

Težave pri obdelavi podatkov in statističnem preiskovanju

dr. Andreja Drobnič Vidic
Univerza v Ljubljani, Fakulteta za matematiko in fiziko

Izvleček

Statistične vsebine, s katerimi se mladi v vzgojno-izobraževalnem procesu srečajo že v vrtcu in jih pogosto srečujejo v vsakdanjem življenju, učiteljem matematike pri poučevanju povzročajo nemalo težav. Poleg nizkega števila ur, namenjenih tem vsebinam, se soočajo z neenotnostjo poimenovanj in definicij ter nelagodjem ob premalo poznanih empiričnih preiskavah, v katerih naj bi mladi uporabili pridobljeno statistično znanje. Namen prispevka je ob navedbi strukturirane in čim bolj učinkovite obravnave statističnih vsebin v gimnaziji izpostaviti težave, na katere lahko naletijo učitelji in dijaki, in jih odpraviti. Z njim želimo vzpostaviti vez med statističnimi vsebinami v osnovni in srednji šoli ter pomagati učiteljem statistične vsebine jasno in čim učinkoviteje predstaviti mladim.

Ključne besede: statistika, empirična preiskava, učni načrt, težave pri poučevanju in učenju

Problems with Data Processing and Statistical Investigation

Abstract

Statistical contents, which young people encounter in the educational process as early as kindergarten and which they often come across in their daily lives, cause quite a few problems to mathematics teachers in their teaching practice. Besides the small number of hours devoted to such contents, teachers also deal with inconsistent names and definitions, and with unease due to a lack of knowledge about empirical investigations in which young people are to apply their acquired statistical knowledge. The purpose of the article is to provide a structured and maximally effective discussion of statistical contents in grammar school, while highlighting potential problems that teachers and secondary school students may face and eliminating them. The aim is to establish a link between statistical contents in primary and secondary schools, and help teachers to introduce them to their students as clearly and effectively as possible.

Keywords: statistics, empirical investigation, curriculum, problems with teaching and learning

1 Uvod

Od matematičnih vsebin je statistika posebna, saj se v visokošolskem izobraževanju pogosto »odcepi« od matematike in jo obravnavamo samostojno. Njeni temelji so oblikovani z elementi verjetnosti, vendar je sčasoma postala tako vsestransko uporabna in tako hitro razvijajoča se veda, da pogosto seže le v uporabo in se tako uči samostojno (brez matematičnih temeljev). Tudi način poučevanja se pogosto razlikuje od običajnih metod pri čisto matematičnih vsebinah, kar prikazuje Preglednica 1.

Razlike se odražajo tudi v zahtevah, izpostavljenih procesih učenja, in v zastavljenih nalogah oziroma problemih.

Za vsebinski sklop Statistika je po učnem načrtu za matematiko za program gimnazija predvidenih 10 ur in priporočajo obravnavo v 1. letniku; vsebine za statistiko v gimnaziji so naslednje (Žakelj in sod., 2008):

Preglednica 1: Razlike med učenjem matematičnih in statističnih vsebin.

MATEMATIKA:	STATISTIKA:
doslednost	preglednost
natančno podajanje	povzemanje
trditve – prepričanost	domneve – zanesljivost
dokazovanje	posploševanje
problem – ena ali več poti – rešitev	problem – več poti – več (odprtih) rešitev
idealizacija	realizacija

- Osnovni statistični pojmi;
- Vrste podatkov;
- Zbiranje podatkov;
- Urejanje in strukturiranje podatkov;
- Prikazovanje podatkov (stolpčni, pozicijski, tortni diagram, histogram, razsevni diagram, linijski in krivuljni diagram, škatla z brki);
- Aritmetična sredina, mediana, modus;
- Variacijski razmik, standardni odklon, medčetrtnski razmik;
- Statistična naloga.

Dijaki naj bi samostojno izvedli statistično nalogo tako, da bi obravnavali realističen problem, povezali pri matematiki pridobljeno znanje o osnovah statistike z drugimi predmeti in ga morda vključili v projektni teden ter rešili z uporabo digitalne tehnologije.

Kot bomo spoznali v prispevku, učitelji pri poučevanju statističnih vsebin naletijo na nemalo težav. Poleg nizkega števila ur, namenjenih tem vsebinam, naletijo na težave v neenotnosti poimenovanj in definicij ter v premalo poznani izvedbi statistične naloge za uvajanje v realno empirično preiskavo, v kateri uporabimo pridobljeno statistično znanje.

Izpostavili bomo, da je večina zapisanih vsebin že del osnovnošolskega učnega načrta, tako da v gimnaziji nekatere vsebine lahko nekoliko hitreje ponovimo in strukturirano predstavimo dijakom ter tako pridobimo čas. V prispevku bomo ob navedbi strukturirane obravnave vsebin izpostavili težave, na katere lahko naletijo učitelji in dijaki in jih seveda skušali odpraviti. Za kakovostno izvedbo empirične preiskave je sicer potrebno tudi znanje, pridobljeno pri matematiki ob koncu srednješolskega izobraževanja, ko dijaki spoznajo osnove verjetnostnega računa. Toda »kvazi« empirično preiskavo lahko dijaki v obliki statistične naloge izvedejo že v 1. letniku, ko običajno obravnavamo zapisane statistične vsebine. S prispevkom želimo vzpostaviti vez med statističnimi vsebinami v osnovni in srednji šoli ter pomagati učiteljem statistiko jasno in čim učinkoviteje predstaviti mladim.

2 Statistično predznanje

O obdelavi podatkov in verjetnosti govorijo že v vrtcu, saj kurikulum za vrtce omenja, da otrok v vrtcu napoveduje rezultat (ali bo na sprehodu opazil kakšno lužo ali ne), pridobiva izkušnje o tem, kaj se pogosto in kaj se vedno zgodi (pogosto je oblačno, voda je vedno mokra), razporeja predmete v skupine po eni ali dveh lastnostih in razporeja skupine predmetov (Bahovec, Bregar, Čas in sod., 1999).

V osnovni šoli do konca 6. razreda učenci sistematično zbirajo podatke (beleženje štetja in meritev), strukturirajo podatke (urejanje podatkov po velikosti, razvrščanje), razporejajo podatke v smiselne skupine po enem ali dveh kriterijih, se seznanijo s predstavitvijo podatkov (npr. s preglednico, prikazi), spoznajo računalniške preglednice in znanje o obdelavi podatkov uporabijo v krajši preiskavi. Nato znanje sistematično predstavijo v zadnji triadi (Kmetič in Sirnik, 2010, str. 233):

7. razred: Zbiranje, urejanje in predstavitev podatkov: krožni diagram, razsevni diagram, črtni (linijski) diagram, empirična preiskava; Merila za sredino in razpršenost: aritmetična sredina.

8. razred: Zbiranje, urejanje in predstavitev podatkov: grafi (odvisnost diskretnih spremenljivk z grafi, odvisnost zveznih spremenljivk z grafi), empirična preiskava; Merila za sredino in razpršenost: pomen aritmetične sredine.

9. razred: Zbiranje, urejanje in predstavitev podatkov: vprašalniki, uporaba (računalniških) orodij, empirična preiskava; Merila za sredino in razpršenost: pomen aritmetične sredine, modus, mediana, škatla z brki; Izkušnje s slučajnimi dogodki: poskus, dogodek (nemogoč, gotov, slučajni), izid, verjetnost dogodka na osnovi poskusa (statistična verjetnost); Empirična preiskava: načrtovanje, zbiranje podatkov, urejanje in analiziranje, interpretacija, predstavitev.

Podoben učni načrt imajo tudi v Italiji, Veliki Britaniji in ponekod drugod po Evropi. Doseganje ciljev učnega načrta v slovenskih osnovnih šolah preverjamo enotno z nacionalnimi preverjanji znanja. Pregled nalog iz Nacionalnega preverjanja znanja (NPZ) iz matematike v 9. razredu kaže, da je na preverjanju vselej vsaj ena naloga iz obdelave podatkov, naloge iz osnovnošolskih statističnih vsebin pa so z leti vse bolj zahtevne (Državni izpitni center, NPZ). Največjo vlogo pri preverjanju ima branje podatkov iz preglednic, stolpčnih prikazov, linijskih diagramov ali krožnih diagramov in njihovo pretvarjanje iz ene oblike v drugo. Tako znanje zahteva prav vsaka statistična naloga na NPZ. Naloge v zadnjih letih zahtevajo tudi izračun aritmetične sredine, določitev modusa in mediane, nekaj nalog pa je tudi iz verjetnosti.

Za zgled bomo podali dve nalogi iz NPZ, eno iz leta 2006 in drugo iz leta 2021 ter preiskovalno nalogo za osnovnošolce. Prek njih bomo nakazali naraščanje zahtev po znanju statistike in težave, s katerimi se v praksi srečujejo tako učenci kot učitelji.

2.1 Naloga na NPZ iz leta 2006

Naloga iz obdelave podatkov je predstavljena na Sliki 1.

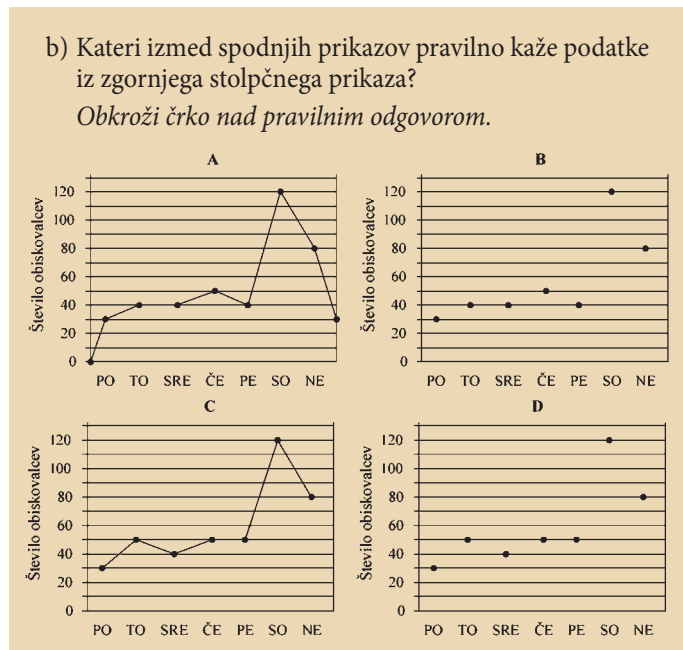


Slika 1: Naloga o obdelavi podatkov na NPZ leta 2006.

Prvemu vprašanju sledi še vprašanje na Sliki 2:

2.2 Naloga na NPZ iz leta 2021

Naloga iz obdelave podatkov v šolskem letu 2021 zajema podatke iz realnega sveta:



Slika 2: Možni prikazi za dani stolpčni prikaz na NPZ leta 2006.

Prvo vprašanje je zelo enostavno, drugo pa z nanizanimi štirimi prikazi na Sliki 2 kar malo zavajajoče. Najprej učenec sploh ne ve, na kaj naj bo pri prikazih pozoren. Šele daljše opazovanje privede do spoznanja, da prikaza C in D odpadeta, saj za torek ni enakega števila obiskovalcev kot je navedeno v stolpčnem prikazu. Izbira prikazov med A in B zopet ni posrečena. Frekvence podatkov pogosto prikazujemo z linijskim diagramom (frekvenčnim poligonom), kot bomo spoznali v nadaljevanju. Linijski diagram je lomljena črta med točkami, katerih prva koordinata predstavlja vrednosti spremenljivke, druga pa njihove frekvence. Tako bi bil sicer primeren graf A, vendar je lomljena črta na levi speljana do točke (0,0), kar je v mnogih statističnih knjigah tudi svetovano storiti (Košmelj, 2008), a na desni strani ni storjeno enako. Kot pravilen je podan graf B, četudi predstavitve z razsevnim diagramom pri eni spremenljivki ni običajna.

V nalogi je dodano še vprašanje:

- c) Ob sobotah in nedeljah je cena vstopnice 1000 SIT, druge dni pa 800 SIT. Kateri tortni prikaz kaže primerjavo skupnega zasluga v soboto in nedeljo s skupnim zaslužkom preostalih dni v tednu?

Učenci izbirajo med štirimi tortnimi prikazi. To vprašanje zahteva tudi izračun skupnega zasluga med tednom in čez vikend.

Naloga 7 iz obdelave podatkov na NPZ iz leta 2006 torej zahteva branje, preoblikovanje in interpretacijo prikazov, zadnje vprašanje pa tudi malo računanja.

Slovensko reprezentanco so na Evropskem prvenstvu v košarki 2017 zastopali igralci, ki so zapisani v spodnji preglednici (Slika 3).



Številka na dresu	Ime in priimek	Igralno mesto	Višina v cm	Starost v letih
0	Anthony Randolph	krilni center	211	28
1	Matic Rebec	branilec	180	22
3	Goran Dragić	branilec	194	31
6	Aleksej Nikolić	branilec	191	22
7	Klemen Prepelič	branilec	191	25
8	Edo Murić	krilo	202	26
11	Jaka Blažič	branilec	196	27
14	Gašper Vidmar	center	211	30
17	Saša Zagorac	krilni center	206	33
22	Žiga Dimec	center	211	24
31	Vlatko Čančar	krilo	203	20
77	Luka Dončić	branilec	201	18

Slika 3: Podatki ob nalogi iz obdelave podatkov na NPZ v letu 2021.

9. a) Kolikšna je bila mediana starosti igralcev slovenske reprezentance na Evropskem prvenstvu v košarki 2017? Reševanje: Mediana je _____ let. (2 točki)
9. b) Igralci slovenske reprezentance, ki so imeli na dresu eno izmed števil 0, 3, 11, 14 in 77, so tekmo začeli. Modus višine igralcev, ki so začeli tekmo, je _____ cm. (1 točka)
9. c) Eden izmed snemalcev tekme je kamero naključno usmeril v enega izmed igralcev slovenske reprezentance, ki so začeli tekmo. Verjetnost, da je kamero usmeril v branilca, je _____. (1 točka)
9. d) Povprečna višina igralcev slovenske reprezentance, ki so začeli tekmo, je _____ cm. (1 točka)
9. e) Edo Murić je v 5. minuti zamenjal igralca, ki je začel tekmo. Zaradi tega se je povprečna višina slovenskih košarkarjev na igrišču spremenila na 203,8 centimetra. Zapiši ime in priimek igralca, ki ga je Edo Murić zamenjal. _____ (1 točka)

Naloga iz obdelave podatkov na NPZ iz leta 2021 zahteva branje in interpretacijo prikaza, ki vključuje različne vrste spremenljivk v preglednici, računanje mer sredine (osredinjenosti ali srednje vrednosti) in računanje verjetnosti, zadnje vprašanje pa tudi nemalo računskih spretnosti.

Primerjava nalog iz vsebin statistike na NPZ iz prvega in zadnjega leta spletne dostopne verzije NPZ do tega prispevka kaže na porast zahtev statističnega znanja. Obravnavani nalogi ravno tako kažeta, da se moramo učitelji za poučevanje statističnih vsebin nenehno izobraževati, da postopoma delamo manj napak.

2.3 Empirična preiskava

Pri empiričnih preiskavah realnega sveta pridejo v ospredje značilnosti statistike, navedene v Tabeli 1: preglednost, povzemanje, domnevanje, posploševanje in realizacija. Kot učitelji matematike morda nismo vajeni pri učenju izpostavljati teh značilnosti, zato nemalokrat delamo napake (potemtakem moramo biti še toliko bolj strpni do učencev). Na primeru opisane osnovnošolske mini raziskave bomo izpostavili značilnosti pri poučevanju statistike in nekatere pogoste napake, ki pa so sestavni del učenja. Primer take mini raziskave z naslovom *Najbolj priljubljen šport na šoli* je podan v priročniku za učitelje (Kmetič in Sirnik, 2010, str. 245).

Šola organizira pet vrst športnih dejavnosti. Razišči, kateri šport je med učenci najbolj priljubljen.

Podan je tudi postopek reševanja:

Razmislimo, kaj vse lahko vpliva na priljubljenost posameznega športa, nato pa se odločimo, s kakšnimi vprašanji bomo raziskali dani izziv ... Pojavi se veliko vprašanj. Na vsa ne bomo mogli in znali odgovoriti. Odločimo se za eno vprašanje, in sicer: Koliko učencev trenira posamezni šport? in v sklopu dobljenih rezultatov na koncu poskušamo komentirati izsledke.

Pri empirični preiskavi, ki je raziskava realnega sveta, je podobno kot v fiziki treba pretehtati, kateri podatki so za reševanje problema ključni, katere bomo zanemarili. Pri tem pa moramo biti previdni. Že res, da je v dani raziskavi priljubljenost športa na šoli povezana s številom učencev, ki se danega športa na šoli udeležujejo, vendar je udeležba pri športu odvisna še od mnogih drugih dejavnikov: od danega termina vadbe, od lokacije, od nevarnosti športa in podobno in ne le od priljubljenosti. Pri predlagani **poenostavitvi**, da priljubljenost merimo s številom učencev, ki se danega športa udeležujejo, se je zato treba vsaj zavedati, da vprašanje o številu učencev, ki trenirajo posamezni šport, ne more neposredno odgovoriti na vprašanje, kateri šport je na šoli najbolj priljubljen. Tako lahko učencem že ob začetku raziskave pojasnimo, da ne bomo iskali najbolj priljubljenega športa na šoli, ampak šport, ki ga izbere največ učencev. Tako smo na podlagi mnogih lastnosti, ki vplivajo na priljubljenost športa, **povzeli** eno lastnost – udeležba, ki jo bomo lahko kakovostneje

raziskali. Seveda bi lahko ostali pri vprašanju učencem: *Razišči, kateri šport je med učenci najbolj priljubljen*, ki je vsekakor **ena od možnih poti reševanja problema** oziroma preiskovanja in bi prinesla lahko **drugačen** odgovor.

Sledi pridobivanje podatkov. Kako pridobiti verodostojne podatke? V Kmetič in Sirnik (2010, str. 245) navajajo, da podatke lahko zberemo z vprašalnikom: *Z anketo bomo zbrali, koliko učencev trenira posamezni šport. Med učenci, ki se ne udeležujejo nobenega športa, ne bomo delali ankete. Če se učenec ukvarja z več športi, se v anketi opredeli le za enega, najljubšega.* V tej mini raziskavi opozorijo, da med učenci, ki ne obiskujejo nobenega športa, ne izvedemo ankete, torej prvega vprašanja o tem, ali učenec sploh kaj trenira, po dogovoru ne štejemo v anketo. Ne ukvarjajo se s pomembnim vprašanjem, ali uspešno anketirati vse učence, kajti taka **realizacija** je zelo malo verjetna. Tako v njej ni **posploševanja** iz vzorca na celotno populacijo, ki je sicer sestavni del statistične preiskave, saj za posploševanje osnovnošolci nimajo dovolj znanja.

Po zbranih podatkih in izračunih moramo pri interpretaciji raziskave učence navajati na natančnost izražanja pri postavljanju **domnev**. V podani preiskavi je opredeljeno, naj bi odgovarjali le učenci, ki trenirajo vsaj en šport. V primeru, da trenirajo več športov, naj bi se opredelili za najljubšega. Podani so rezultati ankete: košarka 30, namizni tenis 15, nogomet 20, ples 15 in plavanje 10 in interpretacija rezultatov: *Največ učencev trenira košarko*. Ta domneva ni nujno pravilna, ker so učenci, ki trenirajo na šoli več športov, navedli najljubšega. Lahko bi se namreč zgodilo, da bi poleg navedenega športa še 11 učencev treniralo nogomet, vendar ta ni njihov najljubši šport. V primeru, da na primer košarke ne trenira nihče drug od navedenih 30, bi največ učencev treniralo nogomet – kar 31 in v tem primeru bi bil podani odgovor napačen. Bolj **zanesljiva** je trditev: *Največ učencev je za najljubši šport, ki ga trenirajo, navedlo košarko*. Pri tem privzamemo, da je v primeru obiskovanja enega športa ta hkrati najbolj priljubljen. Še bolje je, če zaključimo: *Rezultati so pokazali, da je med športi na šoli, ki se jih učenci udeležujejo, največ učencev navedlo košarko*.

Osnovnošolski učitelji naj opozarjajo na težave, se z učenci pogovarjajo in povedo, da so tudi opažanja teh težav del statistične preiskovanja.

3 Obdelava podatkov v gimnaziji in z njimi povezane težave

Z **namenom učinkovite razlage** zahtevanih statističnih pojmov v srednji šoli smo v sklopu pouka na daljavo predstavili **statistične vsebine preko enega samega Primera 1**, z njim pa tudi izvedli zgled statistične naloge, ki predstavlja uvod v empirično preiskovanje.

Primer 1 izzove pri dijakih znanje, ki so ga pridobili v osnovni šoli in je dalje podlaga za bazo podatkov in nove naloge, ki jih uporabimo pri učenju statističnih vsebin, vse to najdete v spletni učilnici Zavoda Republike Slovenije za šolstvo, namenjeni podpori učiteljem matematike v gimnazijah v sklopu poučevanja na daljavo (Drobnič Vidic, 2020): <https://skupnost.sio.si/course/view.php?id=65§ion=1>. Ker je v nekaterih srednješolskih

Primer 1: Učitelj gimnazije Bič se v šolskem letu 2020/2021 odloči dati anketni vprašalnik dijakom 1. b oddelka, ko so bili v februarju vsi pri pouku. Dijaki so dobili naslednji vprašalnik, ki so ga izpolnili.

Obkrožite po en ponujeni odgovor ali odgovor napišite na črto:

- Vaš spol: moški ženski
- Kako priljubljen predmet vam je matematika v gimnaziji:

1: Sploh ne maram. 2: Ne maram preveč.

3: Imam še kar rad. 4: Imam zelo rad.

- Število desetic doseženih odstotkov pri merjenju znanja matematike na zadnjem testu v tem šolskem letu (pri 38 % obkrožite 3):

0 1 2 3 4 5
6 7 8 9 10

Številka dijaka	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Spol	ž	ž	ž	m	ž	m	m	m	ž	m	ž	ž	ž	m	m	ž	m	m	ž	m	ž	m	ž	ž	m	m	m	m	m	ž	m	m
Priljubljenost	4	3	2	3	3	2	4	2	3	2	4	2	2	2	2	3	2	2	1	2	3	1	3	3	3	3	1	1	2	3	1	3
Dosežki	9	7	4	8	5	7	9	3	8	8	9	6	4	5	0	6	7	5	0	4	7	6	8	7	6	7	7	4	6	9	5	6
Aktivnost	13	10	7	4	5	4	4	3	11	8	8	7	7	6	2	5	4	3	3	2,5	3	2	2	6	2	2,5	1	0	2	1	0	2

- Čas, ki ga v tem letu običajno namenite domačemu delu pri matematiki tedensko: _____ ur.«

Učitelj je od dijakov dobil odgovore, ki jih prikazuje preglednica 1.

Nato je dijakom podal vprašanja, na katera so odgovorili!

- a) Določite delež moških in žensk, ki so odgovorili na vprašalnik.
- b) Kateri odgovor se pri vprašanju o priljubljenosti matematike največkrat pojavi?
- c) Zapišite krajšo tabelo, v kateri so predstavljeni dosežki dijakov na zadnjem testu.
- d) Izračunajte povprečno število ur, ki jih dijaki iz 1. b običajno namenijo tedenskemu delu za matematiko.

učbenikih razlaga statistične vsebine pomanjkljiva, določeni izračuni obravnavanih pojmov pa celo napačni, podamo tudi razlago pomembnih pojmov in definicij, zahtevanih z učnim načrtom za matematiko, ki jih razvrstimo v 4 skupine:

- (1) Osnovni pojmi statistike pri zbiranju podatkov,
- (2) Urejenost in prikazovanje podatkov,
- (3) Mere srednje vrednosti,
- (4) Mere razpršenosti.

Natančne definicije in razlage najdete na omenjenem spletnem mestu, v prispevku pa se osredotočimo na težave, ki pestijo tako učitelje kot učence.

3. 1 Osnovni pojmi statistike pri zbiranju podatkov

Populacija je končna ali neskončna množica, opredeljena stvarno, krajevno in časovno. To pomeni, da pri opisu populacije znamo odgovoriti na vprašanje: koga ali kaj proučujemo, kje proučujemo in kdaj. V matematiki ponavadi množico sestavljajo elementi (če ni prazna), v statistiki pa populacijo sestavljajo (statistične) enote. Če iz populacije izberemo določeno število enot, dobimo vzorec. Na populaciji ali vzorcu proučujemo določene lastnosti, ki se od enote do enote lahko spreminjajo. Imenujemo jih **statistične spremenljivke** ali kratko spremenljivke. Za izbrano spremenljivko vsaki enoti določimo vrednost spremenljivke. Glede na to, kako opisujemo vrednosti spremenljivk, ločimo **opisne** spremenljivke

in številske spremenljivke. Vrednosti opisnih spremenljivk izražamo z opisi, na primer z besedami, kot opisujemo na primer spol ali stopnjo priljubljenosti predmeta matematika v gimnaziji (Sploh ne maram., Ne maram preveč., Imam še kar rad., Imam zelo rad.), vrednosti številske spremenljivke zapišemo s števili. Številske spremenljivke so lahko **diskretne** ali **zvezne**. Diskretna spremenljivka lahko zavzame le končno ali števno neskončno mnogo vrednosti (ki lahko tvorijo zaporedje). Navadno te spremenljivke opisujejo lastnosti, ki jih dobimo s preštevanjem, kot je na primer število desetic ipd. Za zvezne spremenljivke velja, da lahko zavzamejo katerokoli vrednost iz določenega intervala. Take spremenljivke navadno dobimo z merjenjem ali izračunavanjem. Zvezna je na primer spremenljivka v Primeru 1, ki meri čas, ki ga dijak nameni domačemu delu pri matematiki tedensko (in bi bil lahko teoretično 15 min = 0,25 ure oziroma katerakoli pozitivna vrednost z danega intervala). Glede na urejenost pa ločimo naslednje 4 vrste spremenljivk:

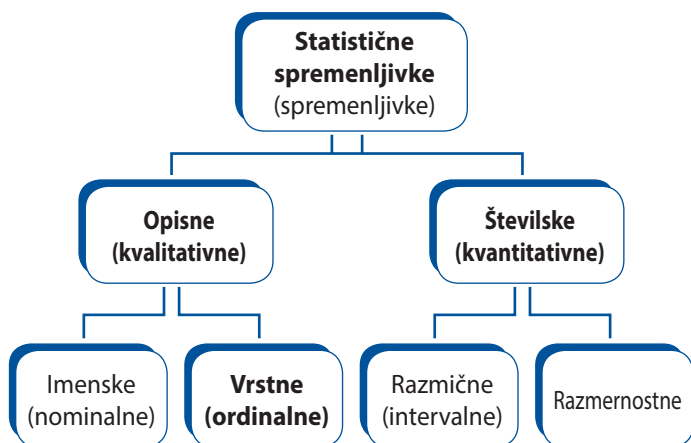
- (1) **imenske ali nominalne:** kjer ni urejenosti, ločimo le enote z različnimi vrednostmi (spol, naziv cigaret, naziv igralnega mesta v košarki, šolski predmeti ...);
- (2) **vrstne ali ordinalne:** kjer vrednosti lahko razvrstimo po nekem vrstnem redu (stopnja izobrazbe, stopnja zadovoljstva z danim predmetom, zaključni uspeh ...);
- (3) **razmične ali intervalne:** izražene številsko, kjer je poljubni dve vrednosti smiselno odštevati, ne pa deliti, saj vrednosti nimajo pomenske ničle (leto rojstva (0 ni pomenska, saj ne

pomeni, da leta rojstva ni ali da oseba ni bila rojena), temperatura v stopinjah Celzija, rezultati testov znanja (število desetic na testu znanja) ...);

- (4) **razmernostne**: izražene številsko, kjer je vrednosti spremenljivke smiselno odštevati in deliti in imajo pomensko ničlo (čas, namenjen matematiki v povprečju na teden, telesna višina, čas reševanja testa, dolžina proizvoda ...).

Urejenost narašča od imenske do razmernostne spremenljivke. Imenske in vrstne spremenljivke so navadno opisne spremenljivke, razmične in razmernostne spremenljivke pa so številske. Vrsta spremenljivke vpliva tako na prikaz podatkov kot na statistične izračune.

V srednješolskih učbenikih in priročnikih zasledimo **različne delitve vrst spremenljivk**. Vrste spremenljivk danih s podatki delimo na opisne (atributivne) in številske (numerične) glede na obliko podajanja vrednosti, torej z opisi ali s številkami, je zapisano v (Felda in sod., 2021; Kmetič in Sirnik, 2010: str. 255); po učnem načrtu za matematiko za program gimnazija dijaki razlikujejo med opisnimi ali kvalitativnimi podatki, vrstnimi ali ordinalnimi ter številkami ali kvantitativnimi podatki (Žakelj in sod., 2008: str. 37). Štiri skupine glede na urejenost so podane v (Kmetič in Sirnik, 2010: str. 255; Košmelj, 2008; Korenjak Černe, 2007). Razne delitve povzemimo s prikazom na Sliki 4.



Slika 4: Različne delitve vrst spremenljivk v srednješolskih učbenikih.

Včasih se zgodi, da zaradi dolgih opisov spremenljivk le-te nadomestimo s številkami znaki (na primer učni uspeh: odličen: 5, prav dober: 4, dober: 3, zadosten: 2, nezadosten: 1). Taka spremenljivka je še vedno opisna, saj razlike med uspehom niso enake, glede na stopnjo urejenosti je to še vedno vrstna spremenljivka.

3.2 Urejenost in prikazovanje podatkov

Vrednosti spremenljivk v vzorcu (ali populaciji) navadno uredimo in podatke organiziramo tako, da so čim bolj pregledni. Urejanje in prikazovanje podatkov je odvisno od vrste spremenljivke, ki ji podatki pripadajo ter od števila in raznolikosti samih podatkov.

Vrednosti imenske spremenljivke (1), ki so navadno podane opisno, uredimo tako, da enote v populaciji (ali vzorcu) velikosti n z dano enako vrednostjo (oziroma z enakim opisom) preštejemo. Število enot, ki ustreza dani vrednosti x_i ($i = 1, \dots, k$ za k različnih vrednosti) imenujemo frekvenca (tudi absolutna frekvenca)

z oznako f_i . Podatke take spremenljivke uredimo v **frekvenčno tabelo**, za večjo preglednost pa jih lahko grafično prikažemo.

V šolski praksi zasledimo **neenotno poimenovanje prikazov**. Učni načrt za matematiko za program gimnazija (Žakelj in sod., 2008), priročnik za učitelje (Kmetič in Sirnik, 2010), statistični terminološki slovar (Košmelj in sod., 2002) in uveljavljeni univerzitetni učbenik (Košmelj, 2008) se močno razlikujejo v poimenovanjih prikazov, zahtevanih v učnem načrtu za matematiko (izpustimo pozicijski diagram, ki se uporablja večinoma na osnovnošolskem nivoju):

Preglednica 2: Različna poimenovanja prikazov iz učnega načrta za gimnazije v slovenskem jeziku.

Učni načrt za matematiko	Priročnik za učitelje	Statistični slovar	Visokošolski učbenik
Tortni diagram, tortni prikaz	Krožni prikaz, strukturni krog	Strukturni krog	Strukturni krog
Stolpčni diagram, stolpčni prikaz	Diagram s stolpci	Prikaz s stolpci	Razrezani strukturni stolpec
Linijski diagram	Frekvenčni poligon, linijski poligon	Frekvenčni poligon	Poligon, linijski grafikon
Histogram	Histogram	Histogram	Histogram
Škatla z brki	Škatla z brki	Grafikon kvantilov	Okvir z ročaji
Razsevni diagram, razsevni prikaz	Razsevni diagram, razsevni prikaz	Razsevni diagram, korelacijski grafikon	Razsevni grafikon

Tudi učbeniki si v poimenovanjih niso enotni, le ime histogram je enotno. Različna poimenovanja niso problematična, a dijaki več novih besed za pojme lažje pomešajo. Pri definicijah lahko navedemo več imen, vendar se pri obravnavi in uporabi opredelimo le za eno. V prispevku smo se pri opisih v nadaljevanju

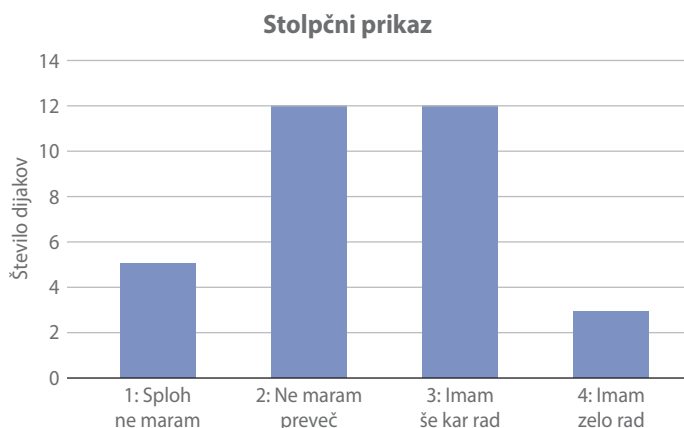


Slika 5: Tortni prikaz za spremenljivko Spol iz Primera 1.

držali imen, zapisanih v gimnazijskem učnem načrtu, razen za linijski diagram smo uporabili ime frekvenčni poligon.

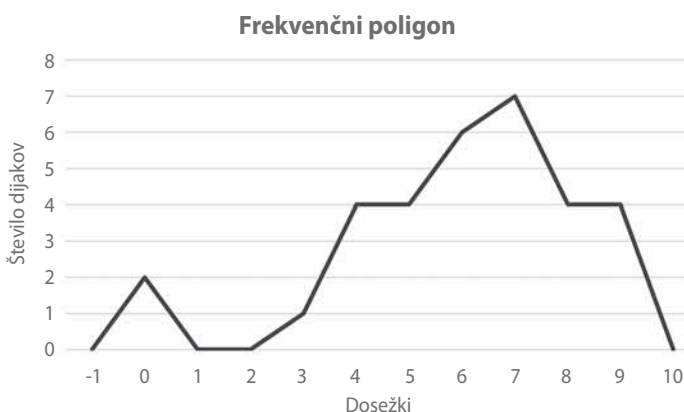
Za grafični prikaz imenskih spremenljivk navadno uporabimo **tortni prikaz**, ki je krog, razdeljen na izseke, katerih ploščine ustrezajo posameznim zastopanim deležem (Slika 5). Krog poudarja, da vrednosti spremenljivk niso nujno razvrščene po vrstnem redu, kar lahko storimo z vrednostmi vrstnih spremenljivk (2) in drugih bolj urejenih v urejenostni lestvici.

Pri bolj urejenih spremenljivkah vrednosti lahko razporedimo v ranžirno vrsto po danem vrstnem redu. Pri tem navedemo vse podatke, tudi če se ponavljajo. Če je precej podatkov z enako vrednostjo, za prikaz v primeru vrstnih spremenljivk lahko uporabimo **stolpčni prikaz** na Sliki 6. Na abscisno os naneseemo različne vrednosti, ki jih spremenljivka lahko zavzame (urejene po vrsti), na ordinatno os pa frekvenca (ali tudi relativne frekvenca) posameznih vrednosti. Pri stolpčnem prikazu do višine frekvenc vrisemo ozke stolpce, katerih širina je lahko poljubno izbrana.



Slika 6: Stolpčni prikaz za spremenljivko Priljubljenost iz Primera 1.

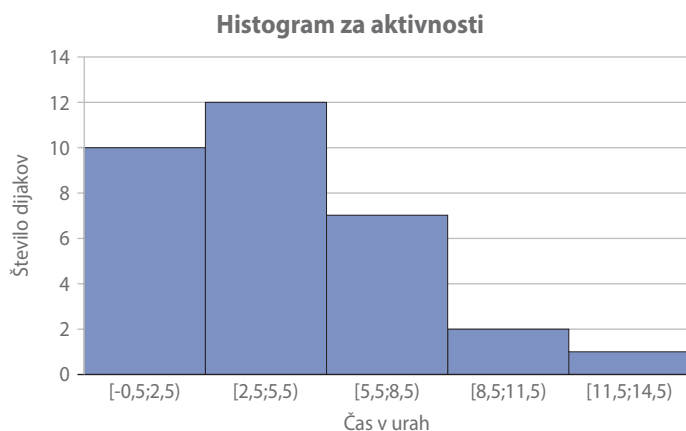
Številске podatke (vrednosti razmičnih (3) ali razmernostnih spremenljivk (4)), razvrščene po frekvencah v frekvenčno tabelo, lahko prikažemo s **frekvenčnim poligonom** (v učnem načrtu linijski diagram) na Sliki 7. Oznake vrednosti na abscisi so enako oddaljene druga od druge. V primeru, ko so frekvenca pri mnogih vrednostih na abscisni osi nič ali so le-te pogosto zastopane le enkrat, podatke lahko ob izračunu mer srednjih vrednosti in razpršenosti pregledno predstavimo s škatlo z brki. Meje škatle, ki je poljubno široka določata prvi in tretji kvartil, z navpičnico



Slika 7: Frekvenčni poligon za spremenljivko Dosežek iz Primera 1.

v njej predstavimo tudi mediano, brki z vodoravnicama na vsako stran pa se končajo z navpičnicama pri najmanjši in največji vrednosti. Celotno škatlo z brki lahko tudi zasukamo za kot 90 stopinj kot kaže Slika 10.

Pri velikem številu različnih številskih podatkov postane ranžirna vrsta nepregledna in vrednosti spremenljivk pri urejanju razporedimo v navadno enako velike (frekvenčne) razrede. Kadar so podatki razvrščeni v razrede, jih navadno prikazujemo s **histogramom**. Histogram je predstavljen s pravokotniki, ki jih je toliko kot razredov. Vsak pravokotnik ima dolžine stranic enake širini razreda (razliki med zgornjo mejo razreda z_i in spodnjo mejo razreda s_i , $i = 1, \dots, r$) in frekvenci razreda (Slika 8). Izbira razredov je odvisna od naloge in od količine podatkov. Navadno priporočamo, naj bo razredov od 5 do 15, toliko informacij človek lahko naenkrat absorbira. Za število razredov r lahko uporabimo Sturgesovo formulo: $r = 1 + 3,322 \log n$, kjer je n število podatkov, ki jih razvrščamo v razrede (Sturges, 1926). V Primeru 1 s podatki na Sliki 8 je $n = 32$ in dobimo $r = 6,0001$. Formula priporoča 6 razredov, vendar v tem primeru meje (ali sredine) razredov ne bi bile celoštevilčne, ali pa bi bil kak razred prazen. Za primer 5 razredov smo blizu priporočljivega števila in sredine razredov so naravna števila (slika 8), kar omogoča hitro računanje mer »na roko«.



Slika 8: Histogram za spremenljivko Aktivnost iz Primera 1.

Razsevni diagram uporabimo za prikaz vrednosti dveh spremenljivk na enoti v populaciji (ali vzorcu) in je prikazan na Sliki 11.

3.3 Mere srednje vrednosti

Tako kot se z urejenostjo spremenljivk večja možnost predstavitve podatkov, podobno z urejenostjo narašča število različnih mer, ki jih danim podatkom določamo. Dijakom predstavimo tri **mere srednje vrednosti** (oziroma mere osredinjenosti po gimnazijskem ali mere sredine po osnovnošolskem učnem načrtu), ki so jih spoznali že v osnovni šoli, v gimnaziji pa skušamo podati razlike med njimi v uporabi.

Z mero srednje vrednosti skušamo z eno vrednostjo čim bolj določiti (opisati) vrednosti vseh podatkov. Za različne vrste spremenljivk in različne namene imamo različne mere. Katero mero bomo uporabili, je odvisno od vrste spremenljivk (imenske, vrstne, razmične, razmernostne) in od tega, čemu naj bi izračunana mera služila.

Od treh mer srednjih vrednosti, ki jih učenci spoznajo že v osnovni šoli, lahko vsem spremenljivkam določimo le modus,

mediane ne moremo določati imenskimi spremenljivkam, aritmetične sredine pa niti imenskimi niti vrstnim spremenljivkam.

Modus Mo lahko določimo vsem spremenljivkam v primeru 1: za Spol je $Mo =$ moški, za Priljubljenost sta modusa dva: $Mo =$ Ne maram preveč., $Mo =$ Imam še kar rad., za Dosežek dobimo $Mo = 7$ (desetic), torej je največ dijakov na testu dobilo med 70 % in 79 %, in za Aktivnost velja $Mo = 2$. V največjem številu so dijaki zapisali, da v povprečju tedensko namenijo 2 uri za učenje matematike.

Medino Me določimo spremenljivkam Priljubljenost: $Me =$ Ne maram preveč., Dosežek: $Me = 6$ in Aktivnost: $Me = 3,5$. Aritmetično sredino $\bar{x}$ določimo le številskim spremenljivkam. Za Dosežek dobimo $\bar{x} = 6$ in za Aktivnost $\bar{x} = 4,375$.

Vsaka od srednjih vrednosti ima prednosti in slabosti. Mediana se na primer ne more računati za imenske spremenljivke, a je lahko boljša od aritmetične sredine, kadar se v veliki količini podatkov pojavljajo kakšne ekstremne vrednosti (osamelci), ki močno popačijo izračun aritmetične sredine. Nekaj prednosti in slabosti smo povzeli po (Korenjak Černe, 2007).

Prednosti modusa so:

- določamo ga vsem vrstam spremenljivk, če se vsaj kakšna vrednost ponovi,
- dobimo ga s preštevanjem enakih vrednosti (brez računanja),
- ni odvisen od vrednosti, ki za populacijo niso tipične.

Pomanjkljivosti modusa so:

- premalo občutljiv za spreminjanje posameznih vrednosti,
- modusov je lahko več in tedaj informacija ni popolna (določanje modusa je lahko vprašljivo tudi pri zaokroževanju vrednosti).

Prednosti mediane so:

- lahko razumljiva,
- za določanje ni treba poznati vseh vrednosti (le vrstni red enot in tiste enote, ki ležijo okoli sredine v ranžirni vrsti).

Pomanjkljivosti mediane so:

- za asimetrične porazdelitve podatkov je mediana različna od večine vrednosti,

- ni zelo občutljiva za spremembe vrednosti (vrednost mediane se spremeni le, če so spremembe vrednosti spremenljivk take, da vrednosti preidejo iz ene polovice ranžirne vrste v drugo).

Prednosti aritmetične sredine:

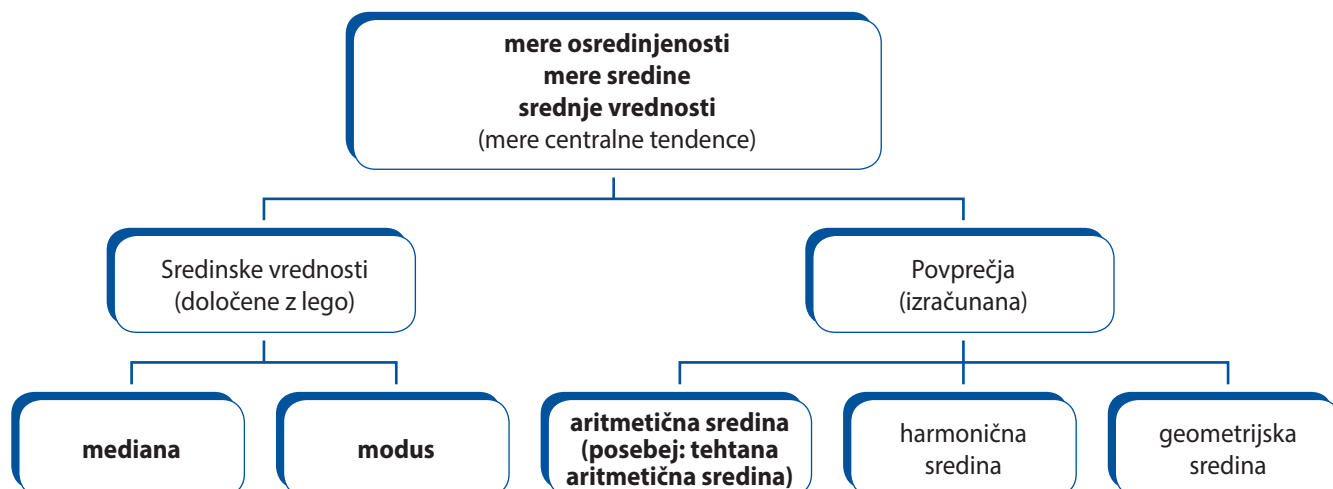
- občutljiva na spremembe vrednosti (nanjo vplivajo vse vrednosti podatkov),
- dobro povzame celoto podatkov,
- računamo jo lahko tudi za velike količine podatkov, razvrščenih v razrede.

Pomanjkljivosti aritmetične sredine:

- izračunamo jo lahko le za številske spremenljivke,
- osamelci (ekstremno velike ali majhne vrednosti) močno popačijo njeno vrednost.

Kadar imamo možnost za dano spremenljivko določiti več mer srednje vrednosti, nam njihova primerjava daje še več informacij o spremenljivki. Spremenljivka Dosežek iz Primera 1 ima vse mere zelo blizu skupaj: $Mo = 7$, $Me = 6$, $\bar{x} = 6$. Za spremenljivko to pomeni, da je vrednost z največjo frekvenco blizu sredine vseh vrednosti razvrščenih v ranžirno vrsto in da sta obe polovici porazdelitve podatkov dokaj simetrični, saj je tudi aritmetična sredina blizu tej vrednosti. Pri spremenljivki Aktivnost pa se mere bolj razlikujejo med seboj ($Mo = 2$, $Me = 3,5$, $\bar{x} = 4,375$), kar nakazuje, da porazdelitev podatkov ni simetrična.

Namesto besed modus, mediana ali aritmetična sredina v medijih, pri pouku in v **vsakdanjem življenju pogosteje zasledimo izraz povprečje**. Wikipedija (2021) ima preprosto definicijo: aritmetična sredina ali povprečje je ena izmed srednjih vrednosti. Tudi v srednješolskem učbeniku Matematika 1 (Felda in sod., 2021: str.123) enačijo povprečno vrednost ali povprečje z aritmetično sredino. Korenjak Černe (2007) ravno tako enači pojma: povprečje ali aritmetična sredina; dalje pa mere srednje vrednosti deli na sredinske vrednosti (to sta modus, mediana) in povprečja (to pa so aritmetična sredina ali povprečje, harmonična sredina, geometrijska sredina, ki jih lahko računamo za številske spremenljivke). Povprečje je torej uporabljeno dvakrat: kot sopomenka za aritmetično sredino in kot nadpomenka za številske mere srednje vrednosti. Košmelj (2008) v visokošolskem učbeniku podobno uporabi povprečje za nadpomenko in pravi: Pov-



Slika 9: Mere srednje vrednosti, ki se pojavijo v srednješolskih učbenikih.

prečje se računa le za številske spremenljivke in je vrednost, za katero velja: če bi bili vsi podatki enaki, bi bili enaki povprečju. Predstavimo izraze na Sliki 9.

Morda je sedaj bolj razumljiva definicija v Statističnem terminološkem slovarju (Košmelj in sod., 2002: str. 98), ki pravi, da je povprečje srednja vrednost, ki je izračunana iz vseh vrednosti številske spremenljivke, npr. aritmetična sredina, harmonična sredina, geometrijska sredina, povprečni verižni indeks. Na Wikipedijo ali spletne razlage se pač ne smemo vselej zanesti, na to je dobro opozarjati tudi dijake.

Različno poimenovanje v statistiki dijaka lahko pošteno zmede, še bolj problematične pa so **neenotne definicije** določenih pojmov za vrednosti, ki jih dijaki računajo. Tako se pojavijo v srednješolskih učbenikih različne definicije za **kumulativno frekvenco** ali imenovano tudi kumulativna frekvenc s simbolom F_i :

- a) Kumulativna frekvenca F_i , $i = 1, \dots, r$ podatkov razvrščenih v r razredov nam poda število enot, ki so manjše od zgornje meje i -tega razreda in se določi s formulo:

$$F_i = F_{i-1} + f_i; \quad i = 2, \dots, r; \quad F_1 = f_1$$

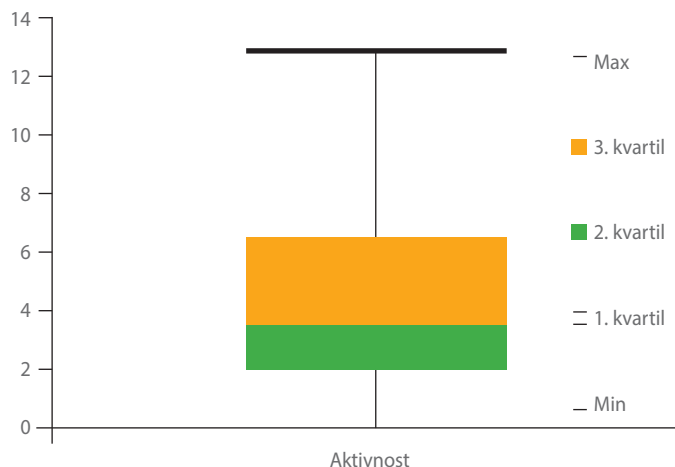
- b) Kumulativna frekvenca je število enot F_i , ki imajo manjše vrednosti od spodnje meje i -tega razreda:

$$F_i = F_{i-1} + f_{i-1}; \quad i = 2, \dots, r; \quad F_1 = 0.$$

S prvo definicijo sovpadajo tudi definicije v učbenikih (Pavlič in sod. 2013), str. 158 in (Felda in sod., 2021). Druga definicija je navedena v priročniku za učitelje (Kmetič in Sirnik, 2008: str. 266) z dodano opombo, da je mogoča tudi definicija z zgornjo mejo razreda. Začetno kumulativno frekvenco 0 najdemo tudi v definiciji spletnega učbenika Vega 1 (2013), kjer preberemo: *Kumulativna frekvenca F_i pove, koliko podatkov je manjših od zgornje meje i -tega razreda. Enaka je vsoti vseh frekvenc do i -tega razreda, vključno z njim.*

3. 4 Mere razpršenosti

Z merami razpršenosti želimo povedati, kako so razvrščeni številski podatki okrog mere srednje vrednosti; ali so podatki blizu



Slika 10: Škatla z brki za podatke spremenljivke Aktivnost v Primeru 1 s kvartili na mejah obarvanega polja, ki kažejo nesimetričnost.

srednje vrednosti (majhna razpršenost) ali daleč (velika razpršenost). Obravnavamo naslednje mere razpršenosti podatkov: variacijski razmik (VR), medčetrtnski razmik (Q) in standardni odklon (σ), ki jih lahko določimo le številskim spremenljivkam.

V mnogih srednješolskih učbenikih ni natančne definicije za kvartile, ki so potrebni za izračun medčetrtnskega razmika Q, ki je predpisan v učnem načrtu. Kvartili so uvedeni tudi v obravnavo pri matematičnih vsebinah v osnovni šoli in so predpisani v predmetnem izpitnem katalogu za poklicno matura. Z njimi lahko narišemo škatlo z brki.

Za pojem kvartil so v učbenikih raznolike trditve:

- Kvartil je vsaka od treh vrednosti številske spremenljivke, s katerimi so opazovane enote, urejene od najmanjše do največje, razdeljene na 4 enake dele (Košmelj in sod., 2002).
- Trije kvartili Q_1 , Q_2 in Q_3 razdelijo podatke v štiri skupine tako, da je v vsaki skupini četrtnina vseh podatkov (Vega 1, 2013).
- Mediana razdeli podatke na dve polovici, lahko pa razdelimo tudi ti dve polovici in dobimo 4 enako močne skupine podatkov. Mejnike posameznih skupin imenujemo kvartile (Pavlič in sod., 2013: str. 166).
- 1. kvartil je mediana prve oz. spodnje polovice podatkov (Pavlič in sod., 2013: str. 166).
- 1. kvartil je mediana vseh podatkov, ki so v nizu dosežkov, urejenih po velikosti, levo od mediane vseh podatkov (Felda in sod., 2021).
- 1. kvartil je vrednost opazovane spremenljivke, od katere ima 25 % enot manjše in 75 % enot večje vrednosti (Košmelj in sod., 2002).

Omenjene definicije so preveč splošne, saj na primer na množici s 15 podatki v ranžirni vrsti (ki se lahko tudi ponavljajo) ne moremo določiti treh vrednosti kvartilov, ki razdelijo podatke v 4 skupine tako, da je v vsaki skupini natanko četrtnina podatkov (da ima 25 % enot manjše vrednosti in 75 % večje vrednosti). Dijakom definicije ne omogočajo konkretnega izračuna kvartilov na posameznih množicah podatkov. Statistični terminološki slovar (Košmelj in sod. 2002) podaja definicijo za velike količine podatkov, toda v srednješolskih učbenikih, kjer dijaki pojme spoznavajo na majhnih količinah, da jih samostojno lahko določijo, je definicija za majhne množice podatkov potrebna. Žal so v nekaterih učbenikih tudi izračuni za kvartile nepravilni.

V večini učbenikov je podana natančna definicija za 2. kvartil, imenovan tudi mediana ali središčnica, vendar za ostala dva kvartila v učbenikih ni podane natančne definicije za enolično določanje. Brez uporabe pojma mediane bi lahko podali naslednjo definicijo:

1. kvartil je vrednost z oznako Q_1 , od katere je vsaj 25 % podatkov manjših ali enakih tej vrednosti in vsaj 75 % podatkov večjih ali enakih tej vrednosti Q_1 (to definicijo najdemo v Vega 1, 2013).

2. kvartil je vrednost Q_2 , od katere je vsaj polovica vseh podatkov manjših ali enakih tej vrednosti in vsaj polovica podatkov večjih ali enakih tej vrednosti (2. kvartil ima tudi ime mediana $Q_2 = M_e$).

3. kvartil je vrednost z oznako Q_3 , od katere je vsaj 75 % podatkov manjših ali enakih tej vrednosti in vsaj 25 % podatkov večjih ali enakih tej vrednosti Q_3 .

Za enolično določanje je treba še dodati, da v primeru, ko po izbrani definiciji kvartilu ustreza vsako število med dvema vrednostma spremenljivke x_k in x_{k+1} v ranžirni vrsti, zanj uporabimo vrednost $(x_k + x_{k+1})/2$.

Za spremenljivko Dosežek v Primeru 1 lahko določimo variacijski razmik kot razliko med največjo in najmanjšo vrednostjo: $VR = 9$, medčetrtnski razmik: $Q = Q_3 - Q_1 = 2,5$ in standardni odklon $\sigma = 2,24$, za spremenljivko Aktivnost pa vrednosti $VR = 13$, $Q = 4,5$ in $\sigma = 3,13$.

Mer razpršenosti ni smiselno primerjati med seboj, saj nam variacijski razmik kaže razliko med največjim in najmanjšim podatkom, medčetrtnski razmik kaže odmik polovice podatkov od mediane in standardni odklon predstavlja koren povprečja kvadratov razdalj posameznih podatkov od aritmetične sredine.

3.5 Empirična preiskava v obliki statistične naloge

Iskanje povezav dveh ali več spremenljivk v učbenikih ni posebej obdelano. A je iskanje odvisnosti ali povezanosti spremenljivk navadno cilj empiričnih statističnih preiskav. Tako je dobro, da napravimo kak zgled, preden damo dijakom samostojno statistično preiskavo. Na spletni strani (Drobnič Vidic, 2020) skozi urejanje, prikazovanje podatkov, določanja mer srednje vrednosti in variabilnosti podatkov iz Primera 1 gradimo statistično preiskavo, ki jo predstavimo dijakom kot statistično nalogo ob koncu obdelane statistične vsebine. Primer 1 nadgradimo na primer tako:

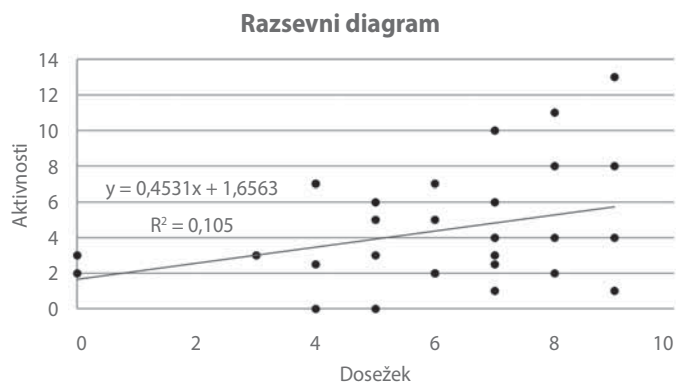
Učitelj 1. b razreda gimnazije Bič iz Primera 1 si je v šolskem letu 20/21 zaradi učenja na daljavo zadal vprašanje, ali spol, priljubljenost matematike (kot predmeta) pri dijakih ali njihove samostojne aktivnosti vplivajo na dosežek pri matematičnem testu. Izvedel je anketo med dijake, podatke pridobil in uredil, prikazal in obdelal ter iskal morebitne povezave. Zapiši ugotovitve, ki jih je na podlagi podatkov v Primeru 1 učitelj lahko dobil!

Podatki, ki jih je učitelj pridobil in jih dijaki pri pouku ali samostojno skozi učne ure statističnih vsebin obdelajo, kratko pa smo jih predstavili tudi v prispevku, kažejo, da ima razred kot celota pri njegovem predmetu dokaj slabo oceno, saj je aritmetična sredina dosežkov na testu 6 desetih odstotkov (kar je med 60 odstotkov in 69 odstotkov celotnih točk in ustreza navadno oceni dobra 2 ali slaba 3. Mediana kaže, da je polovica vseh dijakov dosegla tak dosežek ali manj! Ugotovil je tudi, da je polovica dijakov dosegla uspeh med 5 desetih odstotkov in 7,5 desetih odstotkov, četrtina pa je dosegla 5 desetih ali manj, kar določajo medčetrtnski razmiki na sliki 10. Znanje spodnje četrtine dijakov je torej zelo šibko. Toda kaj je vzrok za slabo znanje na testu? Odgovor na to vprašanje nam da lahko le ugotavljanje povezanosti in odvisnosti spremenljivk, kar pa je za dijake prvih letnikov precej trd oreh.

Razsevni diagram znajo dijaki sicer brati že iz osnovne šole, saj se zahtevajo tudi na NPZ (Državni izpitni center, NPZ), kar predstavlja prvi stik s povezanostjo. Če dijaki prikažejo vrednosti dveh številskih spremenljivk (dva podatka na vsaki enoti) z razsevnim diagramom in ležijo podatki na neki premici, lahko zaključijo,

da podatki spremenljivk kažejo linearno odvisnost (povezanost). Če pa so podatki blizu neke premice, so blizu linearne odvisnosti. Mera za to, kaj pomeni pojem »blizu neke premice« je na primer vzorčni korelacijski koeficient, za njegovo izpeljavo pa potrebujemo znanje odvodov, zato je **natančnejša obravnava povezanosti spremenljivk bolj primerna za 4. letnik**. Če obdelamo statistiko v 1. letniku, potem dijake le seznanimo z računalniškim izrisom take premice (t. i. regresijske premice) v razsevnem diagramu. Tako pravzaprav dijaki nekih oprijemljivih postopkov za ugotavljanje povezanosti pri številskih spremenljivkah tedaj še nimajo. Glede na razsevni diagram lahko postavljajo neke domneve o povezanosti, kakšna je zanesljivost njihovih domnev, pa ostane skrito.

Podatki dijakov v Primeru 1 za spremenljivki Dosežek in Aktivnost so vrisani v razsevni diagram na sliki 11. Točke so precej razpršene in niso nanizane blizu ene same premice, zato diagram ne nakazuje domneve, da bi bili ti dve spremenljivki linearno povezani, torej ne moremo trditi, da samostojno domače delo za matematiko vpliva na slab izid testa. Ali na dosežek morda vpliva spol ali priljubljenost predmeta?

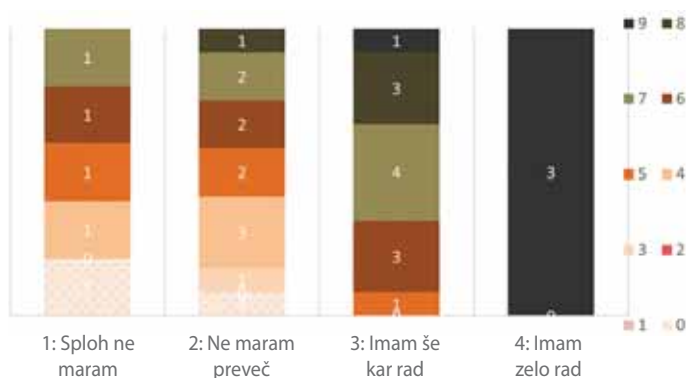


Slika 11: Razsevni diagram s premico za možno povezanost dveh številskih spremenljivk.

Povezanost dveh spremenljivk, od katerih je vsaj ena opisna (imenska ali vrstna) z malo različnimi vrednostmi, lahko raziskujemo grafično (Kmetič in Sirnik, 2010: str. 302) ali z več frekvenčnimi tabelami. Pri eni opisni spremenljivki (z malo različnimi vrednostmi) in drugi številski spremenljivki lahko za vsako različno vrednost opisne spremenljivke določimo frekvenčno tabelo s pripadajočimi vrednostmi številske spremenljivke ter izračunamo aritmetično sredino in standardni odklon. S primerjavo teh mer lahko podamo določene domneve.

Primerjajmo spremenljivki Priljubljenost in Dosežek iz Primera 1. Prva je opisna, druga številska. Ker ima Priljubljenost le 4 različne vrednosti, lahko določimo frekvenčno tabelo za Dosežek pri posameznih enotah z enako vrednostjo opisne spremenljivke in povprečni dosežek za te enote (določimo aritmetično sredino). Dobimo 4 frekvenčne tabele in izračunamo povprečni dosežek 4,4, pri enotah (dijakih), ki matematike sploh ne marajo, povprečni dosežek 4,9 pri dijakih, ki matematike ne marajo preveč, povprečni dosežek 7,0 pri dijakih, ki imajo matematiko še kar radi in povprečni dosežek 9 pri dijakih, ki imajo matematiko zelo radi. Vrednosti povprečij nakazujejo, da pri naraščanju priljubljenosti do predmeta tudi dosežek pri testu tega predmeta narašča. Podoben zaključek lahko podamo iz slike 12. Seveda zopet ne vemo, kako zanesljive so naše domneve o povezanosti.

DOSEŽEK GLEDE NA PRILJUBLJENOST



Slika 12: Prikaz možne povezanosti številске in opisne spremenljivke, kjer so boljši dosežki prikazani z bolj temno barvo.

Prikazi nakazujejo, da dosežek na testu 1. b razreda najbrž ni odvisen toliko od samostojnih matematičnih aktivnosti pri dijaki, kot od priljubljenosti tega predmeta med dijaki tega razreda. O zanesljivosti takih domnev seveda ne moremo povedati nič.

Dijaki nato lahko za samostojno domače delo izvedejo podobno statistično nalogo, kjer podatke pridobijo sami:

Statistična naloga:

Sam naredi anketo med sošolci z vprašanji (4–6) o neki temi, ki te zanima. Vprašanja naj se nanašajo naj vsaj 3 različne vrste spremenljivk, pri katerih te bo zanimala tudi kakšna povezanost pri odgovorih nanje. Vprašanja pošlji po e-pošti ali telefonu in pridobi podatke, postopek v raziskavi opiši. Uredi podatke in jih primerno predstavi (tudi z delitvijo v razrede). Lahko si pomagaš s programom Microsoft Excel. Določi možne mere srednje vrednosti in mere razpršenosti in jih interpretiraj. Skušaj najti povezavo med spremenljivkami. Podaj ugotovitve tvoje raziskave.

Zaključek

Hitro razvijajoče in vsestransko uporabne statistične vsebine mladi začnejo spoznavati že v vrtcu in v osnovni šoli. Da bi v srednji šoli znanje v kratkem času, namenjenim tem vsebinam, čim učinkoviteje nadgradili, lahko celotno vsebino predelamo z enim samim primerom in hkrati gradimo primer statistične naloge, ki naj bi jo dijaki izvedli sami. Pri tem v poplavi različnih imen pazimo na dosledno uporabo izbranega poimenovanja in definicije, ki bodo omogočala pravi izračun statističnih mer. Z dijaki razpravljamo o interpretaciji rezultatov in se neprestano učimo, da delamo čim manj napak.

Literatura

- Bahovec, E., Bregar, K., Čas, M. in sod. (1999). *Kurikulum za vrtce*. Dostopno na: <https://www.gov.si/assets/ministrstva/MIZS/Dokumenti/Sektor-za-predskolsko-vzgojo/Programi/Kurikulum-za-vrtce.pdf>
- Drobnič Vidic, A. (2020). Obdelava podatkov in statistika. Pouk na daljavo. Dostopno na: <https://skupnost.sio.si/course/view.php?id=65§ion=1>.
- Državni izpitni center: Nacionalno preverjanje znanja. Dostopno na: https://www.ric.si/preverjanje_znanja/predmeti/matematika3/
- Felda, D., Bon Klanjšek, M., Škrlec, M., Dvoržak, B. in sod. (2021). *Matematika 1, učbenik za gimnazije*. Državna Založba Slovenije (DZS), Ljubljana.
- Kmetič, S. in Sirnik, M. ur. (2010). *Posodobitve pouka v gimnazijski praksi*. Zavod RS za šolstvo.
- Korenjak Černe, S. (2007). *Statistika 1: gradiva*. Ekonomska fakulteta. Pridobljeno 28 februar 2019. Dostopno na: <https://pdfslide.tips/documents/6-ranzirna-vrsta-rang-kvantilni-rang-in-s-korenjak-terne-statisticne.html>
- Košmelj K. (2008). *Uporabna statistika*. Biotehniška fakulteta, Ljubljana.
- Košmelj, B., Arh, F., Doberšek-Urbanc, S., Ferligoj, A. in Omladič, M. (2002). *Statistični terminološki slovar*. Statistično društvo Slovenije.
- Pavlič, G., Kavka, D., Rugelj, M., Šparovec, J. in sod. (2013). *LINEA NOVA, matematika za gimnazije*. Modrijan založba d. o. o.
- Sturges, H.A. (1926). The choice of class interval. *Journal of the American Statistical Association*, 21 (153): 65–66.
- Vega 1, i-učbenik za matematiko v 1. letniku gimnazij (2013). Dostopno na: <https://eucbeniki.sio.si/vega1/index.html>
- Wikipedia: Aritmetična sredina (Slo). Dostopno na: https://sl.wikipedia.org/wiki/Aritmeti%C4%8Dna_sredina
- Žakelj, A., Bon Klanjšek, M., Jerman, M., Kmetič, S., Repolusk, S., Ruter, A. (2008). *Učni načrt: Matematika - Gimnazija; Splošna, klasična in strokovna gimnazija*. Učni načrt, obvezni predmet in matura (560 ur). Ministrstvo za šolstvo in šport: Zavod RS za šolstvo, Ljubljana. Dostopno na: http://eportal.mss.edus.si/msswww/programi2018/programi/media/pdf/un_gimnazija/un_matematika_gimn.pdf