SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Anja Stanković

MATEMATIČKA STATISTIKA U NASTAVI MATEMATIKE

Diplomski rad

Voditelj rada:
doc. dr. sc. Matija Bašić

Zagreb, rujan 2021.

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Mojim roditeljima, koji su me odmalena učili da obrazovanje svijet može učiniti boljim.
Mojim roditeljima i bratu - hvala što ste unatoč našim
borbama prihvatili da odaberem svoj put.

Mojem dečku, hvala što si me toplo i hrabro držao za ruku u doba promjena.
Mojim bivšim dečkima hvala na bezuvjetnoj podršci i vjeri.

Mentoru i Profesoru Matiji zahvaljujem na slobodi, fleksibilnosti i stručnim savjetima za vrijeme pisanja rada. Hvala što ste me usmjerili da razmišljam izvan okvira!
Prof. dr. sc. Željki Milin Šipuš zahvaljujem na stručnom znanju, izgradnji i jačanju mog nastavničkog i profesionalnog identiteta.

Hvala svim dobrim ljudima, mojim učenicima i prijateljima
koji su me potaknuli da otkrijem zašto.

„Napretkom i usavršavanjem matematike uvjetovano je blagostanje države.“
N. Bonaparte

Sadržaj

Uvod

"Statistički način mišljenja jednog će dana za svakodnevni život građana postati jednako neophodan kao znanje čitanja i pisanja", tvrdio je početkom 20. stoljeća engleski pisac H. G. Wells. Danas živimo u vremenu digitalne revolucije i osnovno znanje matematičke statistike potrebno je za zanimanja 21. stoljeća, uključujući STEM područje. Novom kurikularnom reformom u školske kurikulume uvodi se veći broj sati statistike da bi učenici postali kompetentni za tržište rada i zanimanja budućnosti. U prvom poglavlju opisano je, uz primjere, što se iz statistike uči u osnovnoj i srednjoj školi prema novom kurikulumu, a u drugom poglavlju prikazuje se kako opisivati, uređivati i prikazivati podatke koristeći jednostavne formule i račune. Teorijska pozadina o slučajnoj varijabli i razdiobama dio su 3. poglavlja, a posebno je detaljnije obrađena normalna razdioba. Glavna motivacija bila je dati cjelovit pregled novih dijelova kurikuluma sa svrhovitom primjenom u svakodnevnom životu.

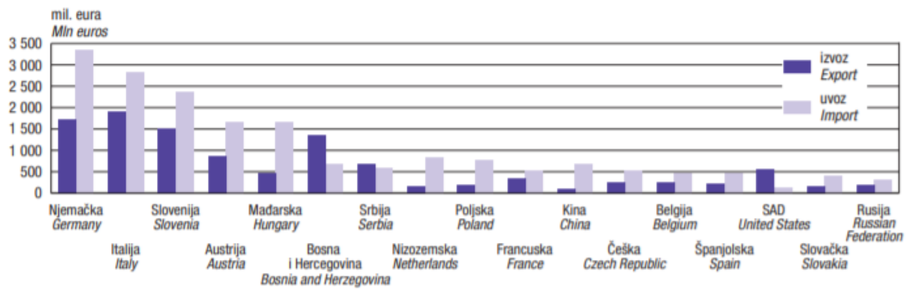
Poglavlje 1

Statistika u kurikulumu

1.1 Potreba za znanjem statistike

U svakodnevnom životu susrećemo se s matematičkom statistikom u različitim područjima poput medicine, IT-a, podatkovne znanosti, fizike, ekonomije i politike. Promotrimo nekoliko primjera koje možemo saznati na službenim mrežnim stranicama Državnog zavoda za statistiku Republike Hrvatske [**14**] (www.dzs.hr, 24. svibnja 2021.):

- Prva procjena pokazuje da je tromjesečni BDP u četvrtom tromjesečju 2020. realno manji za 7,0% nego u istom tromjesečju 2019. To je nešto blaži pad nego što je bio u trećem tromjesečju, kada je iznosio 10,0%.
- Pad u prijevozu robe na unutarnjim vodnim putovima imali su brodovi svih zastava, za 23,8%, kao i željeznički prijevoznici, za 0,3%.
- Pokrivenost uvoza izvozom od siječnja do travnja 2021. iznosila je 64,7%, dok je u istom razdoblju 2020. iznosila 61,3%.



Slika 1.1: Najvažnije zemlje partneri u izvozu i uvozu Republike Hrvatske u 2017. [**14**]

Kao što smo vidjeli na Slici (1.1) statistika nam služi za analizu podataka koje želimo prikazati. Grafičke prikaze možemo čitati na različite načine i interpretirati ih. Uporaba riječi statistika u svakodnevnom životu najčešće je povezana s brojčanim vrijednostima kojima pokušavamo opisati bitne karakteristike nekog skupa podataka [**7**].

Definicija 1.1.1 ([**7**]). *Statistika* kao znanstvena disciplina bavi se razvojem metoda prikupljanja, opisivanja i analiziranja podataka te primjenom tih metoda u procesu donošenja zaključaka na temelju prikupljenih podataka.

Za zanimanja 21. stoljeća, uključujući STEM područje, znanje matematičke statistike jedno je od temeljnih znanja. Prema Craigu ([**11**]) STEM označava znanost, tehnologiju, inženjerstvo i matematiku i odnosi se na sve predmete koji potpadaju pod ove četiri discipline. Postoji i na desetke drugih verzija STEM-a, uključujući STEAM, ali STEM se daleko najviše koristi. U STEM područje svrstavaju se sljedeće djelatnosti: zrakoplovno inženjerstvo, astronomija, biokemija, biologija, kemijsko inženjerstvo, kemija, energetika, informatika, elektrotehnika, matematika, strojarstvo, fizika, psihologija i statistika. Za izgradnju zrakoplova poput Rolls Royce-a, rad na posebnim efektima u Hollywoodu, dizajn nove sportske odjeće ili sudjelovanje u revoluciji u poljoprivrednom sektoru osnovno znanje statistike je neophodan alat za rad. Kako smo i sami u posljednje vrijeme svjesni utjecaja globalnog zatopljenja i klimatskih promjena neke od budućih atraktivnih STEM karijera bit će i iz područja klimatologije.

Istaknuli smo važnost STEM-a u obrazovanju iz čega možemo zaključiti da su ekonomski rast i odgojno-obrazovni sustav međusobno povezani i ključni za blagostanje i bogatsvo države. Matematička statistika sastavni je dio najboljih i najkvalitetnijih kurikuluma pojedinih zemalja koje ne samo da imaju izvrsne rezultate u testiranjima učenika nego imaju i izrazito visok bruto društveni proizvod (BDP). U sljedećem primjeru koristeći statistiku možemo vidjeti koliko ulaganje u STEM može utjecati na budućnost razvoja zemlje i BDP. Pritom uviđamo da interpretacija statističkih pojmova u svakodnevnom životu, kao i argumentacija i prezentiranje neke ideje, čine statistiku zanimljivu ljudima.

Primjer 1. *Zanimljiva STEM statistika iz koje čitatelj lako može uočiti potrebu za znanjem statistike* [**4**].

- Ministarstvo obrazovanja Sjedinjenih Američkih Država nedavno je uložilo 540 milijuna dolara u STEM obrazovanje. Bijela kuća identificirala je potrebu za jačanjem podrške STEM obrazovanju. Ta sredstva uključuju napore za zapošljavanje i osposobljavanje kvalitetnih nastavnika STEM-a, povećanje raznolike zastupljenosti te pružanje nastavnih planova i materijala za škole.

- Predviđa se da će zapošljavanje na STEM radnim mjestima rasti 8,8% u razdoblju od 2017. do 2029. godine.

- U tom razdoblju predviđa se da će i zaposlenost u razvoju softwera porasti 22%.
- Zaposlenost u STEM zanimanjima porasla je 79% od 1990. godine.
- Prosječna satnica za STEM poslove iznosi 38,85 USD.
- Srednja godišnja plaća za STEM zanimanja u 2020. godini iznosila je 89 780 USD.
- 74% djevojčica u srednjoj školi izražava interes za inženjerstvo, znanost i matematiku.
- Od 2019. žene čine samo 27% STEM radne snage.
- 63% djevojčica u srednjoj školi, koje poznaju žene u STEM-u, osjećaju se moćno radeći STEM.
- Djevojke, s kojima su mame pozitivno komunicirale o STEM-u, zainteresiranije su 20-ak bodova više od dječaka za studiranje u STEM području.

Kao što smo vidjeli, statistika je znanstvena grana unutar matematike i iz navedenih razloga javlja se potreba za podučavanjem statistike u nastavi matematike.

1.2 Statistika u obrazovanju

Statistika se kao dio matematike pojavljuje na raznim testovima vanjskog vrednovanja. U 21. stoljeću kvaliteta matematičkog obrazovanje neke zemlje uspoređuje se na osnovi rezultata učenika u testiranjima PISA (Programme for International Student Assessment) i TIMSS (Trends in International Mathematics and ScienceStudy). PISA više od trećinu svojih zadataka iz područja matematičke pismenosti pridaje području podataka, statistike i vjerojatnosti [**15**].

Najbolje rezultate na PISA i TIMMS testiranjima postižu učenici iz Singapura, Finske, Hong Konga, Koreje, Japana,... Bitno je naglasiti da se statistika uz vjerojatnost u tim državama podučava od ranije dobi. Istaknut ćemo kako je statistika uz inovativni pristup podučavanju dio pojedinih uspješnih odgojno-obrazovnih sustava.

Ključni dio singapurskog kurikuluma je rješavanje matematičkih problema, a ovisi o pet međusobno povezanih komponenata: stavovi, metakognicija, procesi, pojmovi i vještine. U singapurskom kurikulumu izražena je primjerenost problema prilikom rješavanja kao i učenikovo preuzimanje odgovornosti za rješavanje i opravdanost odgovora. Naglasak je stavljen na psihološke aspekte rješavanja. Neki od koncepata koji se protežu kroz singapurski kurikulum jesu vjerojatnost, statistika i analiza, a jedna od vještina koju učenici trebaju razviti je analiza podataka.

U Finskoj se procjena postignuća učenika može načelno podijeliti u tri kategorije. Prvu čini procjena učenikovih postignuća u razredu, koju provodi nastavnik. Ona obuhvaća dijagnostičke, formativne i sumativne procjene koje su dio učenja i poučavanja. Ova vrsta procjene u svim je školama isključiva odgovornost nastavnika. Drugu skupinu procjena postignuća čini opće vrednovanje napretka učenika nakon polugodišta. Treće, u Finskoj postoje i vanjske procjene postignuća učenika. Svake tri ili četiri godine procjenjuje se postignuće učenika u čitanju, matematici, prirodoslovlju i drugim predmetima. Glavno načelo politike općeg obrazovanja u Finskoj bilo je stvaranje jednakih obrazovnih mogućnosti za sve građane [**28**]. Nova finska škola treba biti socijalno, nadahnjujuće i sigurno okruženje u kojem će svi učenici steći socijalne vještine potrebne za život.

U Hong Kongu se početna matematika uči u sklopu općih vještina primjenjivih u različitim životnim situacijama. Kao važna sadržajna domena navodi se Upravljanje podatcima (Data handling) u koju je uključena statistika, a od 6. razreda i vjerojatnost [**22**].

Vjerojatnost i statistika usko su povezani te je znanje vjerojatnosti preduvjet za razumijevanje statistike tako da se u obrazovni sustav uvode zajedno. U Republici Hrvatskoj vjerojatnost i statistika uvode se u prvi put Nastavni plan i program 2006. godine. U sedmom razredu osnovne škole učenici su učili Prikazivanje i analizu podataka i Vjerojatnost slučajnog događaja. Uvođenje vjerojatnosti i statistike u nastavu matematike nastavilo se kao dio Nacionalnog okvirnog kurikuluma 2011. godine. U sklopu matematičkog koncepta Podatci, statistika i vjerojatnost uvrštene su kao teme u sve cikluse obrazovanja u većem opsegu 2011. godine, nego 2006. godine.

Republika Hrvatska u procesu vanjskog vrednovanja od školske godine 2009./2010. godine provodi ispite državne mature na kraju završnih razreda srednje škole. Matematika je dio osnovne skupine predmeta te ju je moguće polagati na višoj ili nižoj razini. Na nižoj, odnosno osnovnoj razini, učenici rješavaju zadatke višestrukog izbora i kratkih odgovora. Treba istaknuti da do školske godine 2021./2022. statistika nije bila zasebno područje ispitivanja, već se znanje statistike ispitivalo kroz područje Brojevi i algebra. Na prvoj državnoj maturi u školskoj godini 2009./2010. područje Brojevi i algebra čini udio ispita od 45% na nižoj, a 20% na višoj razini. U školskoj godini 2020./2021. područje Brojevi i algebra obuhvaća 42,5% ispita na nižoj, a 20% ispita na višoj razini [**20**].

Kako bi poboljšala cjelokupno obrazovanje, kao i rezultate u testiranjima učenika poput PISA-e i TIMSS-a i time ostvarila gospodarski rast, Republika Hrvatska odlučila se na Cjelovitu kurikularnu reformu. Analizirali su se kurikulumi različitih zemalja i samim time su se donosile odluke o implementaciji različitih domena u odgojno-obrazovni proces. Neke od navedenih koncepata i ideja Hrvatska je željela ostvariti provođenjem Cjelovite kurikularne reforme. U sklopu Cjelovite kurikularne reforme (CKR) napravljen je Nacionalni kurikulum nastavnog predmeta matematike. Odluke su objavljene 2019. godine u Narodnim novinama pod nazivom Kurikulum nastavnog predmeta Matematika za osnovne škole i gimnazije i pod nazivom Kurikulum za nastavni predmet Matematika za srednje strukovne škole na razini 4.2. Cjelokupni dokument objavljen je 2020.godine pod nazivom Kurikulumi nastavnih predmeta Matematika za osnovne škole i gimnazije i Matematika za srednje strukovne škole na razini 4.2., a objavljeni dokumenti nisu obuhvatili Cjelovitu kurikularnu reformu u potpunosti.

U odgojno-obrazovnom sustavu Republike Hrvatske posljednje tri godine vrijeme je procesa izmjene kurikuluma. Odluka iz siječnja 2019. godine primjenjuje se za učenike 1. i 5. razreda osnovne škole i 1. razreda srednjih škola od školske godine 2019./2020., za učenike 2., 3., 6. i 7. razreda osnovne škole, 2. i 3. razreda srednjih škola od školske godine 2020./2021., a za učenike 4. i 8. razreda osnovne škole i 4. razreda srednjih škola od školske godine 2021./2022.

Jedna od domena novog kurikuluma je Podatci, statistika i vjerojatnost, a od školske godine 2021./2022. navedena domena bit će jedno od područja ispitivanja na državnoj maturi. Na nižoj razini državne mature udio sadržaja domene Podatci, statistika i vjerojatnost nosit će 10%, a na višoj razini državne mature nosit će 5%.

1.3 Analiza ishoda uz primjere po razredima - osnovna škola

Prema Kurikulumu nastavnog predmeta Matematika za osnovne škole, gimnazije i strukovne škole učenje i poučavanje nastavnoga predmeta Matematika [**24**] ostvaruje se povezivanjem matematičkih procesa i domena. Ta dvodimenzionalnost očituje se u ishodima i doprinosi stjecanju matematičkih kompetencija. Matematički su procesi: prikazivanje i komunikacija, povezivanje, logičko mišljenje, argumentiranje i zaključivanje, rješavanje problema i matematičko modeliranje te primjena tehnologije. Domene predmeta Matematika jesu: Brojevi, Algebra i funkcije, Oblik i prostor, Mjerenje te Podatci, statistika i vjerojatnost.

Domena Podatci, statistika i vjerojatnost bavi se prikupljanjem, razvrstavanjem, obradom, analizom i prikazivanjem podataka u odgovarajućemu obliku. Podatke dane grafičkim ili nekim drugim prikazom treba znati očitati te ih ispravno protumačiti i upotrijebiti. Sve se to postiže koristeći se jezikom statistike. Ona podrazumijeva uporabu matematičkoga aparata kojim se računaju mjere srednje vrijednosti, mjere raspršenja, mjere položaja i korelacije podataka. Nakon prepoznavanja veza među podatcima i promatrajući frekvencije pojavljivanja, dolazi se do pojma vjerojatnosti. Određuje se broj povoljnih i svih mogućih ishoda, procjenjuje se i izračunava vjerojatnost što nam omogućuje predviđanje događaja [**23**].

U ovoj cjelini analizirat ćemo domenu Podatci, statistika i vjerojatnost tako da ćemo izdvojiti odgojno-obrazovne ishode vezane uz statistiku po razredima te dati preporuke za ostvarivanje odgojno-obrazovnih ishoda na primjerima.

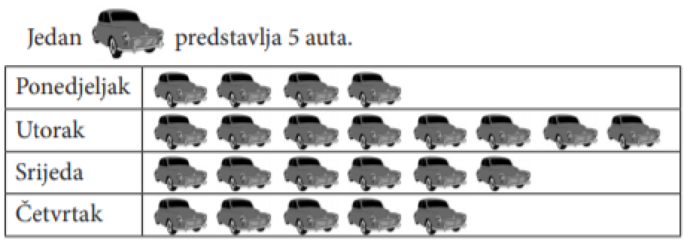
U 1. razredu osnovne škole učenici podatke iz neposredne okoline slikovno (količinski) uspoređuju na crtežima, u skupovima ili piktogramima, a kasnije i brojčano u tablicama radi donošenja jednostavnih i učenicima bliskih zaključaka. Učenicima je važno osvijestiti pojmove redak i stupac.

Primjer 1. *Provođenje slobodnog vremena* Koliko učenika u razredu trenira nogomet? Koliko učenika u razredu svakodnevno koristi tablete?

Učenici će u 2. razredu osnovne škole raditi istraživanja o neposrednoj okolini, bilježiti i razvrstavati podatke te ih prikazivati neformalnim načinima (skupovi, crteži) jednostavnim tablicama ili piktogramima. Od učenika se očekuje da kroz grupni i samostalan rad osjete zadovoljstvo u onom što rade i tumače podatke iz unaprijed pripremljenih jednostavnih tablica i piktograma.

Primjer 2. *Piktogrami* Vozila ispred škole, zanimanja roditelja.

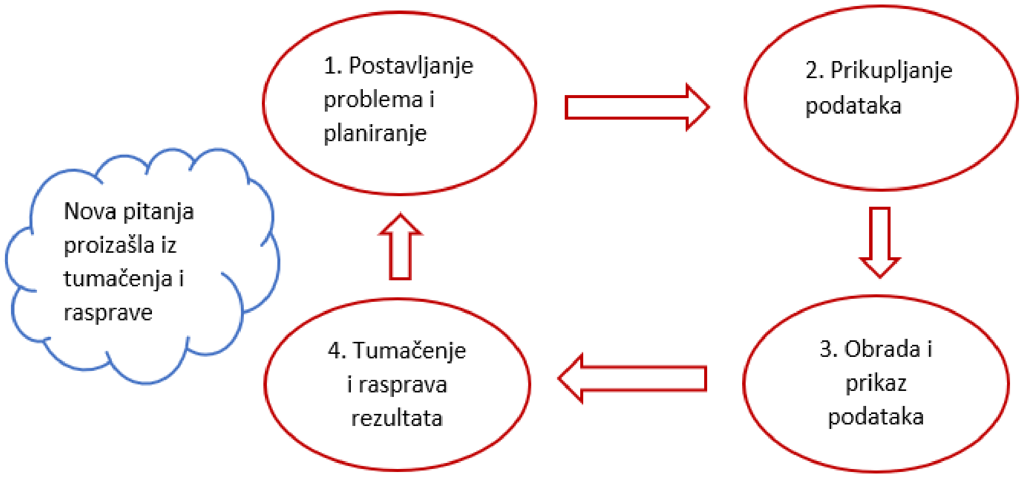
U 3. razredu učenici će na nastavi u različitim situacijama prikazivati podatke, npr. pri rješavanju problemskih situacija, a u poučavanju nastavnik se treba služiti različitim prikazima podataka pri opisivanju i objašnjavanju. Preporuča se korištenje tablice kao reprezentativnog oblika u različitim predmetima i različitim područjima života. Učenicima je važno osvijestiti pojmove: stupac, redak, polje. Korištenjem različitih programa i IKT tehnologije dobro je poticati učenike da prikazuju podatke u tablicama i dijagramima, a također ih je važno poticati na čitanje podataka iz tablica i dijagrama. Posebno se ističe piktogram i stupčasti dijagram. Grafički prikaz na Slici (1.2) pogodan je za uvježbavanje tablice množenja. Dani piktogram pogodan je za uvježbavanje množenja brojem 5.



Slika 1.2: Piktogram. [**22**]

U 4. razredu osnovne škole učenici mogu istraživati problem koji ne mora biti matematički, ali će podatke upisivati i ucrtavati u tablice ili dijagrame. Nastavnici mogu osmišljavati projekte u kojima će učenici prikupljati, razvrstavati i prikazivati podatke. Prema

[**13**] učenici do kraja 4. razreda osnovne škole sposobni su i trebaju praktično primijeniti sve problemske cikluse u statistici. Problemski ciklusi u statistici prikazani su na Slici (1.3).



Slika 1.3: Problemski ciklusi u statistici.

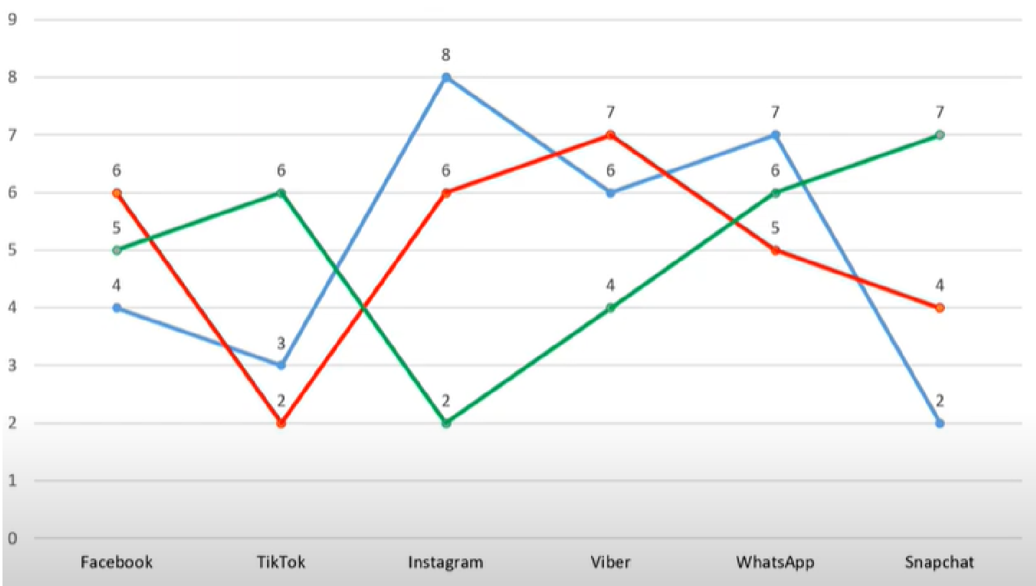
Primjer 3. *Mini projekt* Pratite, promatrajte i bilježite rezultate tijekom tjedan dana. Na kraju tjedna ih objedinite, prikažite i donesite zaključke. Kako provodiš slobodno vrijeme? Koliko se vremena posvećuje igranju igrica, a koliko gledanju i sportskim aktivnostima?

U 5. razredu osnovne škole učenici će baratati podatcima prikazanima na različite načine. Iz zadanoga prikaza učenici će određivati skup objekata, obilježja skupa, broj elemenata skupa s danim obilježjem. Ovaj bi ishod bilo korisno ostvariti provođenjem stvarnih istraživanja u nekome razdoblju: zdrava prehrana, fizičko i mentalno zdravlje, potrošnja hrane,... U pojedinim školama kao izborni sadržaj učit će se i računanje aritmetičke sredine brojčanih podataka.

Primjer 4. *Aritmetička sredina* Marko je na prvoj rukometnoj utakmici postigao 9 golova, na drugoj utakmici 13 golova, na trećoj 7 golova, a na četvrtoj 18 golova. Koliko je prosječno golova po utakmici postigao Marko?

Učenici će u 6. razredu osnovne škole prikupljati i razvrstavati podatke te određivati frekvencije razvrstanih podataka. Prikazuju podatke tablično, linijskim i stupčastim dijagramom frekvencija. U pojedinim školama kao izborni sadržaj interpretirat će rezultat dobiven računanjem aritmetičke sredine. Učenici bi trebali svladati čitanje podataka iz dvostrukoga linijskog grafa (gustoća naseljenosti, vodostaj rijeka). Sljedeći primjer prikladan je za rad u 7. razredu osnovne škole na redovitoj nastavi, ali prilagođen je za i rad u 6. razredu na izbornoj ili dodatnoj nastavi [**30**].

Primjer 5. *Čitanje podataka iz grafa* Promotrite nacrtani linijski graf na Slici (1.4) i odgovorite na pitanja:
a) Odredite ukupan broj ispitanih učenika.
b) Koja je društvena mreža najviše zastupljena među učenicima 6.razreda?
c) Koji je ukupan broj učenika svih 6. razreda koji se koriste mrežom TikTok?



Slika 1.4: Linijski graf. [**30**]

U 7. razredu osnovne škole učenici će organizirati i analizirati podatke prikazane dijagramom relativnih frekvencija. Učenici prikupljaju, razvrstavaju podatke i određuju frekvencije i relativne frekvencije razvrstanih podataka, prikazuju podatke tablično i stupčastim dijagramom frekvencija i relativnih frekvencija te kružnim dijagramom relativnih frekvencija. Učenici će analizirati rezultate i raspravljati o njima te donositi odluke na osnovi prikazanih i analiziranih podataka. Sljedeći primjer primjeren je za rad na nastavi u 7. razredu osnovne škole.

Primjer 6. *Frekvencije i relativne frekvencije*
Svaki čovjek prema spolu pripada jednoj od dviju kategorija (ženskom spolu (Ž) ili muškom spolu (M)), a prema tipu svoje krvne grupe jednoj od četiri kategorija (A, B, AB ili 0). Tablica i sadrži podatke o spolu, a tablica ii sadrži podatke o tipu krvne grupe za deset ispitanika iz nekog medicinskog istraživanja. Iz tablice i vidi se da za svakog ispitanika iz promatranog uzorka vrijednost varijable spol pripada kategoriji M ili kategoriji Ž, a iz tablice ii vide se varijable krvne grupe koje pripadaju jednoj od kategorija A, B, AB ili 0. Informacije koje je moguće dobiti iz prethodne tablice vezane su uz zastupljenost pojedine kategorije u promatranom uzorku. Ako raspolažemo s dodatnim informacijama, moguće je dobiti odgovore na sljedeća i slična pitanja:

Koliko ispitanika ženskog spola ima u promatranom uzorku?

Koliki je udio ispitanika s krvnom grupom 0 u promatranom uzorku?

Koliko ispitanika ženskog spola iz promatranog uzorka ima krvnu grupu A?

Koliki udio ispitanika muškog spola iz promatranog uzorka ima krvnu grupu B ili AB?

Frekvencije i relativne frekvencije svih kategorija varijabli spol i krvna grupa iz primjera prikazane su u tablici i - Slika (1.5) i tablici ii - Slika (1.6).

spol	frekvencija	relativna frekvencija
Ž	6	$6/10 = 0.6 = 60\%$
M	4	$4/10 = 0.4 = 40\%$

Slika 1.5: Tablica i. [7]

krvna grupa	frekvencija	relativna frekvencija
A	3	$3/10 = 0.3 = 30\%$
B	3	$3/10 = 0.3 = 30\%$
AB	2	$2/10 = 0.2 = 20\%$
0	2	$2/10 = 0.2 = 20\%$

Slika 1.6: Tablica ii. [7]

Od velike su važnosti u mnogim istraživanjima i kategorizirane tablice frekvencija i relativnih frekvencija. Frekvencije i relativne frekvencije za izmjerene vrijednosti varijable krvna grupa iz tablice i i varijable spol iz tablice ii kategorizirane su prema spolu ispitanika dane su u tablici iii - Slika (1.7) (za ženski spol) i tablici iv - Slika (1.8) (za muški spol).

spol = Ž		
krvna grupa	frekvencija	relativna frekvencija
A	2	2/6
B	2	2/6
AB	1	1/6
0	1	1/6

Slika 1.7: Tablica iii. [**7**]

spol = M		
krvna grupa	frekvencija	relativna frekvencija
A	1	$1/4 = 0.25 = 25\%$
B	1	$1/4 = 0.25 = 25\%$
AB	1	$1/4 = 0.25 = 25\%$
0	1	$1/4 = 0.25 = 25\%$

Slika 1.8: Tablica iv. [**7**]

Na temelju svih četiriju tablica možemo redom odgovoriti na pitanja postavljena u primjeru:

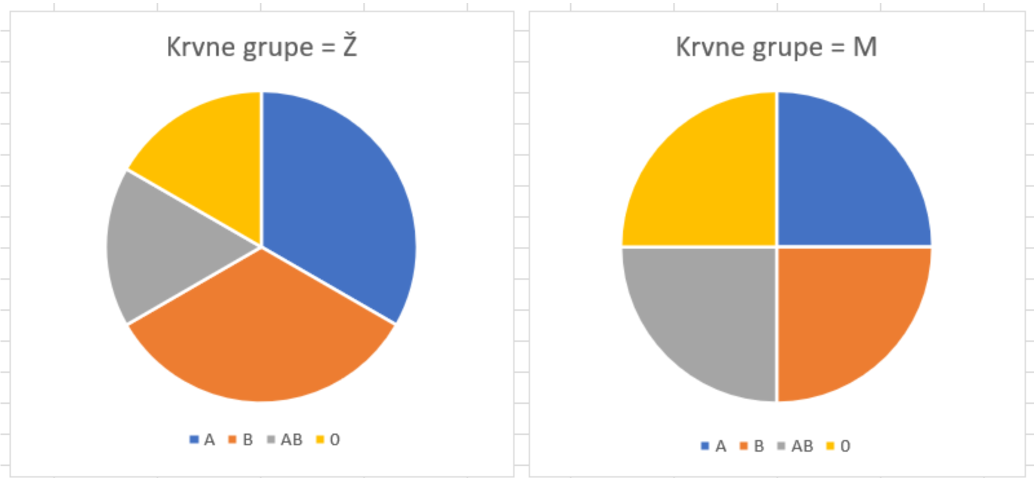
U uzorku ima šest ispitanika ženskog spola (tj. frekvencija žena u uzorku je šest).

U uzorku ima 20% ispitanika s krvnom grupom 0 (tj. relativna frekvencija krvne grupe nula u uzorku je 20%).

U uzorku imaju dvije žene s krvnom grupom A (tj. frekvencija žena s krvnom grupom A u uzorku je dva).

Od svih ispitanika muškog spola njih 50% ima krvnu grupu B ili AB.

Kao dodatak primjeru prikazat ćemo grafičke prikaze podataka u Microsoft Excelu koje su učenici sposobni samostalno prikazati kao na Slici (1.9).



Slika 1.9: Udio ispitanika u pojedinoj krvnoj grupi s obzirom na spol.

Kako bi učenici mogli samostalno riješiti primjer, očekuje se da nastavnik učenicima predstavi i definira osnovne statističke pojmove koji nisu dio kurikuluma, a koriste se u statističkim istraživanjima.

Definicija 1.3.1 ([**18**]). *Statistička populacija* je potpun skup mogućih mjerenja ili podataka o nekom kvalitativnom svojstvu koji odgovaraju cijeloj familiji jedinki o kojoj treba dati zaključak.

Definicija 1.3.2 ([**18**]). *Uzorak* iz statističke populacije je skup mjerenja (podataka) koja su sprovedena u tijekom istraživanja.

Definicija 1.3.3 ([**18**]). *Statističko obilježje ili varijabla X* je svako numeričko ili nenumeričko statističko obilježje koje se opaža (mjeri) tijekom izvodenja eksperimenta ili istraživanja.

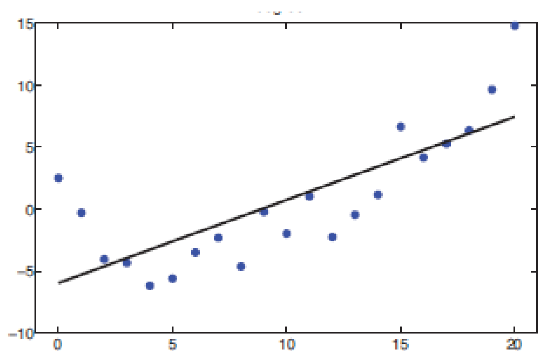
U 8. razredu osnovne škole učenici neće učiti statistiku u redovnom programu.

1.4 Analiza ishoda po razredima - srednja škola

U odgojno-obrazovnom sustavu razlikujemo kurikulum za gimnazijske programe i za srednje strukovne škole. U programima sa 105 i 140 sati nastavnih sati matematike godišnje, u

1. razredu gimnazije učenici će baratati podatcima prikazanim na različite načine. Podatke učenici prikazuju tablično, stupčastim dijagramom, histogramom, dijagramom stablo–list, linijskim dijagramom itd. Određuju statističke parametre: aritmetičku sredinu, mod, medijan, donji i gornji kvartil, varijancu i standardnu devijaciju. Na primjerima iz svakodnevnog života učenici su sposobni interpretirati dobivene podatke te prikazati dobivene mjere dijagramom brkate kutije. Dijagram brkate kutije pogodan je alat za prikazivanje mjera srednjih vrijednosti i lakšu usporedbu više skupova istovrsnih podataka. Navedeni ishodi pojavljaju se i u gimnazijskim programima sa 175 i 210 nastavnih sati matematike godišnje, ali u tim programima dodan je i dodatni ishod vezan uz primjenu normalne razdiobe. U tim programima učenici će crtati krivulju normalne razdiobe, opisivati razdiobu podataka ispod krivulje i rješavati probleme s normalnom razdiobom. U četverogodišnjim strukovnim školama učenici će računati aritmetičku sredinu statističkih podataka prikazanih na različite načine kod primjene računanja u skupu realnih brojeva pri čemu će programi s više sati ostvarivati veću korelaciju sa predmetima.

Zanimljivo je istaknuti da se statistika ne uči u 2. razredu gimnazije niti u jednom od četiriju programa, s obzirom na broj nastavnih sati matematike godišnje. U 3. razredu gimnazije, u programima sa 105 i 140 sati učenici će kao prošireni sadržaj otkriti pravac regresije i primjenjivati ga. U programima sa 175, 210 i 245 nastavnih sati učenici će primjenjivati jednadžbu pravca i tom prilikom će modelirati i interpretirati podatke s pomoću pravca regresije. Oni će naučiti razlikovati i uočavati linearni trend danih podataka. Dane podatke opisuju linearnom vezom, po mogućnosti uz uporabu tehnologije kao u primjeru na Slici (1.10). U problemima je česta korelacija s kemijom. Nastavnicima se preporuča koristiti programe dinamične geometrije te ostalim primjerenim i dostupnim interaktivnim računalnim programimom i alatom. U četverogodišnjim strukovnim školama u programu sa 175 sati godišnje ostvaruju se isti ishodi kao u gimnazijskim programima sa 175 sati, pri čemu se otvaruje korelacija s kemijom i strukovnim predmetima.



Slika 1.10: Pravac regresije. [**25**]

U 4.razredu gimnazije statistiku uče samo učenici u programu za 4.razred gimnazije, s 224 nastavnih sati matematike godišnje. Učenici će:

- primjenjivati binomnu i normalnu razdiobu
- opisivati diskretne i neprekidne slučajne varijable
- računati razdiobu, očekivanje i varijancu diskretne slučajne varijable
- računati funkciju gustoće i funkciju razdiobe neprekidne slučajne varijable
- primjenjivati binomnu i normalnu razdiobu kroz zadatke u kojima će biti česta korelacija s kemijom
- primjenjivat će diskretne i neprekidne slučajne varijable pri rješavanju jednostavnih problema.

U Tablici (1.1) prikazan je kratki pregled statistike u osnovnim i srednjim školama.

RAZRED	SADRŽAJ
1. razred OŠ	Čitanje, tumačenje i prikazivanje podataka. Piktogrami i jednostavne tablice.
2. razred OŠ	Bilježenje, razvrstavanje i prikazivanje podataka na neformalan način.
3. razred OŠ	Prikazivanje podataka (tablice, stupčasti dijagrami).
4. razred OŠ	Upisivanje podataka u dijagrame i tablice nakon istraživanja (ne)matematičkog problema.
5. razred OŠ	Grafovi i dijagrami. Crtanje grafa ili dijagrama. Očitovanje grafa i dijagrama.
6. razred OŠ	Prikupljanje, prikazivanje i tumačenje podataka. Prikazivanje podataka tablično, linijskim i stupčastim dijagramom frekvencija.
7. razred OŠ	Frekvencija. Graf frekvencija. Realtivna frekvencija. Graf relativnih frekvencija.
1. razred SŠ	Prikaz podataka (tablično, stupčastim, kružnim i linijskim dijagramom, dijagramom stabljika-list, histogramom, itd.). Mjere srednje vrijednosti (mod, medijan, gornji i donji kvartil). Standardna devijacija. Brkata kutija.
3. razred SŠ	Pravac regresije.
4. razred SŠ	Binomna i normalna razdioba.

Tablica 1.1: Statistika u osnovnim i srednjim školama.

Poglavlje 2

Opisivanje podataka

2.1 Grane statistike u hrvatskom obrazovanju

S obzirom na temeljna obilježja, statistiku dijelimo na dizajn eksperimenta, deskriptivnu (opisnu) i inferencijalnu (analitičku).

Prema [**18**] dizajn eksperimenta je grana statistike koja se bavi planiranjem eksperimenta i sakupljanjem podataka. Mnogi eksperimenti u znanosti skupi su i zahijevaju velika sredstva. Eksperimentalni dizajn težak je za provođenje u realnom svijetu i ne primjenjuje se u sustavu obrazovanja.

Deskriptivna statistika skup je postupaka kojima se sažimaju informacije sadržane u podatcima i utvrđuju glavna obilježja, odnosno činjenice o pojavi ili skupini pojava predočenih podatcima [**32**]. U metodama deskriptivne statistike rabe se postupci uređivanja, tabličnog i grafičkog prikazivanja podataka. Deskriptivna statistika bavi se također mjerama središnjih vrijednosti i mjerama raspršenosti kao osnovnim statističkim vrijednostima što je prikazano u Tablici (2.1).

Mjere srednjih vrijednosti (centralne tendencije)	**Mjere raspršenosti** (disperzije)
aritmetička sredina	raspon
medijan	gornji i donji kvartil
mod	varijanca
	standardna devijacija
	interkvartilni raspon

Tablica 2.1: Deskriptivna statistika.

Inferencijalna statistika metoda je kojom se donosi zaključak o značajkama osnovnoga skupa na temelju slučajnog uzorka kao njegova podskupa [**32**]. Prema [**18**], statističko zaključivanje obuhvaća procjenu parametara promatrane populacije, testiranje statističkih hipoteza itd.

U hrvatskim osnovnim i srednjim školama uči se deskriptivna statistika, a u gimnazijskim programima i inferencijalna statistika. Deskriptivna statistika služi za analizu podataka kako bismo ih kvalitetnije obradili, a učenici lako uočavaju vezu sa svrhovitom primjenom u svakodnevnom životu.

2.2 Uređivanje i prikazivanje podataka

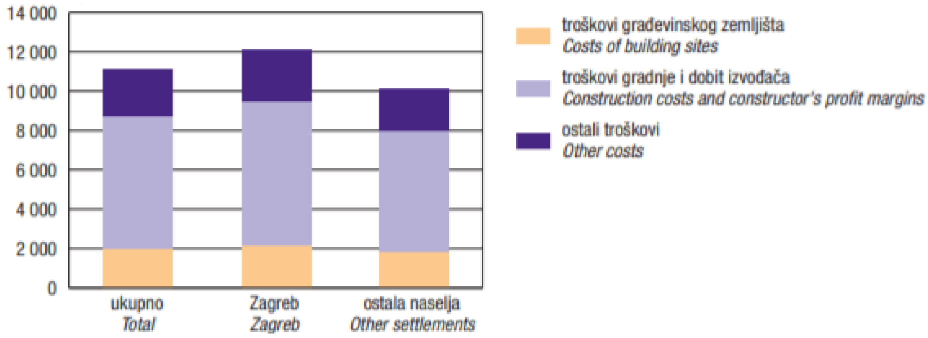
Nakon provedenog istraživanja, sve podatke kojima se raspolaže treba znati urediti na ispravan način, kako bi se donijeli valjani zaključci. Ponekad su nam podatci dani u tablicama ili u odgovarajućem grafičkom prikazu pa ih je lako ispravno interpretirati i tumačiti. No, ponekad su podatci kojima raspolažemo skup podataka koji treba urediti i prikazati. U slučaju da imamo veliki skup podataka najjednostavnije i najprikladnije je koristiti računalne programe kojima se izrađuju grafikoni i statistički računi.

Najčešće se rabe specijalizirani programi za matematiku ili statistiku, kao što su GeoGebra, Wolfram Mathematica, PSPP, IBM SPSS Statistics, zatim programi za proračunske tablice, kao što su Google Tablice, Microsoft Excel, Libre Office Calc, Apple Numbers, no i različiti programi za anketne obrasce, kvizove ili infografike mogu prikazati podatke u nekom grafičkom obliku [**3**].

Da bismo za dani skup podataka odredili najprikladniji grafički prikaz potrebno je upoznati različite grafičke prikaze podataka. U sljedećoj aktivnosti učenici će kroz tri zadatka otkriti različite načine prikazivanja podataka.

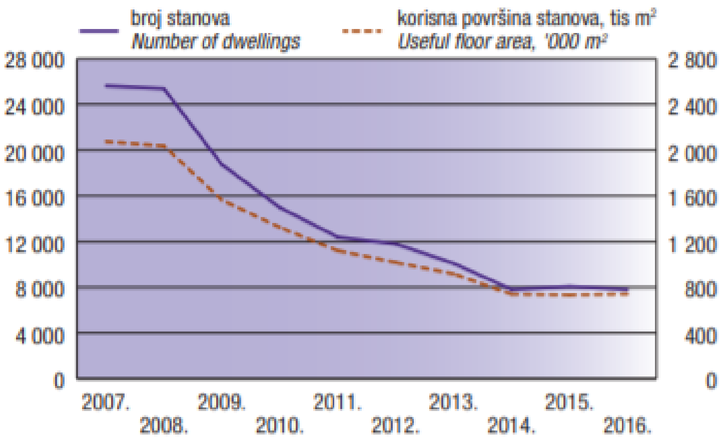
Aktivnost 1. Prikaz podataka
Cilj aktivnosti: otkrivanje prikladnog prikaza podataka
Nastavni oblik: rad u paru
Nastavna metoda: heuristička metoda, metoda usporedbe, metoda dijaloga
Potreban materijal: različiti prikazi podataka iz svakodnevnog života, tablice s grafičkim prikazima podataka

Zadatak 1.
a) Kako bismo nazvali 1. način prikazivanja podataka na Slici (2.1)?
b) Koje karakteristike uočavamo kod stupčastog dijagrama?
c) U kojim slučajevima je stupčasti dijagram primjeren za odabir?



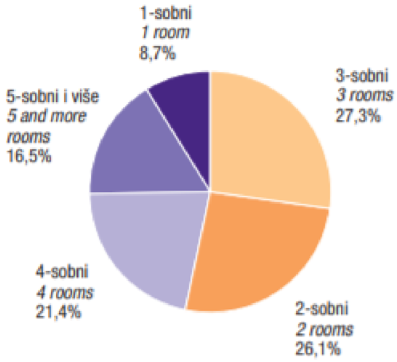
Slika 2.1: Prosječna prodaja 1 m^2 prodanoga novosagrađenog stana u 2017. [**31**]

Zadatak 2.
a) Kako bismo nazvali 2. način prikazivanja podataka na Slici (2.2)?
b) Koje karakteristike uočavamo kod linijskog dijagrama?
c) Što se najčešće prikazuje linijskim dijagramima?



Slika 2.2: Broj korisnika i prodaja završenih stanova od 2007. do 2016. [**31**]

Zadatak 3.
a) Kako se zove 3. način prikazivanja podataka na Slici (2.3)?
b) Kakve karakteristike primjećujemo kod kružnog dijagrama?
c) Što se najčešće prikazuje kružnim dijagramom?



Slika 2.3: Struktura broja završenih stanova prema broju soba u 2016. [**31**]

U prvoj skupini zadataka učenici istražuju karakteristike stupčastog dijagrama. Kod stupčastog dijagrama stupci su međusobno odvojeni i jednake su širine te je visina stupaca određena vrijednostima podataka koji se prikazuju. Stupčasti dijagram je primjeren za prikazivanje različitih vrsta podataka kada imamo mali broj vrijednosti podataka. Stupci mogu biti vodoravni, uspravni ili u obliku različitih geometrijskih tijela.

U drugoj skupini zadataka učenici uočavaju karakteristike linijskog dijagrama. Zaključuju da je ključna karakteristika da su vrijednosti podataka obilježene točkama povezanim linijama. Linijski dijagram je prikladan za prikaz promjene jedne veličine tijekom vremena, npr. promjena temperature, količine padalina, cijena i slično.

U trećoj skupini zadataka učenici iz slike kruga zaključuju da je riječ o kružnom dijagramu. Kod kružnog dijagrama vrijednosti podataka prikazane su veličinama kružnih isječaka. Kružnim dijagramom se najčešće prikazuju dijelovi cjeline ili odnosi dijelova prema cjelini, a najčešće se izražavaju u postotcima.

Stupčasti dijagram je nepraktičan ako imamo velik raspon vrijednosti podataka, npr. od 1 do 10 000. Nekada rabimo više linijskih dijagrama u istome koordinatnom sustavu jer se time omogućuje usporedba prikazanih podataka, npr. promjena temperature zraka u različitim godinama ili praćenje promjena cijena za različite proizvode. Kod kružnog dijagrama najčešće se prikazuje samo jedan skup podataka, najčešće sa šest do deset različitih vrijednosti, kako bi prikaz bio čitljiv. Skup podataka koji se prikazuje na taj način mora činiti cjelinu kako bi raspodjela veličina kružnih isječaka bila smislena. Kružnim dijagramom prikladno je prikazivati frekvencije i relativne frekvencije. U sljedećoj aktivnosti učenici će za dani skup podataka odabrati najprikladniji prikaz uz argumentaciju.

Aktivnost 2. Prikaz podataka
Cilj aktivnosti: odabiranje najprikladnijeg grafičkog prikaza podataka za dani skup podataka
Nastavni oblik: samostalan rad

Nastavna metoda: metoda zaključivanja, metoda dijaloga
Potreban materijal: tablica s danim skupom podataka, bilježnica

	2012.	2013.	2014.	2015.	2016.	2017.	2018.	2019.
RH	10 971	11 756	12 570	10 426	11 465	10 688	11 034	10 735
Grad Zagreb	12 322	12 348	12 524	11 145	11 958	11 797	10 445	12 098
Ostala naselja	9 622	10 855	9 221	9 486	9 280	9 717	9 492	9 634

Tablica 2.2: Prosječne cijene prodanih novih stanova po 1 m^2 (simulirani podatci).

Zadatak Prouči Tablicu (2.2). Što možeš zaključiti koji je grafički prikaz najprikladniji za prikaz danog skupa podataka? Argumentiraj.

Po završetku aktivnosti očekujemo da je učenik sposoban samostalno odabrati grafički prikaz za dani skup podataka prikazan tablicom te je na zadanom primjeru sposoban odrediti prednosti i nedostatke pojedinog grafičkog prikaza.

Za prikaz a u statistici nam je posebno važna posebna vrsta stupčastog dijagrama - *histogram*. Vrlo često nemamo sve podatke o nekoj populaciji nego imamo uzorak koji nam je dovoljan da bismo došli do određenih zaključaka i podataka o cijeloj populaciji. Histogrami se obično koriste u statistici kako bi pokazali koliko se određene vrste varijabli pojavljuju unutar određenog raspona. Histogram je prikaz podataka nalik na stupčasti grafikon koji raspoređuje razrede u stupce duž x - osi. Vodoravna x - os prikazuje skalu vrijednosti koja se koristi za izračun intervala razreda. Svaki razred prikazuje frekvenciju podataka u nekom intervalu. Koliko puta vrijednosti padaju u svakom razrednom intervalu označeno je pravokutnim oblicima na histogramu. Učestalost je predstavljena visinom pravokutnika, dok je interval predstavljen širinom pravokutnika. Širina histogramskih grafova s ujednačenim intervalima klasa obično je ista. Na sljedećem primjeru vidljiva je primjena korištenja histograma.

Aktivnost 3. Primjer korištenja histograma u statistici
Cilj aktivnosti: prikazivanje podataka histogramom
Nastavni oblik: samostalan rad
Nastavna metoda: metoda pismenih i grafičkih radova

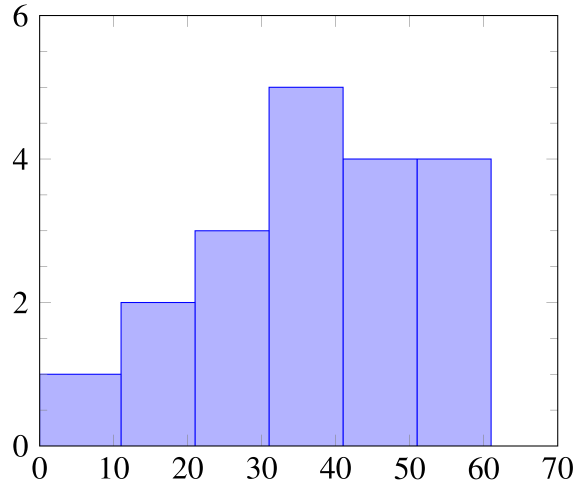
Zadatak Festival glazbe na Lastovu posjećuju posjetitelji različitih dobnih skupina. U tablici je uzet uzorak na 15 ispitanika. Prikažite podatke histogramom.

ID osobe	Broj godina
1	22
2	19
3	23
4	2
5	50
6	29
7	31
8	46
9	59
10	11
11	70
12	43
13	32
14	49
15	64

Prikazat ćemo jedno od mogućih učeničkih rješenja. Podatke možemo grupirati u razrede za prikaz broja ljudi u dobi od $0 - 10$, $11 - 20$, $21 - 30$, $31 - 40$, $41 - 50$ itd.

Razredi	Frekvencija
0-10	1
11-20	2
21-30	3
31-40	5
41-50	4
51-60	4
61-70	3

Podatke prikazujemo histogramom.



Prema [**9**] različite vrste histograma pružaju jednostavna, ali učinkovita rješenja za različita pitanja procesa. Analitičar podataka može histogram prilagoditi na nekoliko načina. Prva je promjena intervala između razreda. U primjeru kojeg analiziramo postoje razredi s intervalom od deset godina. To bi se moglo promijeniti, npr. tri razreda: $0-17$, $18-34$, $35-60$ i +60. Os y predstavlja broj ili postotak pojavljivanja u podatcima za svaki stupac i može se koristiti za vizualizaciju raspodjele podataka. Drugo razmatranje je kako prilagoditi y-os. Najosnovnija oznaka je korištenje učestalosti pojavljivanja uočenih u podacima, ali se također može upotrijebiti postotak ukupne vrijednosti ili gustoće.

Histogrami se uglavnom koriste za prikaz i organiziranje velikog skupa mjerenja ili numeričkih podataka na način prilagođen korisniku. Ciljevi za stvaranje histograma mogu se kategorizirati kako slijedi:

- grafičko sažimanje velikih skupova podataka
- usporedba rezultata s unaprijed definiranim granicama ili ciljevima specifikacija
- provjera je li došlo do promjene procesa iz jednog vremenskog razdoblja u drugo
- utvrđivanje jesu li rezultati dva ili više procesa različiti
- prenošenje stupnja ciljeva
- strateško odlučivanje i predviđanje ishoda.

Navest ćemo nekoliko primjena histograma na koje možemo naići u svakodnevnom životu. Na primjer, pri izgradnji sustava strojnog učenja, koji odlučuje o zdravstvenom riziku pacijenta, uzimamo u obzir puno podataka kao što su: dob, visina, težina, kolesterol, razina šećera itd. [**21**] Statističari također koriste histograme za analizu izbornih podataka: za

prikaz nekoliko numeričkih ocjena, poput udjela glasova određene stranke na svim biračkim mjestima, raspodjele broja važećih glasova po biračkom mjestu, raspodjele broja ljudi uključenih u izborne kampanje, analizu vremena dolaska birača itd. Svi ti čimbenici, prikazani kroz histograme, pomažu kandidatima i odgovarajućim strankama u izradi boljih strategija za njihovu izbornu kampanju. Meteorolozi i klimatolozi koriste histograme za proučavanje varijacija među parametrima: kiša, suša, snijeg,... To im omogućuje da predvide buduće slične događaje, bilo poplave, suše, oluje, mrazeve ili snježne padavine. Štoviše, meteorolozi čak mogu koristiti histograme za mapiranje kada udari grom, a ova tehnika vizualizacije pomaže im razumjeti podatke na nove načine. Histogrami su jedan od najvažnijih alata u epidemiologiji kada je u pitanju pregled i prikaz podataka povezanih s akutnim javnozdravstvenim događajem (svaka epidemija ili druga situacija koja se brzo razvija, a koja može imati negativne posljedice po zdravlje ljudi i zahtijeva hitnu procjenu i djelovanje). S velikom količinom podataka prikladnije je stvoriti razredne intervale i prema njima sortirati podatke. Interval klase može uključivati nekoliko parametara koji se međusobno isključuju, kao što su dob, krvni tlak, tjelesna temperatura, vrijeme izlaganja itd. Trgovci dioničarima proučavaju histograme koji predstavljaju ekonomsko stanje dionica tijekom vremena. Na temelju tih podataka oni stvaraju nove histograme koji predstavljaju obećavajuću imovinu. Jedan od najistaknutijih histograma na tržištu dionica je histogram konvergencije divergentnih pomičnih prosjeka (MACD), tehnički pokazatelj koji ilustrira promjene u snazi, smjeru, zamahu i trajanju trenda u cijeni dionice. Poslovna analiza je disciplina prepoznavanja poslovnih potreba i iznalaženja rješenja za različite poslovne probleme. Organizacija može iskoristiti poslovnu analizu za postizanje svojih strateških ciljeva identificiranjem i provedbom posebnih promjena. Histogram može pomoći u vizualizaciji podataka na način koji je lako razumljiv te objasniti drugima teške ideje. To će osigurati da svi u organizaciji, a ne samo visoko rangirani rukovoditelji s opsežnim vještinama analize podataka, razumiju posljedice određenih radnji. Zato su histogrami toliko prisutni u poslovnim prezentacijama [**17**].
Detaljno su navedeni primjeri uporabe histograma u svakodnevnom životu za rješavanje različitih problema. Svi navedeni primjeri bogat su izvor materijala kako bi učenici produbili razumijevanje statistike i primjereni su za interdisciplinarnu nastavu. Različite razdiobe podataka prikazujemo histogramom, o čemu raspravljamo u trećem poglavlju.

2.3 Mjere srednje vrijednosti

Detaljno ćemo pogledati koje su mjere srednje vrijednosti i kako ih odrediti. U svakodnevnim situacijama često određujemo jedan podatak koji najbolje reprezentira skup. Takav podatak nazivamo srednja vrijednost. "Prosjek" ili "prosječna vrijednost" najčešće intuitivno predstavlja srednju vrijednost. Kao što smo pokazali u primjeru 5. u 1. poglavlju, najčešće kada računamo "prosjek", računamo aritmetičku sredinu.

Definicija 2.3.1 ([**18**])**.** *Aritmetička sredina* je numerička karakteristika koja spada u mjere centralne tendencije, tj. ona mjeri "srednju vrijednost" podataka. Aritmetička sredina niza podataka $x_1, x_2, ..., x_n$ iz varijable X definirana je izrazom:

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

Aritmetičku sredinu možemo izračunati i za dani skup podataka u Tablici (2.3).

Primjer 1. U tablici su prikazani podatci o plaćama u jednoj IT firmi u Bjelovaru koja ima deset zaposlenika.

8 345,67	12 345,4	11 387,12	10 120,1	5 606,6
6 129,59	8 125,5	15 698,34	10 345,23	6 794,4

Tablica 2.3: Plaće zaposlenika (simulirani podatci).

Osim aritmetičke sredine mjere srednje vrijednosti za podatke iz tablice možemo računati mod i medijan.

Definicija 2.3.2 ([**18**])**.** *Mod* je vrijednost od X s najvećom frekvencijom.

Definicija 2.3.3 ([**18**])**.** *Medijan* je vrijednost od X za koju vrijedi da je 50% podataka manje od ili jednako toj vrijednosti i 50% podataka je veće od ili jednako njoj.

Postavlja se pitanje hoće li nam mod i medijan u nekim slučajevima dati bolju informaciju o promatranom skupu nego aritmetička sredina. Mod je najčešća vrijednost i jednak je i aritmetičkoj sredini i medijanu, ako je riječ o konstantnom nizu numeričkih vrijednosti.

Na primjeru plaća često u svakodnevnom životu možemo vidjeti da je bolje računati medijan, iako većina ljudi kao "prosjek" računa aritmetičku sredinu. U Prijedlogu direktive Europskog parlamenta i Vijeća o primjerenim minimalnim plaćama u Europskoj uniji uočavamo da nije jasna razlika između prosječne plaće i medijalne plaće. S matematičkog stajališta, i medijalna plaća je prosječna plaća: "Međunarodno priznata razina od 60% medijana bruto plaće i 50% prosječne bruto plaće može pomoći u procjeni primjerenosti minimalne plaće u odnosu na bruto razinu plaća" [**26**]. Da bismo lakše prebrodili učeničku miskoncepciju o mjerama srednje vrijednosti, prema kojoj prosjekom najčešće smatramo aritmetičku sredinu, a "prosjekom" možemo nazvati i mod i medijan, nastavnik može napraviti sljedeću aktivnost.

Aktivnost 4. Mjere srednje vrijednosti
Cilj aktivnosti: otkrivanje prednosti i nedostataka mjera srednjih vrijednosti
Nastavni oblik: diferencijalna nastava u obliku rada u skupini
Nastavna metoda: metoda usporedbe, metoda dijaloga, metoda zaključivanja
Potreban materijal: radni listić

Zadatak 1
Kupac sa sjevera želi kupiti kuću s bazenom u Komiži. Pronašao je 12 potencijalnih kuća za kupnju. Cijene tih kuća, izražene u eurima, dane su u Tablici (2.4). Koja mjera srednje vrijednosti najviše odgovara agentu nekretninama, a koja kupcu?

1 234 567	823 234	887 567	898 123	989 678	1 134 929
789 989	687 578	988 765	1 102 345	1 011 123	912 397

Tablica 2.4: Cijene nekretnina (simulirani podatci).

Zadatak 2
a) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena blizu aritmetičke sredine, koliko iznose srednje vrijednosti?
b) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena dalje od aritmetičke sredine, koliko iznose srednje vrijednosti?

Zadatak 3
a) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena blizu medijana, koliko iznose srednje vrijednosti?
b) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena dalje od medijana, koliko iznose srednje vrijednosti?

Zadatak 4
a) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena jednaka modu, koliko iznose srednje vrijednosti?
b) Ako znamo da je kupac pogledao još jednu kuću, čija je cijena različita od moda, koliko iznose srednje vrijednosti?

Nakon što na klasičan način izračunaju mjere srednje vrijednosti, učenici su sposobni zaključiti koja mjera srednje vrijednosti najviše odgovara kupcu, a koja agentu nekretnina. Učenici s nastavnikom diskutiraju i donose zaključke o dodavanju novih vrijednosti. Kod aritmetičke sredine učenici će uočiti da, ukoliko dodamo vrijednost koja je blizu aritmetičke sredine, rezultat se neće promijeniti. Ako se podacima dodaje vrijednost koja znatno odstupa od aritmetičke sredine postojećih vrijednosti, aritmetička sredina promijenit će se u velikoj mjeri. Kod medijana je obrnuta situacija. Ako dodamo vrijednost koja je blizu ili dalje od medijana, rezultat se neće u velikoj mjeri promijeniti pa možemo zaključiti da medijan ignorira udaljene vrijednosti. Aritmetička sredina osjetljiva je na dodavanje novih vrijednosti, dok medijan nije toliko osjetljiv na dodavanje novih vrijednosti. Prema definiciji mod je najfrekventnija vrijednost i glavno pitanje kod moda je može li uopće mod predstavljati "sredinu". Dodajemo li nove vrijednosti, koje su jednake ranije određenome modu, traženi mod se neće promijeniti. U slučaju da dodamo nove vrijednosti, koje su različite od ranije određenog moda, traženi mod se može promijeniti. Analogne zaključke učenici bi donijeli kada bi uklanjali vrijednosti iz niza podataka.

2.4 Mjere raspršenosti

Za analizu podataka i za dobar opis skupa podataka, potrebno je odrediti i parametre koji predstavljaju raspršenost podataka. Jedan od parametara raspršenosti podataka je raspon podataka. Iako ne govori ništa o podatcima u nizu, govori o razlici između najveće i najmanje vrijednosti podatka. Interkvartilni raspon je parametar koji sadrži raspon od 25. do 75. postotka. Pomoću interkvartilnog raspona možemo saznati koliko je studenata u 25% najboljih koji su napisali ispit ili koliko je učenika u 25% koji ne znaju skijati. Možemo saznati tražene podatke i o gornjem, a i o donjem kvartilu. Interkvartilni raspon predstavlja 50% podataka disperzije, a u stanju je identificirati i grupirati skupove podataka na oba kraja. Osjetljiv je na nekoliko ekstremnih podataka, što znači da ekstremne vrijednosti mogu uvelike promijeniti vrijednosti interkvartilnog raspona.

Varijanca i standardna devijacija u statističkoj analizi važni su parametri za izvođenje statističkih zaključaka. U gospodarskoj statistici navedeni parametri jedni su od pokazatelja za prosuđivanje rizika pri donošenju poslovnih odluka.

Definicija 2.4.1 ([7]). *Varijanca* niza izmjerenih vrijednosti $x_1, x_2, ..., x_n$ varijable X definirana je izrazom:

$$\bar{v}_n = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2.$$

Definicija 2.4.2 ([7]). *Standardna devijacija* je kvadratni korijen varijance:

$$\bar{s}_n = \sqrt{\bar{v}_n^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}.$$

U sljedećoj aktivnosti učenici će istražiti kako dodavanje i uklanjanje novih podataka utječe na varijancu i standardnu devijaciju. Aktivnost je temeljena na Core Maths Support Programme [**10**].

Aktivnost 5. Mjere raspršenosti
Cilj aktivnosti: otkrivanje utjecaja podataka na varijancu i standardnu devijaciju
Nastavni oblik: individualan rad
Nastavna metoda: metoda usporedbe, metoda dijaloga, metoda zaključivanja
Potreban materijal: radni listić

Učenici će aktivnosti izvesti po koracima i izvesti valjane zaključke.
Korak 1. Proizvoljno izaberite 5 brojeva. Izračunajte varijancu i standardnu devijaciju po formuli.
Korak 2. Iz niza brojeva izaberite 1 broj i uklonite ga te u novi niz dodajte jedan manji broj. Izračunajte varijancu i standardnu devijaciju.
Korak 3. Iz niza brojeva iz Koraka 1. uklonite isti broj koji ste uklonili u Koraku 2. i dodajte u niz jedan veći broj. Izračunajte varijancu i standardnu devijaciju.
Na proizvoljnom primjeru učenici će u zadatku rješenje dobiti iz formule. Učenici mogu zadatak istraživati na papiru, a ukoliko imaju pristup IKT tehnologiji mogu se poslužiti istom. Jedno od mogućih učeničkih rješenja prikazano je u Tablici (2.5).

1. broj	4	4	4
2. broj	27	18	38
3. broj	45	45	25
4. broj	23	23	23
5. broj	76	76	76
varijanca	590	631,76	575,76
st.dev.	24,28992	25,13484	23,995

Tablica 2.5: Jedno od mogućih učeničkih rješenja u Microsoft Excel-u.

Kroz diskusiju učenici će zaključiti da, ako u niz dodaju manji broj od uklonjenog, varijanca i standardna devijacija će se povećati. Analogno će zaključiti, da ako u niz dodaju veći broj od uklonjenog, varijanca i standardna devijacija će se smanjiti. Da bi učenici mogli provesti sljedeću aktivnost vezanu uz investiranje u tvrtku te pritom donositi smislene i valjane zaključke, očekuje se da nastavnik učenike upozna s jednostavnom formulom. Ideja aktivnosti temelji se na podatcima iz literature [**33**].

Formula standardne devijacije portfelja

$$\sigma_{SDP} = T_A \cdot \sigma_A + \mathrm{T}_B \cdot \sigma_B,$$

pri čemu je T_A = težina tvrtke A, T_B = težina tvrtke B, σ_A = standardna devijacija Društva A, σ_B = standardna devijacija Društva B, σ_{SDP} = standardna devijacija portfelja.

Aktivnost 6. Standardne devijacije
Cilj aktivnosti: uspoređivanje standardnih devijacija na zadatku iz ekonomije
Nastavni oblik: rad u paru
Nastavna metoda: heuristička metoda, metoda usporedbe, metoda dijaloga
Nastavni materijal: radni listić

Standardna devijacija u slučaju dvije tvrtke u portfelju
Obične dionice tvrtke A prodaju se po 28 USD po dionici, a iste ponude slijede nakon isplate za sljedeću godinu, što je prikazano u Tablici (2.6).

Ekonomija	Vjerojatnost	Cijene dionica	Dividenda
Procvat ekonomije	0,45	$20	$1,00
Normalna ekonomija	0,35	$30	$1,50
Recesijsko gospodarstvo	0,2	$38	$5,00

Tablica 2.6: Dionice Društva A.

Obične dionice tvrtke B prodaju se za 93 USD po dionici, a iste ponude slijede za sljedeću isplatu što je prikazano u Tablici (2.7).

Ekonomija	Vjerojatnost	Cijene dionica	Dividenda
Procvat ekonomije	0,45	$200	$7,00
Normalna ekonomija	0,35	$105	$5,50
Recesijsko gospodarstvo	0,2	$4	$2,00

Tablica 2.7: Dionice Društva B.

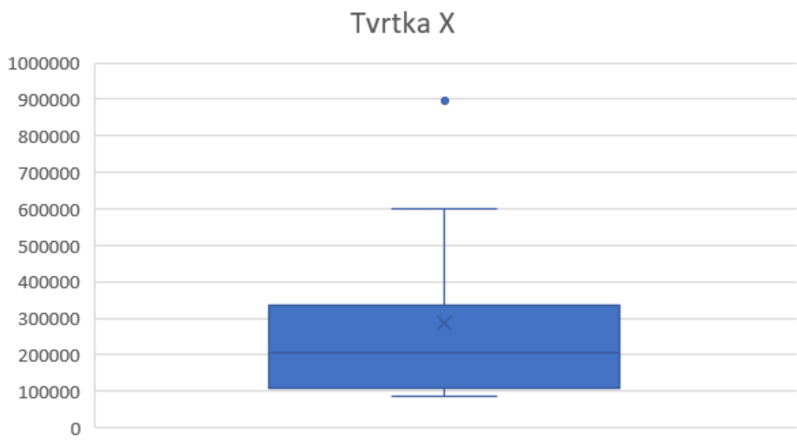
Zadatak
(a) Izračunajte standardnu devijaciju Društva A.
(b) Izračunajte standardnu devijaciju Društva B.
(c) Izračunajte standardnu devijaciju portfelja, ako se polovica ulaganja izvrši u Društvo A, a ostatak pola u Društvo B.

Analiza rješenja
Za Društvo A varijanca iznosi 895,15, a standardna devijacija 29,92.
Za Društvo B varijanca iznosi 6783,65, a standardna devijacija 82,36.
Izračun standardne devijacije portfelja polovično uloženog u Društvo A i polovično uloženog u Društvo B.
Standardno odstupanje Društva A = 29,92
Standardno odstupanje Društva B = 82,36
Težina Društva A = 0,5
Težina Društva B = 0,5
Standardna devijacija portfelja = (0,5 $\cdot 29,92) + (0,5 \cdot 82,36)$
Standardna devijacija portfelja = 56,14
Standardna devijacija portfelja niža je nego za pojedine dionice jer su dionice raznovrsne u različitim dionicama. Standardna devijacija mjeri disperziju skupa podataka, koja je relativna u odnosu na njegovu srednju vrijednost. Što je standardna devijacija vrijednosnog papira veća, veća će biti razlika između svake cijene i srednje vrijednosti, što pokazuje da je raspon cijena velik. Aktivnost je dobar primjer za ispitivanje razumijevanja standardnih devijacija.

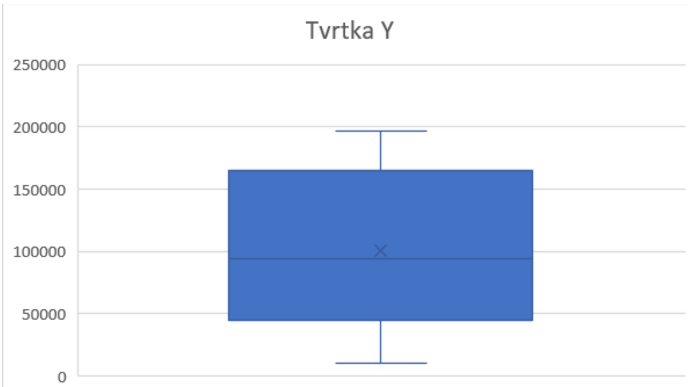
2.5 Prikaz mjera podataka

Da bismo interpretirali odnos mjera podataka, često koristimo dijagram brkate kutije koji je prikadan za uspoređivanje skupova podataka. Prema [7] brkatom kutijom ili kutijastim dijagramom prikazujemo odnos pet numeričkih karakteristika skupa izmjerenih vrijednosti: minimalnu vrijednost, donji kvartil, medijan, gornji kvartil i maksimalnu vrijednost. Na brkatoj kutiji također se označavaju takozvane stršeće vrijednosti, ako postoje.

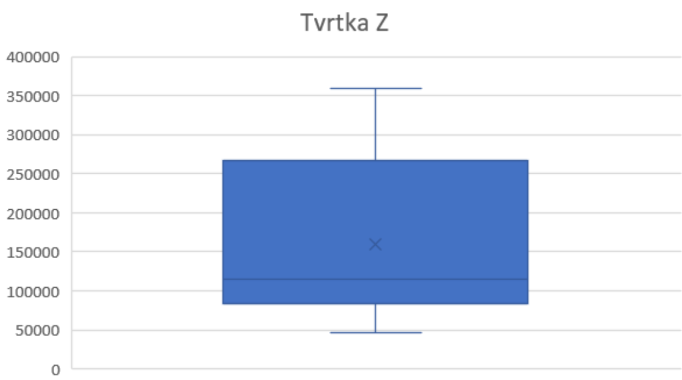
Aktivnost 7. Brkata kutija
Cilj aktivnosti: interpretacija podataka iz brkate kutije
Nastavni oblik: individualan rad
Nastavna metoda: heuristička metoda, metoda usporedbe, metoda dijaloga
Nastavni materijal: grafički prikaz u nekom od statističkih programa



Slika 2.4: Tvrtka X.



Slika 2.5: Tvrtka Y.



Slika 2.6: Tvrtka Z.

Učenicima su dana tri grafička prikaza, na kojima su prikazani pokazatelji godišnjih dobiti Tvrtki X - Slika (2.4), Tvrtki Y - Slika (2.5) i Tvrtki Z - Slika (2.6) u eurima, tijekom posljednjih 10 godina. Učenici trebaju odrediti sve pokazatelje godišnjih dobiti tvrtki koje mogu uočiti na temelju dijagrama brkate kutije i protumačiti ih. Trebaju obrazložiti u koju tvrtku bi ulagali i zašto. Učenicima treba istaknuti da grafički prikazi često ne posjeduju sva obilježja koja su nužna za ispravno donošenje odluka pa u svakodnevnom životu, najčešće medijima, možemo vidjeti manipulaciju prikazom podataka. Kao poseban primjer može se istaknuti pandemija COVID-19, za vrijeme koje se na temelju različitih grafikona rade analize i donose različiti zaključci. Štoviše, različiti analitičari na temelju podataka i grafičkih prikaza iznose i prediktivne analize, vezane uz razvoj pandemije u budućnosti.

Poglavlje 3

Slučajne varijable i razdiobe

Inferencijalna statistika, koja je dio novih gimnazijskih kurikuluma, školske godine 2021./ 2022. ulazi u 4. razrede prirodoslovno-matematičkih gimnazija, tj. u program s 224 sata nastave matematike godišnje. Ta činjenica glavna je motivacija za sadržaj trećeg poglavlja. Na početku je dana ideja eksperimenta bacanja novčića, kojom se na metodički način u nastavi, može uvesti formalni pojam slučajne varijable. S obzirom da razlikujemo dva tipa slučajnih varijabli, definirali smo diskretne i kontinuirane slučajne varijable. Za oba tipa slučajnih varijabli dani su primjeri iz svakodnevnog života te su definirani pojmovi očekivanja, varijance i standardne devijacije. Posebno je detaljnije obrađena najpoznatija razdioba podataka - normalna razdioba. Na kraju poglavlja definiran je Centralni granični teorem, važan alat koji otkriva svojstva diskretnih ili neprekidnih slučajnih varijabli, ako nam i nije poznata njihova razdioba.

3.1 Bacamo novčić

Od osnovne škole s učenicima razgovaramo kolika je mogućnost ako *bacamo novčić da se okrene na stranu na kojoj je pismo ili glava* te su upoznati s osnovama vjerojatnosti i definicijom pojmova poput *događaj, vjerojatnost događaja i skup svih mogućih događaja.* Isto tako, u predmetima poput fizike i kemije, susreću se s eksperimentima u kojima donose zaključke, a ponekad se i u nastavi matematike preporuča provođenje eksperimenata, pogotovo pri uvođenju apstraktnih pojmova.

Za početak, promatramo pokus koji ima *konačno mnogo* ishoda, a te ishode ćemo nazivati *elementarnim događajima*. Skup svih elementarnih događaja označavamo s Ω. Svaki podskup skupa Ω, koji sadržava sve povoljne ishode, nazivamo *događajem* i možemo mu pridružiti vjerojatnost. Kasnije ćemo komentirati i situacije u kojima imamo beskonačno mnogo elementarnih događaja.

Sljedeći eksperiment preporuča se provesti s učenicima kao bazu za razvoj statističkog i vjerojatnosnog mišljenja. U eksperimentu učenici će otkriti kako na jednostavan način odrediti vjerojatnost pojavljivanja nekog događaja ako imamo konačno mnogo ishoda i *svi se podjednako često pojavljuju.*

Primjer 1. (Bacanje simetričnog novčića jednom)
Promatramo slučajni pokus jednog bacanja novčića koji može imati dva moguća ishoda: pismo i glava (skraćeno: P i G). Dakle, skup svih elementarnih događaja je:

$$\Omega = \{pismo,\ glava\}.$$

Pretpostavljamo da je novčić koji bacamo pravilan ili simetričan, tj. da je vjerojatnost da se okrene pismo jednaka vjerojatnosti da se okrene glava te je taj broj jednak

$$P(pismo) = P(glava) = \frac{1}{2}.$$

Modelirali smo situaciju koristeći klasičnu (a priori) definiciju vjerojatnosti.

Definicija 3.1.1 (Vjerojatnost *a priori*)**.** *Neka imamo slučajni pokus s konačno mnogo elementarnih događaja i neka su svi ti elementarni događaji jednako mogući. Tada je vjerojatnost proizvoljnog događaja, vezanog uz taj pokus, broj elementarnih događaja povoljnih za taj događaj podijeljen s ukupnim brojem elementarnih događaja* [**29**].

Pitanje je možemo li situaciju iz Primjera 1. elegantno zapisati korištenjem matričnog zapisa. Preporuča se učenike uputiti da u prvi redak zapisuju sve elementarne događaje, a da u drugi redak zapisuju moguće vjerojatnosti za pojedini događaj. Time dobivamo tzv. *razdiobu vjerojatnosti.* U primjeru 1. razdioba izgleda ovako:

$$\begin{pmatrix} pismo & glava \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

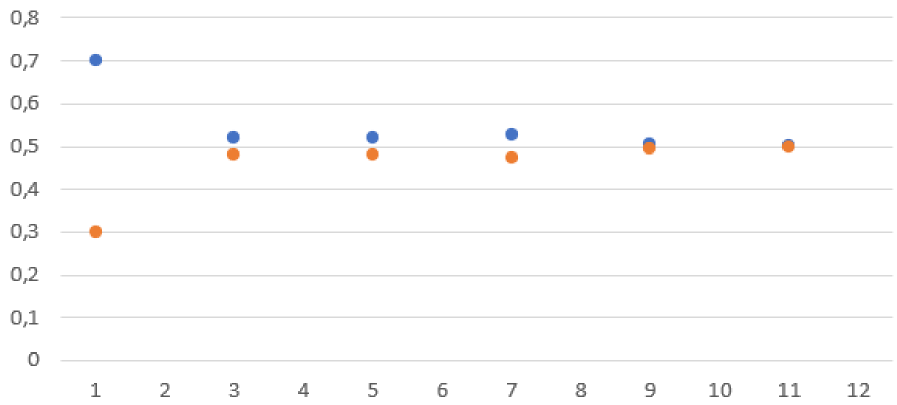
Prikazali smo koliko ima mogućih ishoda i kolika je vjerojatnost za svaki ishod ako simetričan novčić bacamo jednom. Drugačiji pristup vjerojatnosti možemo ilustrirati ako novčić bacamo više puta.

Primjer 2. (Bacanje simetričnog novčića n puta)
Promatramo slučajni pokus bacanja simetričnog novčića n puta. Brojimo koliko se puta "okrenulo pismo". Relativnu frekvenciju događaja dobivamo izračunavanjem omjera $\frac{n_P}{n}$, pri čemu je n broj bacanja novčića, a n_P broj bacanja u kojem se okrenulo pismo.
Simulacije bacanja simetričnog novčića prikazane su u Tablici (3.1).

Tablica 3.1: Bacanje simetričnog novčića n puta.

n	n_P	n_G	n_P/n	n_G/n
10	7	3	0,7	0,3
50	26	24	0,52	0,48
100	52	48	0,52	0,48
500	264	236	0,528	0,472
1 000	506	494	0,506	0,494
10 000	5 027	4 973	0,503	0,497

Bacamo li novčić veći broj puta, relativne frekvencije će se grupirati oko broja $\frac{1}{2}$. Ako se s povećanjem broja izvođenja pokusa relativne frekvencije događaja grupiraju oko nekog realnog broja, prikazuje se da je skup tih relativnih frekvencija otprilike jednak vjerojatnosti tog događaja. Prema [**29**] to svojstvo nazivamo svojstvom statističke stabilnosti relativnih frekvencija i posljedica je Zakona velikih brojeva. Na Slici (3.1) prikazane su relativne frekvencije iz simulacija za ranije opisani primjer. Na horizontalnoj osi prikazan je broj bacanja simetričnog novčića tako da smo redom n - ovima iz Tablice (3.1) pridružili neparne prirodne brojeve. Na vertikalnoj osi označene su relativne frekvencije događaja A i B. Događaj A definiramo kao "okrenulo se pismo", a događaj B kao "okrenula se glava". Plave točkice predstavljaju relativne frekvencije događaja A, a narančaste relativne frekvencije događaja B. Na grafu vidimo da se relativne frekvencije pojedinog događaja grupiraju oko jednog broja, a taj broj onda zovemo vjerojatnost tog događaja. To potvrđuje raniju tvrdnju da je vjerojatnost da se okrene pismo jednaka vjerojatnosti da se okrene glava.



Slika 3.1: Relativne frekvencije iz simulacija u slučaju bacanja simetričnog novčića n puta.

Modelirali smo situaciju koristeći eksperimentalnu (a posteriori) definiciju vjerojatnosti temeljenu na velikom broju ponavljanja pokusa.

Definicija 3.1.2 (Definicija *a posteriori*)**.** *Ako slučajni pokus zadovoljava uvjet statističke stabilnosti relativnih frekvencija, tada se vjerojatnost a posteriori proizvoljnog događaja A vezanog za taj pokus definira kao realan broj* $P(A)$, $0 \leqslant P(A) \leqslant 1$, *oko kojeg se grupiraju, odnosno kojemu teže relativne frekvencije tog događaja* [**29**].

Dalje promotrimo što se događa ako novčić nije simetričan, tj. što ako svi ishodi nisu jednako vjerojatni. Kao i za simetričan novčić prvo promatramo jedno ponavljanje.

Primjer 3. (Bacanje nesimetričnog novčića jednom)
Kada bacamo nesimetričan novčić, znači da ne bacamo novčić kod kojeg je jednaka vjerojatnost da se okrene pismo ili glava. Razmotrimo slučaj "okrenulo se pismo" s vjerojatnošću p, $0 \leqslant p \leqslant 1$. Nadalje, znamo da tada vjerojatnost ishoda "okrenula se glava" iznosi $q = 1 - p$. Razdioba vjerojatnosti u bacanju nesimetričnog novčića glasi:

$$\begin{pmatrix} pismo & glava \\ p & 1-p \end{pmatrix}.$$

Ovu razdiobu imaju svi pokusi s dva ishoda i nazivamo ju *Bernoullijeva razdioba*.

Promotrimo situaciju u kojoj nesimetričan novčić bacamo više puta. Prirodno se nadovezuje pitanje koliko ima mogućih ishoda i kolika im je vjerojatnost ako novčić bacamo više puta.

Primjer 4. (Bacanje nesimetričnog novčića n puta)
Zbog jednostavnosti pretpostavimo da smo novčić bacali pet puta i pitamo se kolika je vjerojatnost da se pismo okrenulo točno tri puta. Prebrojavanjem dobivamo da za dani problem postoji $2^5 = 32$ različita niza, npr. $\{PPPPG\}$, $\{PPPGP\}$, $\{PPGPP\}$, $\{PGPPP\}$, $\{PPPPP\}$, $\{GPPPP\}$, $\{PPPGG\}$, $\{PPGPG\}$, $\{PGPPG\}$, $\{GPPPG\}$, ...
Među njima je tada $\binom{5}{3} = 10$ nizova u kojima su tri slova P i dva slova G. Budući da se pokusi ponavljaju i nezavisni su, vjerojatnost bilo kojeg niza u kojem imamo tri uspjeha (pismo) i dva neuspjeha (glava) iznosi:

$$p^3 \cdot q^2.$$

Ukupno, vjerojatnost da se tri puta okrenulo pismo iznosi

$$\binom{5}{3} \cdot p^3 \cdot q^2.$$

Različite vjerojatnosti dobivene slučajnim pokusom nakon svakog bacanja novčića za pet bacanja možemo zapisati na sljedeći način, na elegantan način koristeći matrični prikaz:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \binom{5}{0} \cdot p^0 \cdot q^5 & \binom{5}{1} \cdot p^1 \cdot q^{5-4} & \binom{5}{2} \cdot q^2 \cdot q^3 & \binom{5}{3} \cdot p^3 \cdot q^2 & \binom{5}{4} \cdot p^4 \cdot q^1 & \binom{5}{5} \cdot p^5 \cdot q^0 \end{pmatrix}.$$

Analogno bismo u slučaju ponavljanja pokusa n puta zaključili vjerojatnost događaja da se točno k puta dogodio uspjeh jednaka

$$P(k\ uspjeha) = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

Time je dobivena razdioba pokusa koju bismo nazvali *binomna razdioba.*

3.2 Slučajne varijable

Uočimo da smo u prethodnim primjerima imali pridruživanja koja su elementima skupa Ω (elementarnim događajima) pridruživala realne brojeve. Za primjer bacanja nesimetričnog novčića definirajmo funkciju X na Ω tako da varijablama P i G pridružimo 1, ukoliko se "okrenulo pismo" te 0, ukoliko se "okrenula glava", tj.

$$X(P) = 1,\ X(G) = 0.$$

U primjeru bacanja nesimetričnog novčića n puta svakom smo od 2^5 nizova pridružili broj ponavljanja slova P, npr.

$$X(PPGPG) = 3.$$

Ovakva pridruživanja nazivamo *slučajne varijable*, a matrični prikazi za razdiobe su zapravo razdiobe tih slučajnih varijabli. Uvođenjem pojma slučajne varijable modeliramo veličine, čije se brojčane vrijednosti pridružuju pripadajućim događajima slučajnog pokusa, što znači da u trenutku njihova proučavanja ne možemo sa sigurnošću znati koja će se od tih vrijednosti realizirati. Neformalno, slučajna varijabla se opisuje kao varijabla čije vrijednosti ovise o ishodima slučajnog fenomena. Formalnije, slučajna varijabla se definira kao funkcija definirana na prostoru vjerojatnosti koji se preslikava iz prostora uzorka u stvarne brojeve. Značenje vjerojatnosti dodijeljenih potencijalnim vrijednostima slučajne varijable nije dio same teorije vjerojatnosti, već je povezano s filozofskim argumentima o tumačenju vjerojatnosti poput onog da je statistika proizašla iz vjerojatnosti i da bez vjerojatnosti nema statistike. Uobičajeno je da ishodi ovise o nekim fizičkim varijablama koje nisu predvidljive. Slučajne varijable označavat ćemo velikim slovima, recimo

X, Y, Z. U prethodnom potpoglavlju promatrali smo samo slučajeve u kojima je Ω konačan. Ako želimo raditi s (neprebrojivo) beskonačnim skupom elementarnih događaja potrebno je definirati kojim podskupovima možemo pridružiti vjerojatnosti. Te podskupove nazivamo događajima i oni moraju zadovoljavati određena skupovno teorijska svojstva. Stoga se uvodi pojam σ-algebre.

Definicija 3.2.1. Kažemo da je familija $\mathcal{A}$ podskupova od Ω *σ-algebra* ako vrijedi

1. $\emptyset \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$
3. $A_i \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}, \ i \in \mathbb{N}$.

Primijetimo da smo definirali familiju podskupova od Ω, koja je zatvorena na skupovne operacije. Kažemo da je σ-algebra familija podskupova od Ω, za koju ima smisla promatrati vjerojatnost. Elementi σ-algebre čine skup svih događaja. Ako je Ω konačan ili prebrojivo beskonačan skup, onda najčešće uzimamo $\mathcal{A} = \mathcal{P}(\Omega)$.

Definicija 3.2.2. Uređenu trojku $(\Omega, \mathcal{A}, P)$ zovemo *vjerojatnosni prostor* ako je $\mathcal{A}$ je σ-algebra na Ω, a P funkcija vjerojatnosti za koju vrijedi:

1. $P(A) \geqslant 0, \ \forall A \in \mathcal{A}$,
2. $P(\Omega) = 1$,
3. $\forall A_i \in \mathcal{A}, i \in \mathbb{N}, \ A_i \cap A_j = \emptyset, \ i \neq j \Rightarrow P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Pogledajmo ponovno bacanje novčića n puta. U tom primjeru su elementi od Ω nizovi, tj. svi nizovi P i G. Definiramo funkciju $X : \Omega \to \mathbb{N}$ koja svakom nizu pridružuje broj pisama. Funkcija X svakom nizu iz Ω pridružuje broj pisama. Uočimo da je slika funkcije skup prirodnih brojeva od 0 do n. Dalje tražimo vjerojatnost p_i da se "okrenulo pismo" i puta, tj. vjerojatnost da X iznosi i zapisujemo:

$$P(X = i) = p_i.$$

Te brojeve p_i smo izračunali u primjeru 4. i možemo primijetiti da vrijedi

$$\sum_{i=0}^{n} p_i = 1.$$

Nakon što smo postupno došli do pojma slučajne varijable, uz objašnjenja na primjerima, učenicima možemo predstaviti formalnu definiciju slučajne varijable.

Definicija 3.2.3. Neka je $(\Omega, \mathcal{A}, P)$ vjerojatnosni prostor. Funkcija $X : \Omega \to \mathbb{R}$, koja ishodima pokusa (elementima skupa Ω) pridružuje realne brojeve, je *slučajna varijabla* ako je $\{w \in \Omega : a \leqslant X(w) \leqslant b\} \in \mathcal{A}$ za sve $a, b \in \mathbb{R}$.

Uočimo da u slučaju kad je skup Ω konačan ili prebrojivo beskonačno, a σ-algebra $\mathcal{A}$ jednaka partitivnom skupu od Ω, slučajna varijabla je bilo koja funkcija $X : \Omega \to \mathbb{R}$. Razlikujemo diskretne i neprekidne slučajne varijable o kojima ćemo diskutirati u nastavku. Diskretne slučajne varijable su one u kojima je skup mogućih vrijednosti konačan ili prebrojivo beskonačan (npr. ishodi bacanja kocke), a kontinuirane slučajne varijable su one u kojima skup mogućih vrijednosti nije prebrojiv (npr. ishodi mjerenja).

3.2.1 Diskretne slučajne varijable

Slučajna varijabla je diskretna kad je skup vrijednosti (slika) konačan ili prebrojivo beskonačan. Primarno su nam važne samo vrijednosti i pripadne vjerojatnosti, tj. razdioba. Stoga se ponekad samo zadaju funkcije razdiobe, dok se vjerojatnosni prostor može konstruirati generički. No, o tome nećemo govoriti u ovom radu.

Definicija 3.2.4 ([**6**], [**7**]). *Diskretna slučajna varijabla* X je slučajna varijabla s: konačnim $\{x_1, x_2, ..., x_n\}$ ili prebrojivim $\{x_1, x_2, ..., x_n, ...\}$ skupom vrijednosti $R(X)$. Razdioba (distribucija) diskretne slučajne varijable X u potpunosti je zadana skupom $R(X)$ i pripadnim nizom $(p_i,\ i = 1, ..., n)$ (odnosno nizom $(p_i,\ i = 1, ..., n)$ ako je $R(X)$ prebrojiv skup).

Definicija 3.2.5. Za diskretnu slučajnu varijablu koja može poprimiti vrijednosti $x_1, x_2, \ldots$ *funkcija razdiobe vjerojatnosti* je funkcija za koju vrijedi:

$$f(x_i) = P\{X = x_i\} = p_i.$$

Definicija 3.2.6. Za svaki realni broj x možemo promatrati vjerojatnost događaja da je $X \leqslant x$, tj.

$$F(x) = P(X \leqslant x),$$

i tako dobivamo funkciju $F : \mathbb{R} \to \mathbb{R}$ koja se naziva *kumulativna funkcija razdiobe diskretne slučajne varijable X.*

Za kumulativnu razdiobu vrijedi da je nepadajuća i da su sve vrijednosti između 0 i 1. Razdiobe često uključuju statističke analize "brojanja" ili "koliko puta" se događaj dogodi.

Definirali smo funkciju razdiobe vjerojatnosti, koja nam je nužna za daljnje proučavanje razdioba podataka. Ako prikazujemo razdiobe matricama, pišemo npr.

$$X \sim \begin{pmatrix} 1 & ... & k... \\ p_1 & ... & p_k... \end{pmatrix}.$$

što znači

$$P(X = k) = p_k.$$

Model diskretne vjerojatnosti je statistički alat koji uzima podatke nakon diskretne razdiobe i pokušava predvidjeti ili modelirati neki ishod, poput cijene ugovora s opcijama, ili koliko je vjerojatan tržišni šok u sljedećih pet godina. U financijama se diskretne razdiobe koriste za određivanje cijena opcija i predviđanje tržišnih šokova ili recesije. Razdioba je statistički koncept koji se koristi u istraživanju podataka [**16**]. Oni koji žele identificirati ishode i vjerojatnosti određene studije, ucrtat će mjerljive točke podataka iz skupa podataka, što će rezultirati dijagramom razdiobe vjerojatnosti. Drugi nazivi za riječ razdioba jesu podjela i distribucija, a dolazi iz lat. distribution. U statistici razdiobom nazivamo matematički izraz koji opisuje vjerojatnost da će neka veličina poprimiti određenu vrijednost ili skup vrijednosti, raspodjela frekvencija kojima se u nekom skupu rezultata, poredanih od manjih prema većima, pojavljuju pojedini rezultati. Oblik krivulje razdiobe omogućuje interpretaciju rezultata. Poznat primjer razdioba za diskretnu slučajnu varijablu je Poissonova, a za kontinuiranu varijablu t-razdioba i F-razdioba. Teorijske razdiobe za diskretnu varijablu, koje koristimo u radu, su Bernoullijeva i binomna razdioba, a od kontinuiranih varijabli u radu detaljno obrađujemo normalnu razdiobu. Primjer jednog bacanja novčića primjer je Bernoullijeve razdiobe, a primjer bacanja novčića n puta primjer je binomne razdiobe. Navest ćemo još nekoliko zadataka u aktivnosti za uvježbavanje primjene koncepta binomne razdiobe.

Aktivnost 8.: Primjena binomne razdiobe
Cilj aktivnosti: primjenjivanje koncepta binomne razdiobe
Nastavni oblik: rad u paru
Nastavna metoda: metoda dijaloga, metoda rada s tekstom
Nastavni materijal: radni listić, bilježnica, statistički program

Učenici mogu u paru rješavati zadatke kako bi lakše utvrdili koncept binomne razdiobe na različitim primjerima. Kroz rješavanje diskutiraju o mogućim nedoumicama i provjeravaju vlastito razumijevanje koncepta.

Zadatak 1.
Jedna tvornica proizvodi čokoladice. Vjerojatnost da je jedna čokoladica oštećena iznosi 10%. Koja je vjerojatnost ako uzmemo uzorak od 50 čokoladica da ćemo dobiti 4 oštećene čokoladice?
Rješenje:
Na početku rješavanja treba uočiti da se radi o binomnoj razdiobi s $p = 0.1$.

Zadani podatci su:
n = broj proizvedenih čokoladica = 50
k = broj oštećenih čokoladica = 4.
Neka je X = broj oštećenih čokoladica.
Formula glasi:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Uvrstimo li podatke u formulu za binomnu razdiobu slijedi:

$$\binom{50}{4} \cdot 0.1^4 \cdot 0.9^{46} = 0.1809.$$

Vjerojatnost dobitka 4 oštećene čokoladice jednaka je 0.1809.

Zadatak 2.
Operacija uspijeva u 75% slučajeva. Koliko je vjerojatno da 75% od 32 pacijenta ima uspješnu operaciju?
Rješenje:
Na početku rješavanja treba uočiti da se radi od binomnoj razdiobi s $p = 0.75$. Izračunamo da 75% od 32 iznosi 24. Uvrstimo li podatke u formulu za binomnu razdiobu slijedi:

$$P(X = 24) = \binom{32}{24} \cdot 0.75^{24} \cdot 0.25^{12} = 0.161.$$

Zaključujemo da 0.161 pacijenata ima uspješnu operaciju.

Zadatak 3.
Upotrebom interaktivnog alata odredite vjerojatnost da će se u 100 bacanja igraće kockice petica okrenuti više od 42 puta. Je li vjerojatnije da se petica okrene barem 42 puta ili najviše 100 puta?
Vidjeli smo da, kada imamo mali broj ponavljanja, rezultate možemo ručno izračunati. No, s obzirom da u zadatku imamo veći broj ponavljanja i dobivene vjerojatnosti nije lako zbrojiti, predlaže se zadatak riješiti upotrebom interaktivnog alata. Zadatak ostavljamo za samostalno istraživanje i učenicima bi trebao služiti kao motivacija za rad s velikim brojevima.

Još jedan poznati primjer diskretne razdiobe je i bacanje kockice. U eksperimentu je utvrđeno da, kada se kockica baci 100 puta, šanse da se dobije "1" su 15-18%, a ako bacimo kockicu 1 000 puta, šanse da se dobije "1" opet su iste , što u prosjeku iznosi 16,7%. Ovakav način razmišljanja povezan je s a posteriori definicijom vjerojatnosti, što smo komentirali ranije u radu. Koristeći a priori definiciju vjerojatnosti znamo da je vjerojatnost da se dobije "1" točno $\frac{1}{6}$. Ako bacamo dvije kockice, imamo ukupno 36 mogućih ishoda i svi su jednako vjerojatni. Za daljnje proučavanje i ilustraciju pojma slučajne varijable možemo se pitati npr. kolika je vjerojatnost da, ako bacamo dvije kockice, zbroj na kockicama bude manji od 4, kolika je vjerojatnost da zbroj na kockicama bude točno 7 ili kolika je vjerojatnost da zbroj na kockicama bude veći od 9 [**1**]. Sljedeću aktivnost može se provesti s učenicima za uvježbavanje.

Aktivnost 9.: Bacamo dvije simetrične kockice jednom
Cilj aktivnosti: primjenjivanje koncepta binomne razdiobe
Nastavni oblik: samostalan rad
Nastavna metoda: metoda dijaloga, metoda rješavanja problemskih radataka
Nastavni materijal: radni listić. bilježnica, statistički program

Zadatak 1.
Zapiši razdiobu diskretne slučajne varijable zbroja dviju simetričnih kockica.

Zadatak 2.
Razdiobu diskretne slučajne varijable zbroja dviju simetričnih kockica prikaži histogramom.

Zadatak 3.
Ako bacamo dvije simetrične kockice, kolika je vjerojatnost da će zbroj na kockicama biti manji od 5? Riješi zadatak postupno kroz korake.
a) Ispiši sve moguće kombinacije uređenih parova.
b) Definiraj elementarni događaj za zadani problem i izračunaj vjerojatnost traženog događaja.
c) Izračunaj kumulatativnu funkciju razdiobe diskretne slučajne varijable X za zadani problem.

Prikazano je jedno od mogućih učeničkih rješenja.

Rješenje:
Zadatak 1.
Definirajmo slučajnu varijablu $X : \Omega \to \mathbb{R}$ tako da je $X(i, j) = i + j$. Funkcija razdiobe vjerojatnosti za navedeni problem jest:

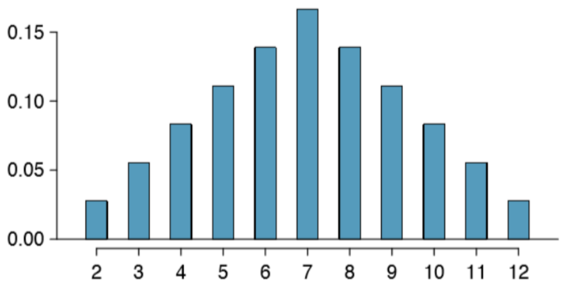
$$\begin{gathered}
\mathrm{P}(X{=}2){=}\tfrac{1}{36}\\
\mathrm{P}(X{=}3){=}\tfrac{2}{36}\\
\mathrm{P}(X{=}4){=}\tfrac{3}{36}\\
\mathrm{P}(X{=}5){=}\tfrac{4}{36}\\
\mathrm{P}(X{=}6){=}\tfrac{5}{36}
\end{gathered}$$

$$\mathrm{P}(X{=}7){=}\tfrac{6}{36}$$
$$\mathrm{P}(X{=}8){=}\tfrac{5}{36}$$
$$\mathrm{P}(X{=}9){=}\tfrac{4}{36}$$
$$\mathrm{P}(X{=}10){=}\tfrac{3}{36}$$
$$\mathrm{P}(X{=}11){=}\tfrac{2}{36}$$
$$\mathrm{P}(X{=}12){=}\tfrac{1}{36}$$

Prikazana je razdioba diskretne slučajne varijable za bacanje dvije simetrične kockice:

$$X \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}.$$

Zadatak 2. Razdioba vjeojatnosti diskretne slučajne varijable za bacanje dvije simetrične kockice prikazana je histogramom na Slici (3.2).



Slika 3.2: Razdioba vjerojatnosti zbroja dviju kockica. [**19**]

Zadatak 3.
a) Traženi uređeni parovi jesu: $(1,1),\ (1,2),\ (2,1),\ (2,2),\ (3,1),\ (1,3)$.
b) Potrebno je definirati skup elementarnih događaja.

$$\Omega = \{(i,j) : 1 \leqslant i, j \leqslant 6\}.$$

Nadalje, potrebno je definirati događaj

$$A = \{\text{"zbroj na kockicama je manji od 5"}\}.$$

S obzirom da imamo ukupno 36 mogućih ishoda, a 6 je povoljnih, vjerojatnost događaja da je zbroj na kockicama manji od 5 iznosi:

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$

c) Tražimo vjerojatnost da je zbroj na kockicama manji od 5.

$$P(X \leqslant 5) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{1}{6}$$

Vjerojatnost da je zbroj na kockicama manji od 5 iznosi $\frac{1}{6}$.

Nadalje, definirat ćemo očekivanje i varijancu za diskretne slučajne varijable. Očekivanjem diskretne slučajne varijable nazivamo broj koji upućuje na sredinu razdiobe te slučajne varijable. Varijancu slučajne varijable definiramo kao nenegativan broj koji opisuje raspršenost razdiobe slučajne varijable u odnosu očekivanje iste varijable.

Definicija 3.2.7 ([**29**])**.** Ako je $(\Omega, \mathcal{A}, P)$ prostor vjerojatnosti, te $X : \Omega \to \mathbb{R}$ diskretna slučajna varijabla koja poprima vrijednosti x_i s pripadajućim vjerojatnostima $P(X = x_i) = p_i$, tada se veličina definirana izrazom

$$E[X] = \mu = \sum_{i=1} f(x_i) p_i$$

naziva *matematičko očekivanje* diskretne slučajne varijable X. Veličina definirana izrazom

$$Var(X) = \sigma^2 = \frac{1}{k} \sum_{i=1} (x_i - \mu)^2 p_i$$

naziva se *varijanca* ili *disperzija*, a veličina

$$\sqrt{Var(X)} = \sigma = \sqrt{\frac{1}{k} \sum_{i=1} (x_i - \mu)^2 p_i}$$

standardna devijacija ili *standardno odstupanje* diskretne slučajne varijable X.

Primjer 1. *Primjer izračuna očekivanja, varijance i standardne devijacije*
Festival glazbe na Lastovu posjećuju posjetitelji različitih dobnih skupina. Uzorak od 15 ispitanika za diskretnu slučajnu varijablu X prikazan je tablicom razdiobe:

$$\begin{pmatrix} 22 & 19 & 23 & 2 & 50 & 29 & 31 & 46 & 59 & 11 & 70 & 43 & 32 & 49 & 64 \\ \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} & \frac{1}{70} \end{pmatrix}.$$

Na danom primjeru slijedi izračun očekivanja, varijance i standardne devijacije.
Rješenje:

$$E(X) = 22 \cdot \frac{1}{70} + 19 \cdot \frac{1}{70} + 23 \cdot \frac{1}{70} + 2 \cdot \frac{1}{70} + 50 \cdot \frac{1}{70} + 29 \cdot \frac{1}{70} + 31 \cdot \frac{1}{70} + 46 \cdot \frac{1}{70} + 59 \cdot \frac{1}{70} + 11 \cdot \frac{1}{70} +$$

$$70 \cdot 7\frac{1}{70} + 43 \cdot \frac{1}{70} + 32 \cdot \frac{1}{70} + 49 \cdot \frac{1}{70} + 64 \cdot \frac{1}{70} = 7.84$$

$$V(X) = 22^2 \cdot \frac{1}{70} + 19^2 \cdot \frac{1}{70} + 23^2 \cdot \frac{1}{70} + 2^2 \cdot \frac{1}{70} + 50^2 \cdot \frac{1}{70} + 29^2 \cdot \frac{1}{70} + 31^2 \cdot \frac{1}{70} + 46^2 \cdot \frac{1}{70} + 59^2 \cdot \frac{1}{70} +$$

$$11^2 \cdot \frac{1}{70} + 70^2 \cdot \frac{1}{70} + 43^2 \cdot \frac{1}{70} + 32^2 \cdot \frac{1}{70} + 49^2 \cdot \frac{1}{70} + 64^2 \cdot \frac{1}{70} = 366.69$$

$$\sigma(X) = \sqrt{V(X)} = 19.15$$

U eksperimentu u kojem smo kockicu bacali n puta otkrili smo svojstvo statističke stabilnosti relativnih frekvencija koja je osnova za Zakon velikih brojeva. Za iskaz teorema potrebno je razumijevanje slučajnih varijabli i binomne razdiobe. Neformalno, Zakon velikih brojeva tvrdi da će prosječni rezultat ponavljanja nekog događaja velik broj puta biti blizu određene vrijednosti ako događaj uključuje slučajne varijable.

3.2.2 Kontinuirane slučajne varijable

Drugi tip slučajnih varijabli jesu kontinuirane pri čemu je skup vrijednosti dan intervalom ili unijom intervala. *Kontinuirana slučajna varijabla* je slučajna varijabla kojoj je skup vrijednosti unija intervala realnih brojeva. Distribuciju neprekidne slučajne varijable X zadajemo nenegativnom *funkcijom gustoće* f, čiji je nepravi integral jednak 1 [**2**].

Definicija 3.2.8 ([**6**])**.** Za slučajnu varijablu X kažemo da je *kontinuirana slučajna varijabla* ako postoji nenegativna realna funkcija $f : \mathbb{R} \to \mathbb{R}$, definirana na skupu realnih brojeva, takva da je za $x, y \in \mathbb{R}$ $(x \leqslant y)$ vjerojatnost

$$P(x \leqslant X \leqslant y) = \int_x^y f(x)dx.$$

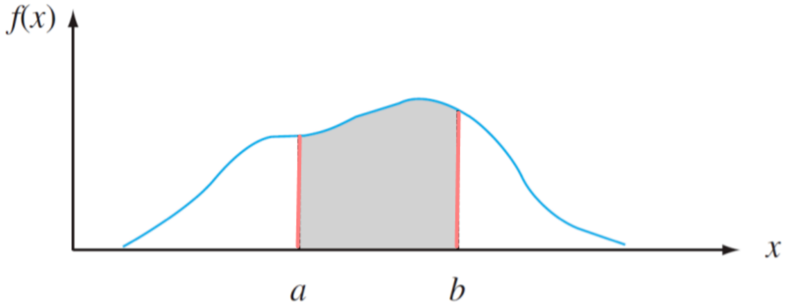
Takvu funkciju f zovemo *funkcija gustoće kontinuirane slučajne varijable* X.

Definicija 3.2.9. Funkcija gustoće vjerojatnosti $f(x)$ je derivacija kumulativne funkcije razdiobe $F(x)$:

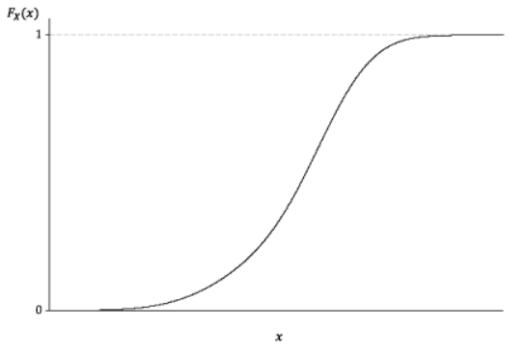
$$f(x) = F'(x)$$

odnosno, *kumulativna funkcija razdiobe* $F(x)$ može se na ovaj način definirati iz funkcije gustoće vjerojatnosti $f(x)$:

$$F(x) = P(X \leqslant x) = \int_{-\infty}^x f(x)dx.$$



Slika 3.3: Primjer grafa funkcije gustoće kontinuirane slučajne varijable.



Slika 3.4: Graf kumulativne funkcije razdiobe. [**12**]

Uočimo da vjerojatnost $P\{a \leqslant X \leqslant b\}$ zapravo predstavlja površinu između x-osi i grafa funkcije f na intervalu [a, b], što je prikazano na Slici (3.3). Na Slici (3.4) prikazan je primjer kumulativne funkcije razdiobe $F(x)$ jedne kontinuirane slučajne varijable. Očekivanje kontinuirane slučajne varijable X s funkcijom gustoće f računamo pomoću određenog integrala funkcije $xf(x)$. Iskazat ćemo formalnu definiciju očekivanja kontinuirane slučajne varijable, a nakon toga i formalnu definiciju varijance kontinuirane slučajne varijable.

Definicija 3.2.10. Ako je za kontinuiranu slučajnu varijablu X dana funkcija gustoće $f(x)$, tada *očekivanje kontinuirane slučajne varijable* definiramo kao:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

Ako je za kontinuiranu slučajnu varijablu X dana funkcija gustoće $f(x)$, tada *varijancu kontinuirane slučajne varijable* definiramo kao:

$$Var(X) = E[X^2] - \mu^2 = \left(\int_{-\infty}^{\infty} x^2 \cdot f(x)dx\right) - \mu^2.$$

Primjer 2. *Primjer izračuna očekivanja, varijance i standardne devijacije*
Neka je X slučajna varijabla čija je vrijednost vrijeme trajanja akumulatora i dana je kumulativnom funkcijom razdiobe

$$F(x) = P(X \leqslant x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x^2, & 0 \leqslant x < 2 \\ 1, & x \geqslant 2. \end{cases}$$

Odredimo funkciju gustoće od X, izračunajmo pripadno očekivanje, varijancu i standardnu devijaciju te odredimo vjerojatnost da se akumulator potrošio u razdoblju od 1.5 godine do 2 godine.

Rješenje:
Imamo

$$f(x) = F'(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \leqslant x < 2 \\ 0, & x > 2. \end{cases}$$

Dakle,

$$R(X) = [0, 2]$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^2 \frac{x^2}{2}dx = \frac{4}{3}$$

$$Var(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \mathbb{E}(X)^2 = \int_0^2 \frac{x^3}{2}dx - \frac{16}{9} = \frac{2}{9}$$

$$\sigma(X) = \sqrt{Var(X)} = \frac{\sqrt{2}}{3}.$$

Tražena vjerojatnost je

$$\mathcal{P}(1.5 < X \leqslant 2) = F(2) - F(1.5) = 0.4375.$$

Slučajan pokus mjerenja tlaka, gdje je izmjerena vrijednost tlaka potpuno slučajna pojava, primjer je kontinuirane slučajne varijable. Jedan od primjera kontinuirane slučajne varijable je i normalna razdioba o kojoj detaljno pišemo u idućem potpoglavlju.

3.3 Normalna razdioba

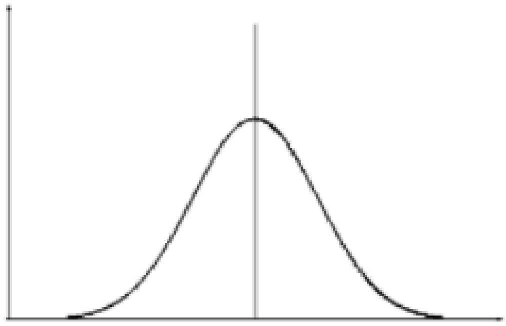
Vidjeli smo da postoje eksperimenti u kojima možemo koristiti eksperimentalnu definiciju vjerojatnosti, no postoje problemi koje ne možemo riješiti pomoću teorije vjerojatnosti. Na pitanja poput kolika je razlika u u prosječnoj težini novorođenčadi u SAD-u i u Brazilu, broj gostiju koji dnevno dolaze na kavu u nekom kafiću, koliko stanovnika nekog grada reciklira papir, a koliko plastiku - potrebno je izvršiti statistička istraživanja. Za jednostavnije probleme dovoljna nam je vjerojatnost, što smo pokazali, dok je za preciznost složenih problema potrebna statistika. Razdioba kod koje vrijednosti obilježja raspoređuju oko prosjeka naziva se normalna razdioba. Na tržištu dionica vrijednosti Forex kamatnih stopa, indeksa cijena i povrata cijena dionica često tvore krivulju normalne razdiobe. Krivulju normalne razdiobe često nazivamo zvonolikom, s obzirom da ima oblik zvona. Normalne razdiobe omogućuju analitičarima i ulagačima statističke zaključke o očekivanom prinosu i riziku dionica. Nakon što smo prikazali različite primjere normalne razdiobe u svakodnevnom životu, pogledajmo kako je krivulja normalne razdiobe otkrivena.

3.3.1 Otkriće Gaussove krivulje kroz povijest

Mnogi autori pripisuju otkriće normalne razdiobe i krivulje Johannu Carlu Friedrichu Gaussu, ali su i drugi matematičari promatrali pojam normalne razdiobe u kontekstu vjerojatnosti. Kada je rješavao kockarski problem, Abraham de Moivre razmišljao je kako aproksimirati binomnu razdiobu za jako veliki broj bacanja kockice. Primijetio je da se povećanjem broja bacanja kockice, tj. povećanjem broja n graf promatrane binomne razdiobe diskretne varijable približava tzv. zvonolikoj krivulji. Moivre je u svojoj knjizi The Doctrine of chances opisao izgled krivulje te zapisao njenu formulu. U isto vrijeme, zbog preciznog mjerenja u astronomiji, krenula je potraga za krivuljom pogreške, a to je bila još jedna od prvih primjena krivulje normalne razdiobe u praksi. Krivulju koja nalikuje na zvonoliku krivulju promatrao je Pierre-Simon Laplace. Možemo reći da je Laplace bio svjestan značenja otkrića krivulje normalne razdiobe jer je dokazao da je površina ispod zvonolike krivulje jednaka 1. Iako Moivre nije dokazao navedenu tvrdnju, ipak se danas teorem u kojem se tvrdi da se binomna razdioba pri velikom broju ponavljanja aproksimira normalnom razdiobom, naziva Moivre-Laplaceov teorem. Nakon Moivrea i Laplacea ipak je Gauss generalizirao taj teorem. Kao posljedicu toga, danas takvu zvonoliku krivulju nazivamo Gaussova krivulja [**27**].

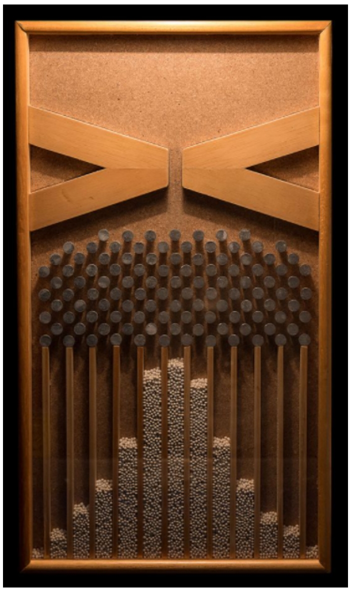
3.3.2 Otkrivanje normalne razdiobe u nastavnom procesu

U sljedećoj aktivnosti učenici će korištenjem pomagala otkriti normalnu razdiobu na nastavnom satu. Na Slici (3.5) dan je primjer krivulje normalne razdiobe.



Slika 3.5: Primjer krivulje normalne razdiobe.

U praktičnoj nastavi postoje različita pomagala kojima se učenicima može približiti normalna razdioba. Jedno od najpoznatijih pomagala je Galtonova "daska s čavlima", prikazana na Slici (3.6).



Slika 3.6: Galtonova "daska s čavlima". [**5**]

"Daska s čavlima" je koso položena plitka kutija, koja u sredini na gornjem dijelu ima lijevak u koji se ubacuju kuglice. U sredini se nalaze čavlići koji sprečavaju da se kuglice bez prepreka spuštaju prema dnu kutije. Na dnu kutije su pregrade koje grupiraju kuglice, tj. možemo precizno odrediti kamo je pojedina kuglica pala. U demonstraciji dobivanja normalne razdiobe učenici trebaju kuglice ubacivati kroz lijevak i promatrati kako putuju prema dnu kutije. Uočavaju histogram, a znaju da histogramom prikazujemo velike količine podataka za koje uzimamo uzorke i donosimo statističke zaključke. Opažaju da se kuglice grupiraju u "razrede". Budući da imamo konačno kuglica, ovo je zapravo histogram diskretne slučajne varijable, a ako imamo puno kuglica, onda se taj histogram približava području ispod zvonolike krivulje. Kasnije ćemo to formalizirati u obliku Centralnog graničnog teorema. U nastavku je prikazana aktivnost otkrivanja Gaussove ili normalne razdiobe.

Aktivnost 10.: Otkrivanje Gaussove ili normalne razdiobe
Cilj aktivnosti: otkrivanje normalne razdiobe i normalne krivulje
Nastavni oblik: praktični rad
Nastavna metoda: heuristička metoda, metoda dijaloga, metoda demonstracije
Nastavni materijal: bilježnica

U aktivnosti otkrivanja Gaussove ili normalne razdiobe učenici trebaju opisati kako se kuglice, koje ubacuju kroz lijevak, raspoređuju na dnu kuglice. Kroz diskusiju učenici zaključuju da se kuglice grupiraju u "razrede" u obliku zvonolike krivulje.

Kroz praktični rad učenici lako mogu uočiti značajke normalne razdiobe:

- kuglica ima najviše u sredini, a prema krajevima u svakoj pregradi ima ih sve manje i manje
- ubacivanje kuglica u lijevak koji se nalazi na sredini, predstavlja tendenciju grupiranja rezultata oko sredine, a čavlići koji ometaju kuglice predstavljaju tendenciju raspršenja rezultata.

3.3.3 Definicija normalne razdiobe

Razdioba prikazana u prethodnom potpoglavlju naziva se *Gaussova ili normalna razdioba*, a krivulja koja prikazuje tu razdiobu, *Gaussova ili normalna krivulja*. Slijedi formalna definicija normalne ili Gaussove razdiobe.

Definicija 3.3.1. Za kontinuiranu slučajnu varijablu X kažemo da ima *normalnu ili Gaussovu razdiobu* s parametrima μ i σ^2 i pišemo $X \sim \mathcal{N}(\mu, \sigma^2)$, ako je njezina funkcija gustoće vrijednosti zadana formulom:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

gdje je: σ - standardna devijacija, σ^2 - varijanca, μ - očekivanje. Takvu slučajnu varijablu zovemo *normalna slučajna varijabla*. Specijalno, ako je $\sigma = 0$, $\sigma^2 = 1$, normalnu slučajnu varijablu zovemo *standardna normalna slučajna varijabla*.
Opisat ćemo postupak standardizacije. Neka je X normalna slučajna varijabla $X \sim \mathcal{N}(\mu, \sigma^2)$, tada je slučajna varijabla

$$Z = \frac{X - \mu}{\sigma}$$

standardna normalna slučajna varijabla (tj. normalna slučajna varijabla s očekivanjem 0 i varijancom 1). Vrijednost varijable Z zapravo predstavlja udaljenost od očekivanja izraženu u dijelovima standardne devijacije.

Definicija 3.3.2. *Kumulativnu funkciju distribucije standardne normalne slučajne varijable* označavamo sa Φ i vrijedi:

$$\Phi(x) = P(Z \leq x).$$

Za $x \leqslant 0$ vrijedi $\Phi(x) = 1 - \Phi(-x)$.

Sažetak svojstava teorijske normalne razdiobe

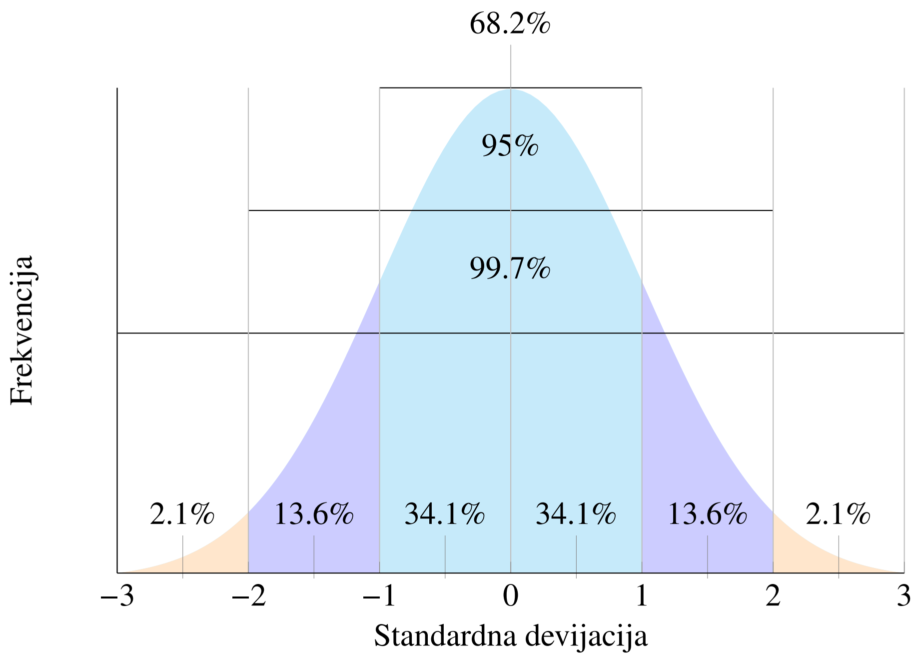
1. Normalna krivulja razdiobe je zvonolika.

2. Krivulja je simetrična u odnosu na srednju vrijednost, što je ekvivalentno reći da je njezin oblik isti s obje strane okomite crte koja prolazi kroz središte. Razlikujemo asimetričnu pozitivnu krivulju ulijevo i udesno.

3. Krivulja nikada ne dodiruje x - os. Teoretski, bez obzira koliko daleko u oba smjera krivulja se proteže, nikada se ne susreće s x - osi, ali se sve više približava.

4. Ukupna površina pod normalnom krivuljom razdiobe jednaka je 1 ili 100%. Ova činjenica može se činiti neobičnom jer krivulja nikada ne dodiruje x - os.

Kao što smo rekli u uvodnom dijelu, normalnu razdiobu koristimo kada tražimo prosjek neke vrijednosti obilježja za veliki broj ljudi. Na danoj populaciji uzima se reprezentativan uzorak da bi se izmjerila pojava slučajnih događaja. Što je uzorak veći, to je bliže stvarnim vrijednostima u populaciji. Skupina na kojoj obavljamo mjerenja morala bi biti heterogena po svojstvu koje se mjeri. Na primjer, mjerimo li težinu novorođenčadi, skupina koju mjerimo treba biti izjednačena po spolu i mjesecu rođenja i eventualno još nekim drugim svojstvima poput mjesta rođenja, ali ne smije biti selekcionirana po težini novorođenčadi, tj. treba izmjeriti sve članove skupine, a ne samo one iznimno teške ili iznimno lake. Podatci o nekoj pojavi najčešće su prikupljeni u različitim situacijama i na različitim uzorcima pa očekujemo da će se podatci u tim skupinama razlikovati u artimetičkim sredinama i standardnim devijacijama. Npr. rezultati testiranja trčanja na 500 m na nastavi tjelesne i zdravstvene kulture u 3. razredu srednje škole. Svaki razred će postići drugačiji broj bodova s nekom aritmetičkom sredinom i standardnom devijacijom. Takve podatke nije moguće uspoređivati ako želimo usporediti uspješnost učenika 3. a i 3. b razreda. U slučaju da su podatci normalizirani (standardizirani), kako bi imali isti prosjek i standardnu devijaciju, tada ih je moguće uspoređivati.



Područje ispod dijela normalne krivulje, koje se nalazi unutar jedne standardne devijacije, od prosjeka je približno 0,68 ili 68%; unutar dvije standardne devijacije od prosjeka je približno 0,95 ili 95%; unutar tri standardne devijacije od prosjeka je približno 0,997 ili 99,7%. Općenito, za sve pojave koje se mogu opisati normalnom razdiobom, izračunamo koliki se postotak od ukupnog broja podataka nalazi u određenim intervalima. Takvi intervali zovu se intervali pouzdanosti.

Kad crtamo graf normalne razdiobe, tada su vrijednosti varijable X označene na osi apscisa, dok os ordinata označava frekvenciju (tj. gustoću) pojavljivanja varijable X. Stoga je vjerojatnost da vrijednost slučajne varijable X upadne u interval [a, b] jednaka površini ispod grafa funkcije gustoće na tom intervalu. Najviše podataka je u intervalima oko prosjeka (očekivanja), gdje krivulja doseže najveću visinu. Kako se krivulja pomiče od središta, tako se vrijednosti smanjuju. Ta pojava govori nam da je vjerojatnije da će se dogoditi događaj oko prosjeka (očekivanja), a da je manja vjerojatnost da će se dogoditi događaji koji su udaljeniji (ili lijevo ili desno) od prosjeka. Standardna devijacija σ određuje raspršenost od prosjeka, a očekivanje μ određuje točku kroz koju prolazi os simetrije. Ako se vrijednost standardne devijacije σ povećava, krivulja se snižava i širi, a ako se vrijednost standardne devijacije smanjuje, krivulja se povisuje i sužava uz konstantno očekivanje. Ako se povećava vrijednost očekivanja μ, krivulja se translatira udesno, a ako se vrijednost očekivanja μ smanjuje, krivulja se translatira ulijevo, pri čemu standardna devijacija σ ostaje jednaka. U nastavku rada objasnit ćemo kako koristiti tablicu z - vrijednosti koja nam je potrebna za računanje normalne razdiobe, a onda ćemo prikazati primjenu normalne razdiobe.

3.3.4 Tablica z - vrijednosti

	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Slika 3.7: Tablica z - vrijednosti. [**18**]

Tablica z - vrijednosti prikazana je na Slici (3.7). Standardna normalna radioba omogućuje računanje vjerojatnosti. Pritom koristimo tablicu vrijednosti funkcije razdiobe jedinične normalne razdiobe. No, danas postoje kalkulatori ili različiti interaktivni programi, koji mogu izračunati ono što se traži, tako da nije potrebno računati "ručno". Svakako je potrebno znati i razumijeti kako se očitavaju vrijednosti iz tablice. Npr. za $z = 1,57$ dio broja koji završava s prvom decimalom pronađemo u 1. stupcu, a dio broja koji ima drugu decimalu pronađemo u 1. retku. Na mjestu gdje se sijeku iz 1. stupca 1,5 i iz 1. retka 0,07 nalazi se vrijednost 0,9616, a to znači da je 96,16% većih od podataka čija je z - vrijednost 1,57.

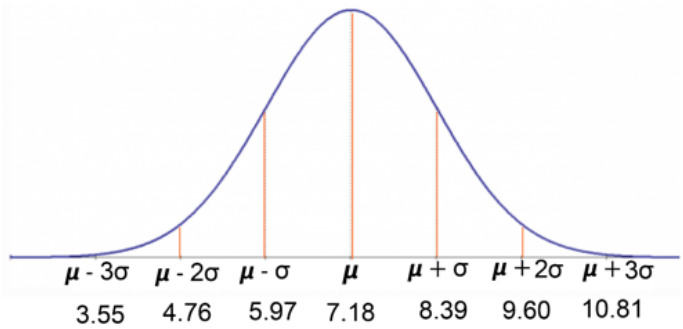
3.3.5 Primjena normalne razdiobe

Kada se rade statistička istraživanja u svakodnevnom životu, mjere se reprezentativni uzorci. Ako se srednjim vrijednostima izmjerenim u uzorcima približavamo očekivanju slučajne varijable za cijelu populaciju, tada zaključujemo da imamo normalnu razdiobu podataka. Navest ćemo primjere primjene normalne razdiobe u svakodnevnom životu. Visina stanovništva primjer je normalne razdiobe. Većina ljudi u određenoj populaciji prosječne je visine. Broj ljudi viših i nižih od prosječne visine ljudi je gotovo jednak, a vrlo mali broj ljudi je ili izuzetno visok ili izuzetno nizak. Međutim, visina nije jedinstvena karakteristika, na visinu utječe nekoliko genetskih i okolišnih čimbenika. Stoga slijedi normalnu razdiobu. Zanimljiva je činjenica da je, prema podacima prikupljenim u SAD-u, prodaja ženskih cipela prema veličini normalno raspoređena jer je fizički sastav većine žena gotovo isti. Spomenuli smo poznati problem određivanja težine novorođenčadi koja se kreće od 2,5 do 3,5 kg. Porođajna težina također slijedi normalnu krivulju razdiobe. Ideje za sljedeće dvije aktivnosti dobivene su iz [**8**] i [**10**].

Aktivnost 11.: Modeliranje normalnom razdiobom
Cilj aktivnosti: primjena svojstava normalne razdiobe
Nastavni oblik: individualan rad
Nastavna metoda: heuristička metoda, metoda dijaloga
Nastavni materijal: radni listić, interaktivni alat

Problem
U Royal Hospital bolnici jučer je rođeno 12 beba, što je dnevni rekord. Jake, sin Jasmine i Jeremy Whitehouse, teži 10,8 Ibs [1] i najteža je beba u bolnici u posljednjih 5 godina. No, je li Jake baš izuzetno težak? Isti dan rođena je i Rose i njena težina iznosi 5,6 Ibs. Je li Rose beba izrazito male težine? Koristeći poznata svojstva normalne razdiobe, učenici će pomoću nacrtane krivulju normalne razdiobe (3.8) donijeti zaključke.

[1] Jedna funta definirana je kao jedinica težine. 1 Ibs = 0.453592 kg.



Slika 3.8: Krivulja normalne razdiobe sa zadanim podatcima. [**8**]

Učenici na temelju slike trebaju zaključiti da prosječno novorođenče ima težinu 7,18 Ibs sa standardnom devijacijom 1,21 Ibs. Na temelju podataka zaključuju da standardna normalna distribucija glasi

$$X \sim \mathcal{N}(5,18,\ 1,21^2).$$

Iz krivulje normalne razdiobe, sa zadanim podatcima, učenici uočavaju da 68% beba ima između 5,97 i 8,39 Ibs. Također, uočavaju da 95% beba ima između 4,76 i 9,6 Ibs kao i da 99,7% beba ima između 3,55 i 10,81 Ibs. S obzirom da Jake ima 10,8 Ibs i njegova težina pripada intervalu unutar tri standardne devijacije s desna, učenici zaključuju da je Jake iznadprosječno teško novorođenče i možemo reći da je Jake izuzetno težak. Rose ima težinu 5,6 Ibs i njezina težina pripada intervalu unutar dvije standardne devijacije s lijeva pa učenici zaključuju da je Rose malene težine, ali ne i izrazito malene težine.

Aktivnost 12.: Zvonolika krivulja u svijetu oko mene
Cilj aktivnosti: primjena koncepta standardne normalne slučajne varijable
Nastavni oblik: rad u paru
Nastavna metoda: metoda pisanog rada, metoda rješavanja problema
Nastavni materijal: bilježnica

Učenici u paru rješavaju zadani zadatak i diskutiraju o dobivenim rješenjima. Kroz rješavanje zadatka učenici će primijeniti definiciju standardne normalne slučajne varijable.

Zadatak
Istraživanje je pokazalo da žene u prosjeku potroše 146,21 USD na kozmetičke proizvode ljetnih mjeseci. Pretpostavimo da je standardna devijacija 29,44 USD. Pronađi postotak žena koji potroše manje od 160,00 USD.

Jedno od mogućih učeničkih rješenja:
Pretpostavimo da je varijabla normalno distribuirana. Da bismo našli rješenje, trebamo

pronaći z vrijednost koja odgovara 160,00 USD. Prema formuli slijedi:

$$z = \frac{X - \mu}{\sigma} = \frac{160,00 - 146,21}{29,44} = 0.47.$$

Dakle, 160,00 USD je 0,47 standardne devijacije iznad srednje vrijednosti od 146,21 USD, što je prikazano u z razdiobi na Slici (3.9).

Slika 3.9: Prikaz normalne razdiobe za zadani zadatak.

U zadnjem koraku u rješavanju zadatka treba pronaći područje pomoću tablice z-vrijednosti. Površina ispod krivulje lijevo od z = 0,47 iznosi 0,6808. Zaključujemo da 0,6808 ili 68,08% žena troši manje od 160,00 USD na kozmetičke proizvode tijekom ljetnih mjeseci.

3.4 Centralni granični teorem

Vidjeli smo ranije da će prosjeci koje izračunamo biti distribuirani normalnom razdiobom. Takav primjer je i tvrdnja iz koje saznajemo informacije iz svakodnevnice: "Svaki mjesec jedna osoba, koja živi u Republici Hrvatskoj troši prosječno 2 500 kuna za hranu". Ako je veličina uzorka dovoljno velika, može se koristiti Centralni granični teorem koji će nam odgovarati na pitanja o diskretnom uzorku aproksimirajući ga normalnom razdiobom. Sljedeći teorem nam tvrdi da, razdioba slučajne varijable s n mjerenja teži normalnoj razdiobi, ako n teži u beskonačnost. Moivre je dokazao Centralni granični teorem, odnosno u posebnom slučaju tvrdio je da se povećanjem broja n Bernoulijevih pokusa, binomna razdioba približava normalnoj. Kako se veličina uzorka n povećava bez ograničenja, oblik razdiobe uzorka znači uzet iz populacije sa srednjom vrijednošću μ i standardnom devijacijom σ težit će normalnoj razdiobi s parametrima. Prosjek mjerenja u n ponavljanja će približno imati razdiobu sa srednjom vrijednosti μ i imat će standardnu devijaciju:

$$\frac{\sigma}{\sqrt{n}}.$$

Da bismo iskazali formalnu definiciju Centralnog graničnog teorema, potrebno je definirati nezavisnost slučajnih varijabli.

Definicija 3.4.1. Za diskretne slučajne varijable X i Y definirane na istom diskretnom vjerojatnosnom prostoru $(\Omega, \mathcal{A}, P)$ kažemo da su nezavisne ako vrijedi

$$P(X \leqslant x \ \cap \ Y \leqslant y) = P(X \leqslant x) \cdot P(Y \leqslant y).$$

Slijedi formalni iskaz Centralnog graničnog teorema.

Centralni granični teorem Neka je $X_n, \ n \in \mathbb{N}$, niz nezavisnih slučajnih varijabli kojima pripada ista vjerojatnosna razdioba s očekivanjem μ i varijancom σ^2. Za $n \in \mathbb{N}$ definiramo $S_n := X_1 + ... + X_n$. Tada za sve $x \in \mathbb{R}$ vrijedi:

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right) = \Phi(x)$$

Možemo reći i da niz funkcija razdiobe varijable $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, $n \in \mathbb{N}$, konvergira funkciji $\Phi(x)$, tj. funkciji razdiobe vjerojatnosti standardne normalne razdiobe $\mathcal{N}(0, \ 1)$.

Prilikom korištenja Centralnog graničnog teorema važno je zapamtiti dvije stvari:
1. Kad se izvorna varijabla normalno distribuira, razdioba uzorka će se normalno distribuirati, za bilo koju veličinu uzorka n.
2. Kada razdioba izvorne varijable možda nije normalna, veličina uzorka od 30 ili više je potrebna za upotrebu normalne razdiobe za približnu razdiobu uzorka. Što je uzorak veći, aproksimacija će biti bolja.

Centralni granični teorem objašnjava zašto srednja vrijednost teži k očekivanju razdiobe cijele populacije. Centralni granični teorem osigurava da je razdioba uzorka približno normalna, tj. da dobivamo poznatu zvonastu krivulju. Rad završavamo zadatkom iz svakodnevnog života, u kojem je prikazana primjena Centralnog graničnog teorema.

Zadatak
Kolika je vjerojatnost da među 1000 novorođene djece bude barem 490 dječaka, ako je za svako dijete jednako vjerojatno da bude muško ili žensko?

Definiramo događaj A = {novorođeno dijete je dječak}.
Označimo sa X broj novorođenih dječaka.
Uočimo da je u ovom primjeru broj pokusa jednak $n = 1000$, a vjerojatnost događaja A je

$p = 0.5$. Slijedi da X ima binomnu razdiobu s $n = 1000$ i $p = 0.5$.
Tražimo $P(X \geq 490)$.

$$P(X \geq 490) = P(490 \leqslant X \leqslant 1000).$$

S obzirom da bi postupak računanja bio dug, u ovom trenutku slijedi aproksimacija binomne razdiobe normalnom razdiobom, kod koje je jednostavno odrediti raspone varijable X unutar traženog područja. Slijedi:

$$P(490 \leqslant X \leqslant 1000) \approx \Phi\left(\frac{1000 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{490 - np}{\sqrt{npq}}\right) = [np = 500, npq = 250] =$$

$$\Phi\left(\frac{1000 - 500}{\sqrt{250}}\right) - \Phi\left(\frac{490 - 500}{\sqrt{250}}\right) = \Phi(31.63) - \Phi(-0.63) = \Phi(31.63) - (1 - \Phi(-0.63)) =$$

$$0.7357.$$

Vjerojatnost da od 1000 novorođene djece bude barem 490 dječaka je 73.57%.

Bibliografija

[1] *9 Real Life Examples Of Normal Distribution*, dostupno na: https://studiousguy.com/real-life-examples-normal-distribution/ (rujan 2021.)

[2] N. Adžaga, A. M. Špoljarić, N. Sandrić: *Vjerojatnost i statistika*, dostupno na: https://www.grad.unizg.hr/_download/repository/VIS%5B1%5D.pdf (rujan 2021.)

[3] S. Antoliš, A. Copić, E. Špalj, *Matematika za 1.razred srednje škole*, dostupno na https://edutorij.e-skole.hr/share/proxy/alfresco-noauth/edutorij/api/proxy-guest/af9b8682-eef4-478e-9b92-edcba4790886/index.html (rujan 2021.)

[4] R. Barone: The state of STEM education told through 18 stats, dostupno na: https://www.idtech.com/blog/stem-education-statistics (kolovoz 2021.)

[5] *Bean machine*, dostupno na: https://en.wikipedia.org/wiki/Bean_machine (rujan 2021.)

[6] M. Benšić, D. Grahovac, N. Šuvak, *STEDy - statistički edukacijski portal*, dostupno na: https://stedy.hr (rujan 2021.)

[7] M. Benšić, N. Šuvak, *Primijenjena statistika*, Sveučilište J.J. Strossmayera, Odjel za matematiku, 2013.

[8] A. G. Bluman, *Elementary Statistics: A Step by Step Approach*, dostupno na: https://ugess3.files.wordpress.com/2016/01/bluman-step-by-step-statistics-8th-edition.pdf (rujan 2021.)

[9] J. Chen, *Histogram*, dostupno na: https://www.investopedia.com/terms/h/histogram.asp (rujan 2021.)

[10] Core Maths Support Programme, dostupno na:
https://www.stem.org.uk/resources/elibrary/reurce (rujan 2021.)

[11] O. Craig, *What is STEM?*, dostupno na:
https://www.topuniversities.com/courses/engineering/what-stem
(rujan 2021.)

[12] *Cumulative distribution functions*, dostupno na: https://amsi.org.au/ESA_Senior_Years/SeniorTopic4/4e/4e_2content_2.html (rujan 2021.)

[13] A. Čižmešija, T. Soucie, N. Radović, R. Svedrec, *Igrajmo se s podatcima*, Zbornik radova IV. kongresa nastavnika matematike (P. Mladinić, R. Svedrec), Hrvatsko matematičko društvo i Školska knjiga, Zagreb, 2010., 127-142

[14] Državni zavod za statistiku, dostupno na: https://www.dzs.hr/ (rujan 2021.)

[15] D. Glasnović Gracin, *Uvođenje statistike u početno učenje matematike*, Zrno (2016), 122-123

[16] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, 2008.

[17] Histogram Uses in Daily Life, dostupno na:
https://studiousguy.com/histogram-uses-in-daily-life/ (rujan 2021.)

[18] M. Huzak, *Matematička statistika*, skripta za kolegij, dostupno na: https://web.math.pmf.unizg.hr/nastava/stat/index.php?sadrzaj=predavanja.php
(rujan 2021.)

[19] *Introductory Statistics*, dostupno na:
https://stats.libretexts.org/Bookshelves/Introductory_Statistics
(rujan 2021.)

[20] Ispitni katalozi za državnu maturu iz matematike, dostupno na:
https://www.ncvvo.hr/kategorija/drzavna-matura/ispitni-katalozi/
(rujan 2021.)

[21] P. Khuarna, *Uses of Histograms in Machine Learning*, dostupno na:
https://prvnk10.medium.com/uses-of-histograms-in-machine-learning-e209bad36cd2 (kolovoz 2021.)

[22] S. Loparić, *Vjerojatnost i statistika - zašto, kada i kako*, Poučak (2019), 45-51

[23] Ministarstvo znanosti i obrazovanja Republike Hrvatske, *Kurikulumi nastavnih predmeta Matematika za osnovne škole i gimnazije i Matematika za srednje strukovne škole na razini 4.2.*, Zagreb, 2019., dostupno na: https://skolazazivot.hr/wp-content/uploads/2020/07/MAT_kurikulum_1_71.pdf

[24] Ministarstvo znanosti i obrazovanja Republike Hrvatske, *Odluka o donošenju kurikuluma za nastavni predmet Matematike za osnovne i srednje škole u Republici Hrvatskoj*, Zagreb, 2019., dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2019_01_7_146.html (rujan 2021.)

[25] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, The MIT, Press, 2012.

[26] Europski parlament, Nacrt zakonodavne rezolucije Europskog parlamenta o Prijedlogu direktive Europskog parlamenta i Vijeća o primjerenim minimalnim plaćama u Europskoj uniji, dostupno na: https://www.europarl.europa.eu/doceo/document/EMPL-PR-689873_HR.pdf (srpanj 2021.)

[27] E. A. S. Pontes, *A Brief Historical Overview Of the Gaussian Curve: From Abraham De Moivre to Johann Carl Friedrich Gauss*, dostupno na: http://www.ijesi.org/papers/Vol(7)i6/Version-5/D0706052834.pdf (rujan 2021.)

[28] P. Sahlberg, *Lekcije iz Finske*, Školska knjiga, Zagreb, 2012.

[29] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 2002.

[30] E. Sekulić, *Prikazivanje podataka - linijski dijagram*, dostupno na: https://www.youtube.com/watch?v=s3KdpFFgWoM&t=960s (srpanj 2021.)

[31] Statistički ljetopis Republike Hrvatske, Zagreb, 2018. dostupno na: https://www.dzs.hr/Hrv_Eng/ljetopis/2018/sljh2018.pdf (srpanj 2021.)

[32] *statistika*, Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2021., dostupno na: http://www.enciklopedija.hr/Natuknica.aspx?ID=57896 (kolovoz 2021.)

[33] M. Thakur, *Standard Deviation Examples*, dostupno na: https://www.educba.com/standard-deviation-examples/ (kolovoz 2021.)

Sažetak

U diplomskom radu "Matematička statistika u nastavi matematike" cilj je bio dati pregled Statistike prema "novom" kurikulumu u osnovnim i srednjim školama. Fokus rada bio je na davanju ideja za dublje razumijevanje koncepata i povezivanje sa svakodnevnim životom. Prvenstveno je bit uvidjeti primjenu znanja statistike u aktualnim zanimanjima, ali i osvijestiti zašto je znanje statistike bitno za budućnost. U prvom dijelu rada čitatelj može pročitati koji se statistički koncepti uče u obrazovnim sistemima drugih zemljama te su detaljno analizirani ishodi kurikuluma iz statistike po razredima. U drugom dijelu rada tema je deskriptivna statistika. Mjere srednjih vrijednosti i mjere raspršenosti obrađene su tako da su uz definicije prezentirane aktivnosti kroz koje će učenici primjenjivati koncepte te uvidjeti primjenu znanja u ekonomiji, trgovanju dionicama i trgovanju nekretninama. Posebno se obrađuje histogram. U trećem dijelu rada postupno je prezentirano kako na metodički način učenicima definirati pojam slučajne varijable. U radu su detaljno obrađene diskretne i kontinuirane slučajne varijable te su dani prijedlozi aktivnosti za primjenu koncepata. Dani su primjeri u kojima je prikazano kako se računaju očekivanje, varijanca i standardna devijacija za neprekidnu i kontinuiranu slučajnu varijablu. Posebno je obrađena najpoznatija razdioba podataka - normalna razdioba: od povijesnog pregleda preko uvođenja normalne razdiobe u nastavi, formalne definicije i svojstava sve do primjene. Rad završava Centralnim graničnim teoremom, jednim od najvažnijih teorema iz statistike, kojim se tvrdi da razdioba slučajne varijable s n mjerenja teži normalnoj razdiobi ako n teži u beskonačnost. Nadamo se da se čitanjem ovog diplomskog rada čitatelji mogu upoznati s osnovnim statističkim pojmovima potrebnim za dublju analizu podataka te dobiti motivaciju za daljnje istraživanje i proučavanje statistike u svakodnevnom životu.

Summary

In graduate thesis "Matematička statistika u nastavi matematike" the goal was to give an overview of statistics by the new curriculum in elementary and high schools. Focus was on giving ideas for deeper understanding of concepts and linkage with everyday life. Primarily the goal is to see application of statistics knowledge in different professions, but also to become aware why statistics knowledge is important for the future. In first part of the thesis the reader can see which statistical concepts are studied in educational systems in other countries where outcomes of curriculum by each class is analyzed in detail. In the second part of the thesis descriptive statistics can be found. Measures of mean values and measures of dispersion are done in a way that activities are presented along the definitions, where students can apply concepts and see application in economy, stock trading and real estate. Special focus is on histogram. In the third part of the thesis students are gradually presented how to methodically define random variables. In thesis discrete and continuous random variables are discussed in most detail, as well as suggestions of activities are given to apply concepts. Examples are given how to calculate expectation, variance and standard deviation for nonstop and continuous random variable. Special focus is on the most famous data distribution - normal distribution: from historical overview to the introduction of normal distribution in class teaching, formal definitions and properties to the application. Thesis ends with The Central Limit Theorem, one of the most important theorems of statistics, which claims that distribution of random variables with n measurement tends to normal distribution if n tends to infinity. We hope that by reading this graduate thesis readers can familiarize with basic statistic terms required for deeper analysis and find motivation for further research and study of statistics in everyday life.

Životopis

Zovem se Anja Stanković. Rođena sam 27.travnja 1992. u Bjelovaru, na dan kada je s emitiranjem počela emisija "Dobro jutro, Hrvatska". Završila sam Drugu osnovnu školu u Bjelovaru i Gimnaziju Bjelovar, opći smjer, s odličnim uspjehom. Matematiku sam zavoljela kao natjecateljica na općinskim, županijskim i regionalnim natjecanjima u osnovnoj školi. Odmalena me privlačio nastavnički poziv. U srednjoj školi pjevala sam u zboru, glumila sam u Bjelovarskom kazalištu i rekreativno se bavila sportom. Godine 2011. kao prvi izbor upisujem preddiplomski studij matematike, nastavnički smjer na Prirodoslovno - matematičkom fakultetu u Zagrebu. 2016. godine odlazim na studentsku ERASMUS razmjenu u Beč i završavam preddiplomski studij. Nakon prosvjeda za Kurikularnu reformu, odlučujem se zaposliti u osnovnoj školi. Dvije godine radila sam u Bjelovaru kao učiteljica matematike u Drugoj i Trećoj osnovnoj školi. Za vrijeme rada u školi prošla sam 23 edukacije iz područja odgoja i obrazovanja, a stručno sam se usavršavala izvan Republike Hrvatske. Godine 2018. odlučujem upisati diplomski studij matematike, smjer: nastavnički na Prirodoslovno - matematičkom fakultetu u Zagrebu. Za vrijeme studija bavila sam se jogom, plesom, trčanjem. Bila sam članica europske studentske agencije AEGEE, sudjelovala sam u radu udruge RADDAR i Centru izvrsnosti za matematiku Bjelovar. Tijekom svojih 20-ih zalagala sam se za Cjelovitu kurikularnu reformu i pozitivne promjene u školstvu, a od prve godine fakulteta proučavam obrazovnu politiku sa željom za promjenom. U rujnu 2021. branim diplomski rad pod nazivom "Matematička statistika u nastavi matematike", kojeg sam pisala u ljeto 2021. godine. Iako se nadam da ću uskoro iskusiti i rad u privatnom sektoru, od 1. rujna 2021. zaposlena sam kao srednjoškolska profesorica Matematike u zagrebačkoj X. gimnaziji "Ivan Supek".