y = \log b + a\log x$, pa zamjenom $\tilde{y} = \log y$, $\tilde{b} = \log b$, $\tilde{x} = \log x$ dobivamo linearnu $\tilde{y} = a\tilde{x} + \tilde{b}$.

Razlomljena ovisnost $y = c/(ax + b)$ zamjenom $\tilde{y} = 1/y$, $\tilde{a} = a/c$, $\tilde{b} = b/c$ također postaje linearna $\tilde{y} = \tilde{a}x + \tilde{b}$, itd.

ZADACI

1. Kakvi podaci će imati $r^2 = 1$ odnosno $r^2 = 0$?
2. Koji pravac najbolje reprezentira točke (0,0), (0,1), (1,0), (1,1), (2,0), (2,1), (3,0), (3,1), (4,0), (4,1)? Nacrtajte sliku.
3. Funkciju $f(x) = x^2$ aproksimirajte linearnom funkcijom na skupu točaka $\{1,2,3,4,5\}$, te odredite varijancu, standardnu devijaciju i koeficijent varijacije te aproksimacije.
4. Funkciju $f(x) = \sqrt{x}$ aproksimirajte linearnom funkcijom na skupu točaka $\{1,4,9,16\}$, te odredite koeficijent determinacije takve aproksimacije.
5. U jednom gradu imamo broj rođenih i umrlih tokom zadnjih pet godina.

Godina	***2011***	***2012***	***2013***	***2014***	***2015***
Rođeni	220	205	180	190	170
Umrli	300	320	305	335	330

   Odredite povezanost rođenih i umrlih linearnom regresijskom jednadžbom. Ako broj rođenih padne ispod 150 godišnje, koliko možemo očekivati umrlih? Što možemo reći o dobivenoj procjeni na temelju koeficijenta determinacije modela?
6. Ako je u jednom gradu, zbog permanentnog ulaganja u turizam, broj turista u posljednje tri godine bio 22400, 24800 i 27800, koliko turista možemo očekivati u ovoj i slijedećoj godini? Kolika je pouzdanost (koeficijent determinacije) procjene?
7. Na mekom tržištu su poznati slijedeći podaci za potražnju proizvoda X i ponudu istog proizvoda od jednog dobavljača u ovisnosti o njegovoj cijeni.

Cijena (kn)	**46**	**48**	**50**	**52**	**54**	**56**	**60**	**64**
Potražnja	320	–	300	–	285	–	271	–
Ponuda	–	250	–	265	–	289	–	320

   (a) Odredite analitički oblik linearne funkcije ponude i potražnje.
   (b) Odredite tržišnu cijenu (to je cijena za koju su ponuda i potražnja u ravnoteži).
   (c) Za koju cijenu možemo očekivati višak od 50 proizvoda na tom tržištu?
   (d) Za koju cijenu možemo očekivati manjak od 50 proizvoda na tom tržištu?
8. Evidentiran je broj korisnika dvaju bankomata B1 i B2 jedne poslovne banke u dvije gradske četvrti tokom tjedna.

Dan	***PON***	***UTO***	***SRI***	***ČET***	***PET***	***SUB***	***NED***
B1	146	120	133	95	205	181	30
B2	55	62	40	75	52	13	32

   Prikažite podatke dijagramom rasipanja. Pomoću koeficijenta determinacije odredite postojanje značajnije povezanosti korištenja bankomata? Ako značajnija veza postoji, odredite regresijsku jednadžbu.
9. Brod za ispitivanje morskih struja izbacio je u more 5 plutača različitih težina na mjestu označenom kao ishodište sa koordinatnim osima usmjerenim prema glavnim

stranama svijeta (pozitivna strana osi x prema istoku a osi y prema sjeveru). Plutače periodično šalju signale sa podacima o svom položaju. Nakon 24 sata koordinate plutača (u km) su bile: $P_1(4,7.8)$, $P_2(6.2,9)$, $P_3(1.5,7)$, $P_4(5.3,8.5)$ i $P_5(8,9.7)$. Koji je smjer (kut) morskih struja u odnosu na strane svijeta? Da li je u promatranom periodu bilo transverzalnog pomicanja struja, za koliko i u kojem smjeru?

10. Potrebno je napraviti projekt glavnog pravolinijskog plinovoda koji će opskrbljivati naselja A, B, C i D. Zemljopisni položaji priključnih stanica u navedenim naseljima u zadanom pravokutnom koordinatnom sustavu su: A(20, 30), B(30,10), C(40,60) i D(50,80), gdje su koordinate izražene u kilometrima. Odredite trasu plinovoda tako da ukupna duljina priključnih vodova za sva četiri naselja bude minimalna. Odredite duljinu priključnih vodova za svako naselje ako se oni spajaju okomito na glavni vod. Nacrtajte skicu plinovoda, naselja i priključnih vodova.
11. Brzi vlak B700 tokom 2016. godine svakodnevno prometuje na relaciji Zagreb – Rijeka prema slijedećem redu vožnje

Stanica	***Zagreb***	***Karlovac***	***Ogulin***	***Moravice***	***Rijeka***
Vrijeme	6.30	7.09	8.06	8.37	10.08
Udaljenost od Zagreba (km)	0	53	109	139	229

Aproksimirajte ovu putnu relaciju linearnim modelom (jednolikim gibanjem) te odredite koeficijent determinacije. Interpretirajte rezultate. Modelom procijenite gdje se vlak nalazi u 9 sati i 30 minuta te usporedite sa voznim redom (Vozni red HŽ putnički prijevoz za razdoblje 13.12.2015. – 10.12.2016.).

12. U jedan proizvodni pogon dostavljane su slijedeće količine (u kg) sirovina x i y: (750, 3050), (900, 3780), (1000, 4000) i (630, 2550). Odredite povezanost količina ovih sirovina linearnim regresijskim modelom. Odredite ukupna (*ST*), protumačena (*SP*) i ne protumačena (*SR*) odstupanja te koeficijent determinacije modela.
13. U laboratorijskim uvjetima provodi ¨se eksperiment uzgoja jedne vrste bakterija te se prati dinamika njihovog razmnožavanja,

Vrijeme (sati)	1	2	3	4
Broj bakterija	100	970	8500	112000

Aproksimirajte navedenu dinamiku eksponencijalnim trendom, $y = b \cdot a^x$. Kada treba završiti eksperiment ako je cilj uzgojiti koloniju od 50 milijardi bakterija,?

14. Veće grupa pacijenata sa značajnije povišenom tjelesnom temperaturom (preko 39°C) primila je različite doze iste vrste lijeka. Nakon jednog sata ustanovljen je slijedeći prosječni pad temperature.

Doza lijeka (mg)	100	200	400	800
Pad temperature (°C)	0.5	0.7	1	1.4

(a) Aproksimirajte pad temperature regresijskom jednadžbom $y = b \cdot x^a$.

(b) Ako neki pacijent sa temperaturom 40°C zbog preosjetljivosti smije primiti najviše 50 mg lijeka na sat, za koje vrijeme može očekivati normalizaciju tjelesne temperature (na 37°C)?

(c) Koliku dozu lijeka bi trebalo dati pacijentu sa temperaturom 39.5°C pa da mu za jedan sat padne na 37.5°C ?

15. Iz zadanog skupa brojeva svaki od vas neka nasumce izabere i zapiše nekoliko brojeva, na primjer iz skupa brojeva od 1 do 30 izaberite ih 8, te zatim u parovima nacrtajte dijagram rasipanja i odredite pojedine pokazatelje povezanosti tako zapisanih brojeva. Pri tome brojeve usporedite: (a) po redoslijedu zapisivanja, (b) po veličini. Što primjećujete?

16. Proizvoljno zadajte jednadžbu pravca $y = ax + b$. Na pravcu uzmite 10 točaka sa apscisama 1,2,…,10. Time ste dobili točke $T_1(1, y_1)$, $T_2(2, y_2)$, ..., $T_{10}(10, y_{10})$, gdje je $y_i = a \cdot i + b$, $i = 1, 2, ..., 10$. Ordinatu točke T_1 povećajte za 1. Sada umjesto T_1 imate točku $P_1(1, y_1 + 1)$. Rasporedite se međusobno u grupe. Prva grupa neka odredi linearnu regresijsku jednadžbu za P_1, T_2, T_3, druga za P_1, T_2, T_3, T_4, treća za P_1, T_2, T_3, T_4, T_5, itd. Tako dobijete 8 regresijskih jednadžbi. Ispišite koeficijente a i b iz dobivenih jednadžbi i dijagramom dvostrukih stupaca prikažite njihovu razliku u odnosu na koeficijente polazne jednadžbe. Umjesto 10 točaka, zadatak možete rješavati sa bilo kojim brojem točaka (m) sa apscisama $1, 2, ..., m$ ili nekim drugim vrijednostima.

RJEŠENJA

1. Za $r^2 = 1 \Rightarrow SR = 0 \Rightarrow y_i = y(x_i),\ i = 1, 2, ..., n$, što znači da svi empirijski podaci pripadaju istom pravcu. Za $r^2 = 0 \Rightarrow SP = 0 \Rightarrow y(x_i) = y_A,\ i = 1, 2, ..., n$, što znači da je regresijski pravac konstanta, $y(x) = ax + b = y_A \Rightarrow a = 0$ odnosno (relacija (20)),

$$\sum_{i=1}^{n} x_i y_i = n x_A y_A \;\Rightarrow\; \sum_{i=1}^{n} x_i y_i = n \cdot \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) \cdot \frac{1}{n} \sum_{i=1}^{n} y_i \;\Rightarrow\; \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right) = n \left(\sum_{i=1}^{n} x_i y_i \right),$$

ili $\sum_{i=1}^{n} (y_i - y_A)(x_i - x_A) = 0$ što znači da su podaci simetrično raspoređeni oko $x = x_A$ i $y = y_A$ (imaju dvije osi simetrije).

2. Imamo $x_A = 2,\ y_A = 0.5$ pa je $\sum_{i=1}^{n} x_i y_i - n x_A y_A = 10 - 10 \cdot 2 \cdot 0.5 = 0$, odnosno $a = 0$ (vidjeti zadatak 1). Točke su simetrično raspoređene oko osi $x = 2$ i $y = 0.5$, pa je regresijski pravac konstanta, $y = 0.5$. Treba nacrtati točke i pravce (osi simetrije).

3. Imamo niz (1,1), (2,4), (3,9), (4,16), (5,25) pa je $x_A = 3$, $y_A = 11$. Iz (20) slijedi

$$a = \frac{225 - 5 \cdot 3 \cdot 11}{55 - 5 \cdot 3^2} = 6, \quad b = 11 - 6 \cdot 3 = -7 \quad \Rightarrow \quad y = 6x - 7,$$

te iz definicije 11, $\sigma_y^2 = \frac{1}{5}(979 - 6 \cdot 225 + 7 \cdot 55) = 2.8$, $\sigma_y = 1.67332...$, $V_y = 0.15212...$

4. Imamo niz (1,1), (4,2), (9,3), (16,4) pa je $x_A = 7.5$, $y_A = 2.5$. Iz (20) i (21) sada slijedi

$$y = ax + b = \frac{25}{129}x + \frac{45}{43} \approx 0.1938x + 1.0465, \qquad r^2 = \frac{125}{129} \approx 0.969.$$

5. Ako su x rođeni a y umrli, imamo $x_A = 193$, $y_A = 318$, iz čega slijedi

$$y = -\frac{31}{79}x + \frac{31105}{79} \approx -0.3924x + 393.734 \quad \Rightarrow \quad (x \le 150 \Rightarrow y \ge 335).$$

   Kako je $r^2 = 62/237 \approx 0.2616$, dobiveni linearni model slabo reprezentira povezanost pojava pa procjena nije pouzdana.

6. Numerirajmo protekle tri godine sa $x = -3, -2, -1$. Imamo $x_A = -2$, $y_A = 25000$, pa je linearna regresijska jednadžba $y = 2700x + 30400$. Tražene procjene su $y(0) = 30400$ i $y(1) = 33100$ a njihova pouzdanost $r^2 = 243/244 \approx 0.996$ je iznimno visoka.

7. Označimo li $p-$ cijena, $q_1 -$ potražnja i $q_2 -$ ponuda, imamo:

   (a) Potražnja je $q_1(p) = -\frac{370}{107}p + \frac{50883}{107}$ ili približno $q_1(p) \approx -3.458p + 475.542$, dok je ponuda $q_2(p) = \frac{156}{35}p + \frac{251}{7}$ ili približno $q_2(p) \approx 4.457p + 35.857$.

   (b) Iz $q_1 = q_2$ slijedi $p = 55.55$ kuna.

   (c) Iz $q_2 = q_1 + 50$ slijedi $p = 61.87$ kuna.

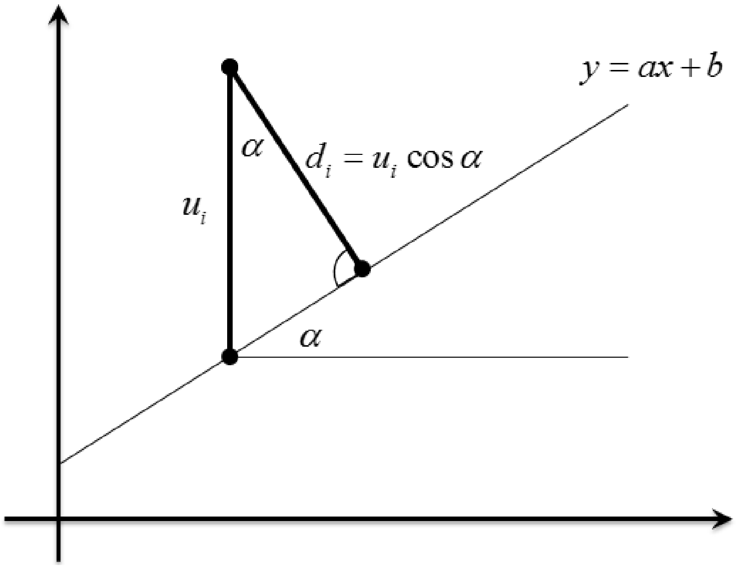
   (d) Iz $q_1 = q_2 + 50$ slijedi $p = 49.23$ kune.

8. Imamo $x_A = 130$ za B1 i $y_A = 47$ za B2, pa je $r^2 = 0.015529...$ što znači da praktično ne postoji nikakva korelacija broja korisnika ta dva bankomata. To se može vidjeti i iz dijagrama rasipanja.

9. Treba odrediti parametre linearne regresijske jednadžbe za zadane točke. Koeficijent smjera a određuje kut a koeficijent b transverzalni (paralelni) pomak. Dobijemo $a = 35/82 = 0.426829...$ i $b = 2569/410 = 6.26585...$. Kako je $a = \tan\alpha$, smjer struje je $\alpha \approx 23^o 7'$ prema sjeveroistoku a pomak oko 6.3 km prema sjeveru.

10. Iz $x_A = 35$, $y_A = 45 \;\Rightarrow\; a = (7300 - 6300)/(5400 - 4 \cdot 35^2) = 2$, $\quad b = 45 - 2 \cdot 35 = -25$, pa je trasa plinovoda $y = 2x - 25$. Rezidualna odstupanja za A, B, C i D su redom: $u_1 = 30 - y(20) = 15$, $u_2 = 10 - y(30) = -25$, $u_3 = 60 - y(40) = 5$, $u_4 = 80 - y(50) = 5$. Kako su priključni vodovi okomiti na glavni vod, računamo duljinu okomica d_i, $i = 1, 2, 3, 4$ (vidjeti sliku). Imamo $d_i = u_i \cos\alpha$ a kako je $a = \tan\alpha = 2$ (koeficijent smjera pravca) slijedi $\cos\alpha = 1/\sqrt{1 + \tan^2\alpha} = 1/\sqrt{1 + a^2} = 1/\sqrt{5}$. Dakle, $d_1 = 6.708$ km, $d_2 = -11.180$ km, $d_3 = d_4 = 2.236$ km (suprotni predznaci ukazuju da su vodovi

sa suprotnih strana plinovoda). Ukupna duljina priključnih vodova je zbroj njihovih apsolutnih vrijednosti, 22.36 km. Treba nacrtati kompletnu sliku.



11. Vrijeme (x_i) izražavamo u decimalnom obliku (zaokruženo na tri decimale): 6.5, 7.15, 8.1, 8.617, 10.133. Imamo: $x_A = 8.1$, $y_A = 106$,

$$a = \frac{4780.07 - 4293}{335.912878 - 328.05} = 61.94551..., \quad b = 106 - a \cdot 8.1 = -395.75864...,$$

pa linearni model glasi $y = 61.946x - 395.759$. Parametar a pokazuje prosječnu brzinu (km/h) vlaka. Nadalje, $r^2 = 30172.2762 / 30272 = 0.9967...$ što pokazuje da je putovanje na ovoj relaciji izuzetno dobro aproksimirano jednolikim gibanjem uz navedenu prosječnu brzinu. U 9 sati i 30 minuta (9.5 sati) je $y(9.5) = 192.728$, pa je vlak približno na 193. kilometru od Zagreba (36 km od Rijeke) a što se točno poklapa sa voznim redom (stanica Drivenik).

12. Dobijemo $x_A = 820$, $y_A = 3345$, pa je $y = 4.0651629x + 11.566416$. Odstupanja su $ST = 1337300$, $SP = 1318738.847$ i $SR = 18561.153$, pa je $r^2 = SP / ST = 0.98612$ što pokazuje visoku kvalitetu (reprezentativnost) dobivenog modela.

13. Tu je x vrijeme a y broj bakterija. Logaritmiramo, $\log y = \log b + x \log a$. Stavimo li $\tilde{y} = \log y$, $\tilde{b} = \log b$ i $\tilde{a} = \log a$, imamo linearni trend, $\tilde{y} = \tilde{a}x + \tilde{b}$ za kojeg su podaci

x_i	1	2	3	4
$\tilde{y}_i = \log y_i$	2	2.9868	3.9294	5.0492

zaokruženo na četiri decimale. Imamo: $x_A = 2.5$ i $\tilde{y}_A = 3.49135$, iz čega dobivamo $\tilde{a} = 5.0451 / 5 = 1.00902$, $\tilde{b} = 0.9688$. Sada je $a = 10^{\tilde{a}} = 10.2099$ i $b = 10^{\tilde{b}} = 9.3068$, što daje približan eksponencijalni trend razmnožavanja bakterija, $y = 9.3 \cdot 10.2^x$. Za $y = 50\,\text{milijardi} = 50 \cdot 10^9 = 5 \cdot 10^{10}$ je $\tilde{y} = \log 5 + 10 = 10.69897$, pa iz linearnog modela slijedi $x = (\tilde{y} - \tilde{b}) / \tilde{a} = 9.6432 \approx 9$ sati i 39 minuta od početka eksperimenta.

14. Tu je x doza lijeka a y pad temperature. Linearizacijom dobivamo $\tilde{y} = a\tilde{x} + \tilde{b}$, gdje je $\tilde{y} = \log y$, $\tilde{b} = \log b$, $\tilde{x} = \log x$, pa su podaci zaokruženi na tri decimale

$\tilde{x}_i = \log x_i$	2	2.301	2.602	2.903
$\tilde{y}_i = \log y_i$	–0.301	–0.155	0	0.146

(a) Iz $\tilde{x}_A = 2.4515$ i $\tilde{y}_A = -0.0775$ je $a = 0.4970099...$, $\tilde{b} = -1.2959199...$, odnosno $b = 10^{\tilde{b}} = 0.05059179...$, što daje traženu aproksimaciju $y = 0.05 \cdot x^{0.497}$.

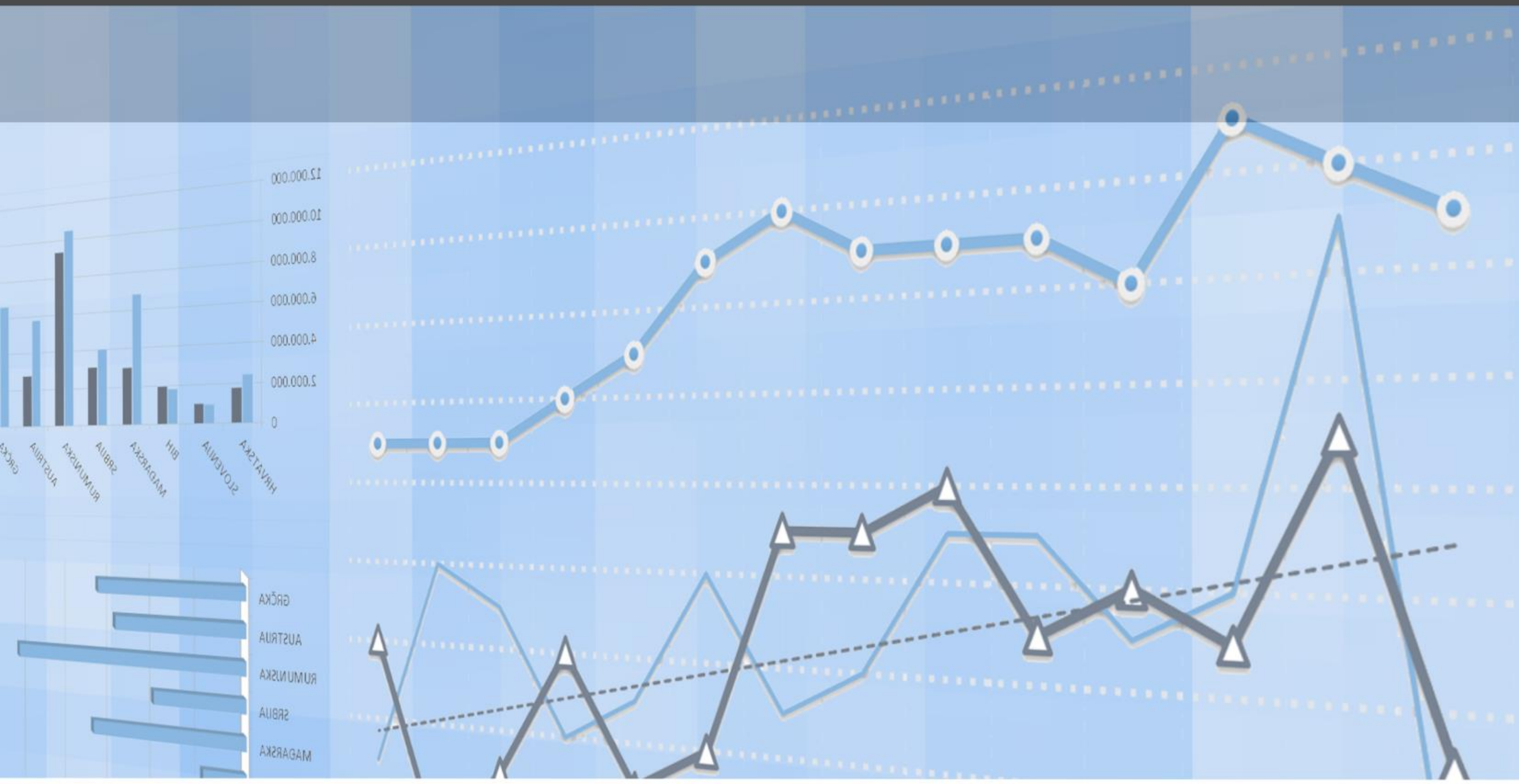
(b) Imamo pad temperature $y(50) = 0.05 \cdot 50^{0.497} = 0.349428... \approx 0.35\,^{\circ}$C po satu, pa normalizaciju možemo očekivati za $(40-37):0.35 = 8.5714...$, dakle, unutar 9 sati.

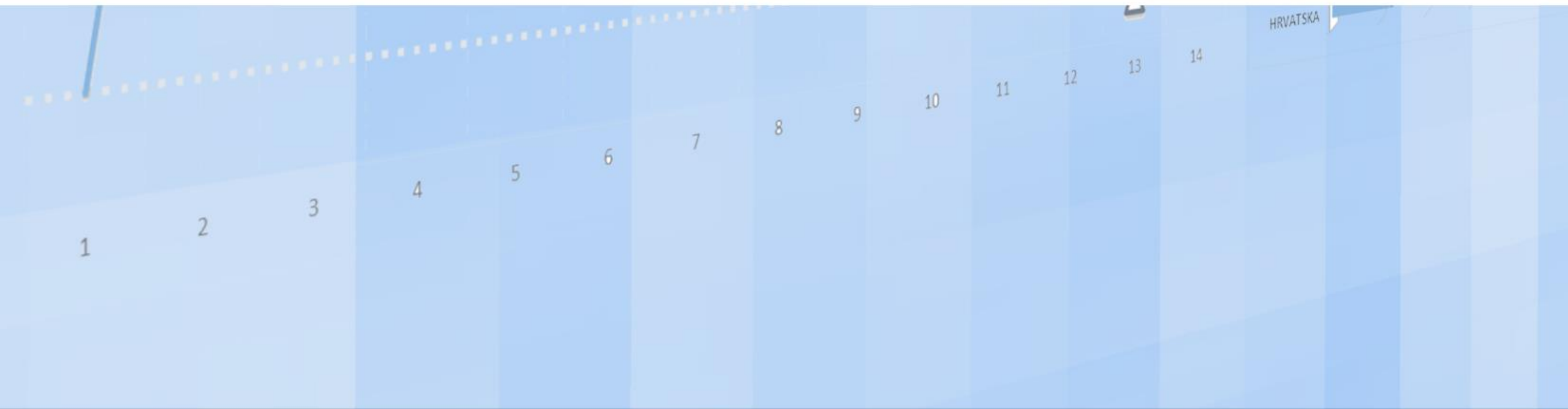
(c) Temperaturu treba smanjiti za 2°C tokom jednog sata. Iz jednadžbe $2 = 0.05 \cdot x^{0.497}$ slijedi $x^{0.497} = 40$, odnosno $x = 1672.86443... \approx 1673$ mg.

15. Koristite stvarne podatke. Povezanost je znatno veća ako uspoređujemo zapisane brojeve po veličini.
16. Koristite izabranu jednadžbu pravca. Ako su razlike koeficijenata za grafički prikaz vrlo male, prilagodite (rastegnite) jediničnu dužinu na ordinati.

LITERATURA

1. Anderson D. R., Sweeney D. J., Williams T. A., Freeman J., Shoesmith E., *Statistics for Business and Economics*, Thomson, London 2007.
2. Bahovec V., Dumičić K., Erjavec N., Čižmešija M., Kurnoga B., …, *Statistika*, Element d.o.o., Zagreb 2014.
3. Matejaš J., *Nastavni materijali (predavanja i vježbe) iz različitih kolegija*, Ekonomski fakultet, Zagreb 2000 – 2015.
4. Mohammed A. S., *Applied Statistics*, ISBN: 978-87-403-1210-2, bookboon.com 2013.
5. Singpurwalla D., *A Handbook of Statistics – An Overview of Statistical Methods*, ISBN: 978-87-403-0542-5, bookboon.com 2013.
6. Šošić I., *Statistika*, Školska knjiga, Zagreb 1997.
7. Tomašević N., *Statistika*, Štamparija nauka i društvo, Beograd 2012.




HRVATSKA
1
2
3
4
5
6
7
8
9
10
11
12
13
14