

Raziskovalna naloga na Evropskih statističnih igrah: Nasveti za pripravo nalog

Dr. Aleš Toman
Univerza v Ljubljani, Ekonomska fakulteta

Izveček

Evropske statistične igre so mednarodno ekipno tekmovanje, ki spodbuja radovednost in zanimanje za statistiko, uporabo uradnih statističnih podatkov, osvajanje dodatnih statističnih znanj in veščin ter skupinsko delo in sodelovanje med mladimi. Na šolski ravni tekmovanja ekipe rešujejo spletni test iz poznavanja statističnih pojmov in podatkov, na državni ravni pripravijo statistično raziskovalno nalogo na osnovi danega podatkovnega niza, na evropski ravni pa pripravijo videoposnetek, v katerem s pomočjo uradnih statističnih podatkov obravnavajo predpisano temo. Avtor prispevka kot član strokovne žirije že 5 let ocenjuje raziskovalne naloge na državni ravni tekmovanja. V prispevku navaja primere dobre in manj dobre prakse, vse z namenom, da bodo lahko ekipe v prihodnje pripravljale še boljše naloge. Poudarek je na tistih vsebinah, ki so tesno povezane s poukom matematike (izračuni in grafikoni). Kljub temu so zapisani nasveti dovolj splošni, da so uporabni pri vsaki raziskovalni nalogi. Prispevek vključuje tudi kratek opis vseh ravni tekmovanja in spodbuja učiteljice in učitelje matematike k mentoriranju ekip na svojih šolah.

Ključne besede: Evropske statistične igre, raziskovalna naloga, analiza podatkov, statistični grafikoni

European Statistics Competition: Tips for Preparing Research Papers

Abstract

The European Statistics Competition is an international team competition that encourages young people to work together and cooperate while using official statistical data and developing their statistical knowledge and skills. At the school level, teams complete an online test on statistical concepts and data; at the national level, they write a research paper based on an actual dataset; and at the European level, they create a video in which they present a given topic using official statistical data. The author is a member of the expert jury and has been evaluating research papers at the national level for five years. The article offers examples of both the best and less effective techniques to help teams produce even better papers in the future. Although the emphasis is on topics directly relevant to mathematics (calculations and diagrams), these broad guidelines adhere to any research paper. The summaries of each competition stage further encourage math professors to mentor school teams.

Keywords: European Statistics Competition, research paper, data analysis, statistical graphs.

1 Opis tekmovanja

Evropske statistične igre so mednarodno ekipno tekmovanje srednješolcev, ki ga organizirajo statistični urad Evropske unije Eurostat in nacionalni statistični uradi sodelujočih držav z namenom, da bi med mladimi spodbudili zanimanje za statistiko in uporabo podatkov uradne statistike ter s tem prispevali k **dvigu statistične pismenosti** v Evropi (<https://www.esc2022.eu>). Prve

Evropske statistične igre so potekale v šolskem letu 2017/18. Tekmovanje se vselej odvija v dveh fazah, prva faza je nacionalna in poteka v nacionalnih jezikih sodelujočih držav, druga faza pa je evropska in poteka v angleškem jeziku. Popularnost iger raste tako v Sloveniji kot tudi v Evropi. Ker je tekmovanje v celoti izvedeno preko spleta, je lahko nemoteno potekalo tudi v času epidemije. O tekmovanju in uspehu ekipe Teglzaróže¹ iz leta 2021 sta članici ekipe skupaj z mentorico že poročali v reviji Matematika v šoli (Rauter Repija in drugi, 2021).

¹ Imena ekip si tekmovalke in tekmovalci izberejo sami. Imena članic in članov ekip niso objavljena.

1.1 Nacionalna faza tekmovanja

Slovenija na Evropskih statističnih igrah sodeluje že od samega začetka, nacionalno fazo tekmovanja organizira Statistični urad Republike Slovenije (SURS) v sodelovanju s Statističnim društvom Slovenije (<https://www.stat.si/igre>). Nacionalna faza tekmovanja se deli na šolsko in državno raven. Na obeh ravneh dijakinje in dijaki tekmujejo v stalnih največ tričlanskih ekipah v dveh starostno ločenih tekmovalnih skupinah, in sicer

- v kategoriji A dijakinje in dijaki zadnjih dveh letnikov srednjih šol (16–18 let),
- v kategoriji B pa dijakinje in dijaki prvih dveh letnikov srednjih šol (14–16 let).

Na šolski ravni tekmovanja ekipe odgovarjajo na 3 sklope po 10 vprašanj izbirnega tipa.

- Sklop a) preverja osnovno znanje statistike.
- Sklop b) preverja poznavanje in uporabo uradnih statističnih virov.
- Sklop c) preverja razumevanje izbrane Eurostatove publikacije s statističnimi podatki.

Šolsko tekmovanje poteka v obliki spletnega testa in traja približno 2 tedna, ekipe pa lahko na vprašanja odgovarjajo v poljubnem tempu in zaporedoma na več napravah (v šoli, dijaškem domu, doma ...). Sodelujoči na šolski ravni tekmovanja lahko osvojijo bronasto priznanje.

Najuspešnejše ekipe se uvrstijo na **državno raven tekmovanja**. Tam prejmejo obsežen podatkovni niz, na osnovi katerega pripravijo **statistično raziskovalno nalogo**, ki jo oceni strokovna žirija. Ekipe so povsem samostojne pri izbiri raziskovalne teme, vprašanj, hipotez, podmnožice podatkov ter metod dela. Časa za pripravo naloge imajo približno 4 tedne, nalogo nato v ocenjevanje oddajo preko spletnega obrazca. Sodelujoči na državni ravni tekmovanja lahko osvojijo srebrno ali zlato priznanje.

1.2 Evropska faza tekmovanja

V evropsko fazo tekmovanja se iz vsake sodelujoče države uvrstita po 2 (prvo leto tekmovanja pa 3) najuspešnejši ekipe iz vsake starostne skupine. Organizator tekmovanja izbere temo, nato vsaka ekipa pripravi **videoposnetek**, v katerem s pomočjo uradnih statističnih podatkov na zanimiv način obravnava izbrano tematiko. Vsebinsko predstavijo v največ 2 minutah, temu sledi še največ 10 sekund za navedbo virov. Videoposnetek je lahko v nacionalnem ali angleškem jeziku, obvezno pa mora biti opremljen z angleškimi podnapisi. Ekipe skupaj z videoposnetkom preko spletnega obrazca oddajo še spremljevalni dokument, v katerem na največ 4 straneh opišejo postopek priprave videoposnetka. Časa za pripravo obeh datotek imajo približno 6 tednov, nato datoteki po vnaprej znanem kriteriju oceni mednarodna strokovna žirija. Sodelujoči v evropski fazi tekmovanja lahko osvojijo pokal ali priznanje.

1.3 Namen prispevka

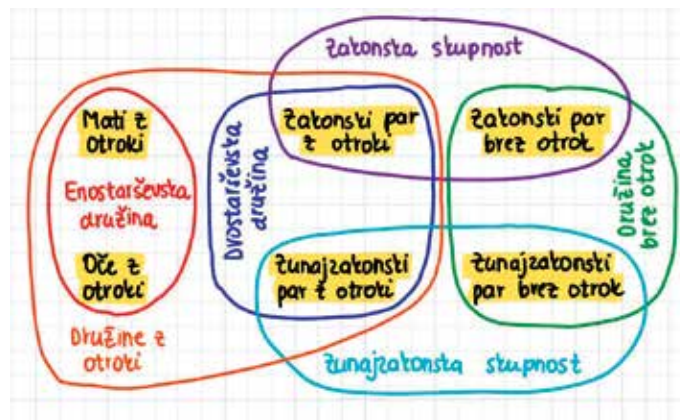
Ekipe na državni ravni tekmovanja pripravijo izredno zanimive in kakovostne raziskovalne naloge, najvišje uvrščene naloge pa vsako leto SURS objavi na svoji spletni strani. Objavljene izdelke

posebej pozorno pregledajo prihodnje generacije tekmovalk in tekmovalcev, saj v njih iščejo ključ do visoke uvrstitve na državni ravni tekmovanja. A tudi v najboljših nalogah se najdejo napake in nespretnosti. V upanju, da se te zaradi objave nalog ne bi širile, sodelavci SURS in člani Statističnega društva pripravljajo izobraževanja za ekipe in njihove mentorice in mentorje. Temu je namenjen tudi ta prispevek. Predstavljeni zgledi izhajajo iz raziskovalnih nalog, ki so jih ekipe pripravile v obdobju 2018–2022. Čeprav nekateri nasveti izhajajo iz priročnika *Making Data Meaningful* (2009), so v prispevku opisana predvsem lastna opažanja.

2 Podatkovni niz in raziskovalno vprašanje

Državna raven tekmovanja se prične, ko SURS ekipam posreduje podatke, na osnovi katerih pripravijo raziskovalne naloge. Podatki so shranjeni v **Excelovi datoteki** z več listi, numerične vrednosti pa so zapisane v tabeli, kjer stolpci predstavljajo spremenljivke, vrstice pa geografske enote. Vrednosti so najpogostejše dane na ravni 212 slovenskih občin, v dodatnih listih pa so lahko podane še agregirane vrednosti za 12 statističnih regij in Slovenijo kot celoto. V nadaljnjih listih Excelove datoteke ekipe najdejo metapodatke ter prostorski šifrant. Metapodatki so podrobni opisi spremenljivk, prostorski šifrant pa kodna tabela, ki jo ekipe potrebujejo, saj so geografske enote v tabeli s podatki določene s kodami in ne imeni.

V letu 2020 so imele ekipe na voljo kar sto spremenljivk, a so vse opisovale *družine v Sloveniji* v točno določenem trenutku. Leto kasneje je bilo spremenljivk skoraj pol manj, a so ekipe imele na voljo vrednosti vsake spremenljivke v 5 zaporednih letih, zato so lahko tematiko *okolja v Sloveniji* spremljale skozi čas. Prva naloga vsake ekipe je, da dane podatke zelo natančno preuči in raziše njihovo kompleksno strukturo. Le tako si lahko zastavijo ustrezna **raziskovalna vprašanja** in kasneje pripravijo ustrezne izračune (Slika 1).



Slika 1: Klasifikacija družin. Razumevanje medsebojne povezanosti 12 spremenljivk (tipov družin) omogoča računanje ustreznih deležev.

Vseh spremenljivk ni mogoče uporabiti v eni raziskovalni nalogi, ekipa mora zato izbrati natanko tiste, ki ustrezajo postavljenemu raziskovalnemu vprašanju. Podatkovni niz preučijo tudi člani strokovne žirije in zato lahko preverijo ustreznost posamezne izbire. Ekipe lahko uporabljajo le podatke danega podatkovnega niza. Uporaba dodatnih podatkov ni dovoljena in vodi v diskva-

lifikacijo. Lahko pa ekipe iščejo dodatne metapodatke ali ideje za pripravo naloge (Slika 2). Prav tako lahko uporabljajo poljubne vire pri učenju statistične analize, dela z računalnikom in izdelavi predstavitev.



Slika 2: Podatek o količini vode, uporabljeni v procesu izdelave enega avtomobila (zgoraj), v podatkovnem nizu ni bil na voljo. Ker pa je služil kot zanimivost in ne kot informacija za nadaljnje analize, uporaba podatka ni sporna. Vir: Deadsirius (2021).

Čeprav vse ekipe analizirajo iste podatke in je njihov vir iz konteksta povsem jasen, SURS ob uporabi podatkov izrecno zahteva **navedbo vira**, pri čemer je dovolj že zapis *Vir: SURS*. Kadar ekipe podatke še naprej preračunavajo (združevanja, povprečja, deleži), je pravilen še pripis *Lastni izračuni*. Vir praviloma navajamo tudi ob tabelarnih in grafičnih prikazih podatkov.

3 Elementi raziskovalne naloge

Pravila za nacionalni del tekmovanja določajo, da ekipe svoje delo predstavijo na največ **8 prosojnicah**, ki jim lahko sledita največ **2 prosojnici z metodološkimi pojasnili**. Vsaka naloga mora vsebovati uvodno prosojnico in zaključek, pri ostalih 6 prosojnicah pa imajo ekipe proste roke. Žirija ima pri ocenjevanju na voljo le prosojnice, zato je pomembno, da je z njih razvidno logično zaporedje vsebin, zlasti kadar je na eni prosojnici prikazanih več analiz. Ekipe lahko urejenost (pa tudi jasnost in razumljivost) nalog preverijo tako, da ob prosojnicah pripravijo predavanje za mentorja ali prijatelje. Pri tem bodo hitro ugotovile, ali so na pravem mestu zapisale vse potrebne informacije.

3.1 Uvodna prosojnica

Uvodna prosojnica obvezno vsebuje ime ekipe, šole, statistične regije in navedbo starostne kategorije. Preostali prostor na uvodni prosojnici lahko ekipa izkoristi za informativen naslov naloge ter privlačno grafiko, ki lahko prispeva k razumevanju naloge. Pri uporabi slikovnega gradiva pazi, da ne krši avtorskih pravic in da v nalogi navede ustrezne vire. Uvodna prosojnica je prvi stik bralca z raziskovalno nalogo, zato se jo splača čim bolj premišljeno izkoristiti (Slika 3).

3.2 Cilji in hipoteze

Na drugi prosojnici ekipe običajno navedejo cilje in hipoteze. **Hipoteze** niso obvezne, če jih ekipa zapiše, pa morajo biti jasne, argumentirane in zapisane pred analizo podatkov. Nedoločene hipoteze je bolje nadomestiti z navedbo ciljev (Slika 4). Cilji po-



Slika 3: Uvodna prosojnica ekipe Statisti1 (zgoraj) poleg obveznih sestavin vsebuje še osrednje raziskovalno vprašanje, tri slike pa kažejo okoljske vidike, ki jih v nadaljevanju obravnava ekipa. Pri uvodni prosojnici ekipe Lnmstorm (spodaj, ime in logotip šole sta zakrita) je nesrečna izbira pisave močno otežila branje. Vir: zgoraj Statisti1 (2021), spodaj Lnmstorm (2020).

magajo ustvariti pregledno strukturo naloge. Pravila ne zahtevajo, da so cilji zbrani na drugi prosojnici. Ekipa se zato lahko odloči, da cilje preplete s statistično analizo.

Zaželeno je, da si ekipe zastavijo **glavni (primarni) cilj**, ki določa zgodbo raziskovalne naloge, nato pa navedejo še **sekundarne cilje**, s katerimi postopoma dosežejo glavni cilj. Pri številu ciljev naj ekipe upoštevajo, da imajo na voljo 8 prosojnic. Količina analiz, ki jih lahko prikažejo, je zato omejena. Tudi sekundarni cilji so lahko dovolj splošni, da so zapisani kratko in jedrnato. Čeprav je cilj zapisan splošno, so lahko z njim povezane ugotovitve veliko bolj podrobne.

3.3 Statistični kazalniki

Ko so cilji določeni, lahko ekipe v podatkovnem nizu poiščejo ustrezne podatke in se lotijo analize. Večina podatkov je rezultat preštevanja (število prebivalcev, gospodinjstev ...), zato po vrednostih med regijami običajno izstopa osrednjeslovenska (Slika 5), med občinami pa Ljubljana. Ker regije in občine niso enako velike, je pred primerjavo treba opraviti ustrezne preračune.

Statistični kazalnik je iz danih podatkov izvedena (izračunana) spremenljivka, ki omogoča primerjavo različnih geografskih enot. Če sta števili zaposlenih in brezposelnih oseb odvisni od velikosti občine ali regije, je stopnja brezposelnosti kazalnik, s

- HIPOTEZE**1. Tipi družin**

- Večina ljudi v Sloveniji živi v družinah.
- Malo družin je brez otrok.
- Prevladujejo dvostarševske družine.
- Največji delež otrok ima poročene starše, najmanjši delež otrok pa živi v enostarševskih družinah
- Večina partnerjev v Sloveniji je poročenih.

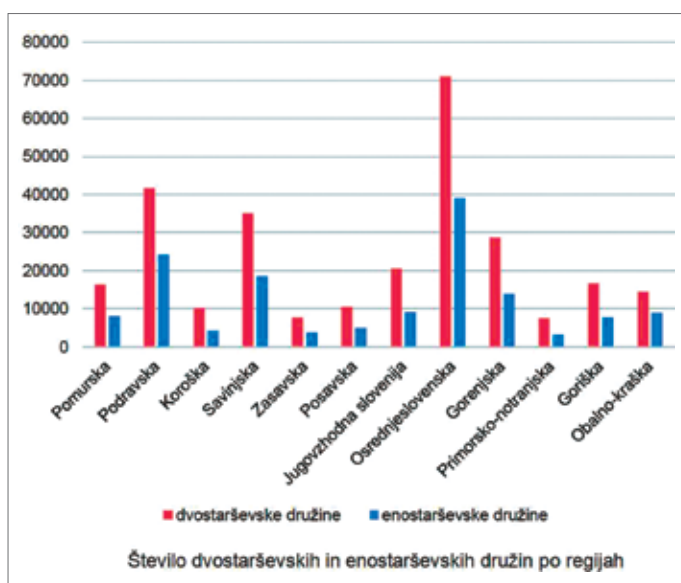
2. PORABA VODE ZA NAMAKANJE ZEMLJIŠČ (2019)

- ★ Najmanj vode za namakanje kvadratnega metra zemljišča se porabi na Gorenjskem, saj je tam velika količina padavin.
- ★ Največ vode za namakanje kvadratnega metra zemljišča se porabi na Pomurskem, saj je na tem področju najmanj padavin.

HIPOTEZE**DEJAVNOSTI**

1. Delež zaposlenih v kmetijskih dejavnostih se skozi leta zmanjšuje.
2. Kmetijske dejavnosti tako v letu 2015 kot v letu 2020 v nobeni občini ne zaposlujejo več kot 50% delovno aktivnega prebivalstva.
3. Delež zaposlenega delovno aktivnega prebivalstva v kmetijskih dejavnostih v letu 2020 presega 30% samo v občinah severovzhodne Slovenije.

Slika 4: Primer presplošnih (zgoraj levo) in ustrezno argumentiranih (spodaj) hipotez. Hipoteze, ki navajajo le izbrana leta in konkretne vrednosti (zgoraj desno), dajejo vtis, da so bile pripravljene po analizi podatkov. Vir: zgoraj levo 20in20 (2020), zgoraj desno Kapljice (2022), spodaj Plan_xy (2021).



Slika 5: Neinformativen prikaz absolutnih vrednosti. V osrednjeslovenski regiji živi približno četrtina prebivalcev Slovenije, zato so tam vrednosti demografskih spremenljivk običajno najvišje. Vir: Teglzaržo (2020).

katerim je mogoče primerjati občine ali regije med seboj. Pri izbiri kazalnikov si ekipe lahko pomagajo s pregledovanjem spletnih strani SURS. Če uporabijo standardni kazalnik, lahko izračunane vrednosti primerjajo s tistimi, ki jih objavi SURS, in preverijo ustreznost izračunov² (Slika 6). Pri izbiri kazalnikov je ekipa lahko zelo izvirna, paziti mora le, da so vsi izračuni dobro dokumentirani (Slika 7).

3.4 Metode dela

Uporabljene metode dela morajo biti skladne z ravno izobrazbo ekipe. Torej lahko ekipe uporabijo vse znanje opisne statistike iz srednješolskih učbenikov. Dovoljeni so kakršnikoli preračuni ali



Slika 6: Število vodovodnih ali kanalizacijskih priključkov na 100 gospodinjstev sta statistična kazalnika. Zapisana ugotovitev nas preseneti. Ekipa je spregledala, da v večstanovanjskih stavbah več gospodinjstev uporablja isti priključek na javni vodovod oziroma kanalizacijsko omrežje. Vir: Kamni (2021).

tabelacije. Če vrednosti statističnih kazalnikov ekipe prikažejo še grafično, nimajo nikakršnih omejitev pri uporabi klasičnih grafičkonov ali modernejših infografik (Slika 6 in Slika 8).

² Možna so manjša odstopanja, če SURS v standardnem kazalniku uporabi podatek za dan, ki v podatkovnem nizu ni na voljo. Ekipe imajo na primer na voljo podatke o številu prebivalcev na začetku leta, SURS pa v standardnem kazalniku uporabi število prebivalcev na sredini leta.

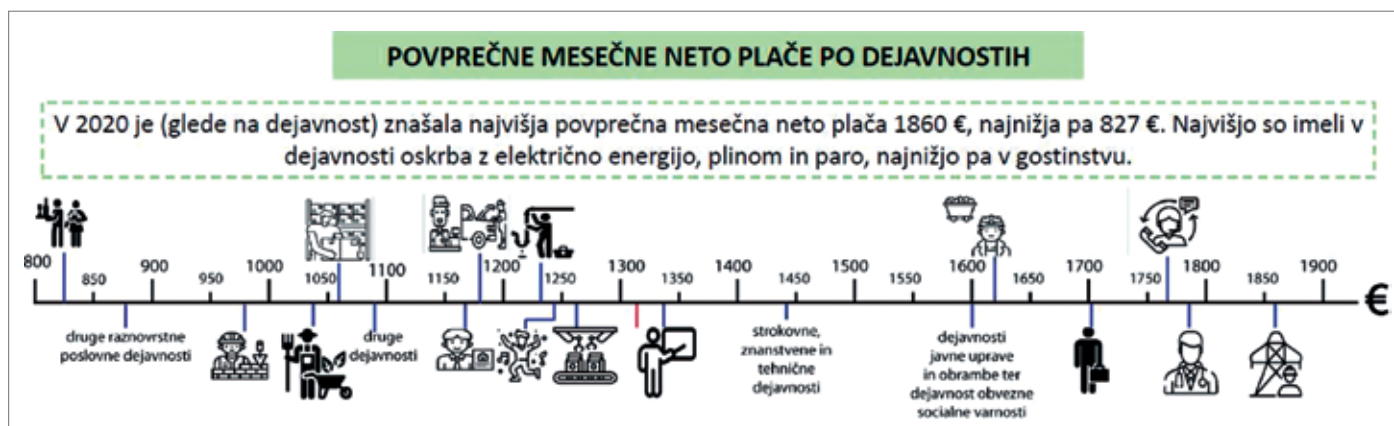
4. OKOLJSKI INDEKS

Okoljski indeks je kazalnik, izračunan na podlagi treh parametrov: količina odpadkov na prebivalca (O); količina porabljene vode na prebivalca (V); delež BDP, ki ga predstavljajo investicije za vlaganje v okolje (I). Pri vsakem parametru smo regije razporedili v 10 rangov (razlaga rangov na prosajnici 10), kjer rang 1 pomeni najboljši rezultat. Okoljski indeks E izračunamo tako, da od 33 odštejemo vsoto vrednosti vseh treh parametrov za dano regijo. Po tej metodi je največje število točk indeksa 30, najmanjše pa 3. Regije z višjim rezultatom indeksa veljajo za okolju bolj prijazne.

Največ otrok živi v osrednjeslovenski regiji (177426), najmanj pa v primorsko-notranjski (17247). Pearsonov koeficient korelacije kaže na to, da je število otrok močno povezano s številom vseh prebivalcev v regiji ($p=0,99$).

Slika 9: Pearsonov korelacijski koeficient, ki je izjemno blizu vrednosti 1, lahko nakazuje, da je preučevana zveza trivialna; v regijah z veliko prebivalci živi veliko otrok. Vir: Teglzarže (2020).

Slika 7: Opis izračuna in interpretacija zapletenega statističnega kazalnika. Vir: Marreb (2021).



Slika 8: Infografika prikazuje povprečne mesečne neto plače po dejavnostih v Sloveniji 2020. Raziskovalna naloga je vsebovala tudi legendo uporabljenih grafik. Vir: Avengers (2022).

Težave lahko nastopijo pri bivariatni statistiki (povezanost dveh spremenljivk). Ekipe pogosto izračunajo in interpretirajo **korelacijski koeficient**, pri čemer se ne zavedajo, da je to zgolj mera linearne povezanosti. Povezanost med dvema spremenljivkama lahko učinkoviteje preučijo in opišejo z razsevnim grafikonom. Včasih razsevni grafikon razkrije, da je preučevana zveza nelinearna ali trivialna (Slika 9).

Večja težava je uporaba **linearne regresije**. Prileganje premice množici točk je vsebina, ki jo srednješolci poznajo iz fizike, kljub

temu pa izračun in interpretacija determinacijskega koeficienta (R^2) običajno spadata med visokošolske statistične vsebine. Tam študenti spoznajo potrebne okoliščine, v katerih je uporaba regresije ustrezna (Slika 10). Uporabo metod, ki jih ekipe ne razumejo dovolj dobro, odsvetujemo.

Najbolj sporna je uporaba **sklepne statistike** (statistični testi), ki je v kontekstu tekmovanja nepotrebna. S sklepno statistiko raziskovalci ugotovijo, do katerih so prišli na vzorcu, posplošijo na celotno populacijo. Podatkovni niz pa ni vzorec, temveč vsebuje



Slika 10: Enostavna linearna regresija predpostavlja linearno povezanost spremenljivk (levo). Uporaba regresije pri deležih (desno) ni nujno smiselna. Prilegana premica zavzame tudi negativne vrednosti, ki pri deležu niso možne. Vir: zgoraj Gimvel (2020), spodaj Latte (2022).

		Delez_brezposelni	
Delez_brezposelni	Pearson Correlation	1	-.013
	Sig. (2-tailed)		.848
	N	212	212
Poraba antidepressivov, izražena v definiranih dnevni odmerkih na 1.000 prebivalcev na dan. Antidepressivi (N06A) se uporabljajo za zdravljenje epizod velike depresije in anksioznih motenj.	Pearson Correlation	-.013	1
	Sig. (2-tailed)	.848	
	N	212	213

Tabela 6.

Po obdelavi podatkov s programom IBM SPSS sva dobila rezultate (Tabela 6), da je Pearsonov koeficient negativen 0.13, kar kaže na veliko razpršenost podatkov z rahlo tendenco padanja. Vrednost testne statistike je statistično velika (Sig. je 0.848).

Slika 11: Uporaba statističnih programov lahko privede do izračunov, ki presegajo srednješolski okvir statistike. Posledično so možne napačne interpretacije. Vrednost Sig. (2-tailed) je narobe interpretirana. Vir: Statisti (2018).

podatke za prav vse enote populacije. Še huje je, če so rezultati sklepne statistike napačno interpretirani (Slika 11).

3.5 Izračuni

O matematičnih napakah v statističnih izračunih sem že pisal (Toman, 2022). Napake se nanašajo predvsem na računanje **povprečij** iz povprečij. Poglejmo si namišljen zgled, ki hkrati pojasni dve napaki. Preučevana regija naj ima le dve občini, v katerih je skupaj zaposlenih 5 oseb. **Tabela 1** prikazuje podatke o njihovih neto plačah v izbranem mesecu ter pripadajoča občinska povprečja.

Tabela 1: Podatki o plačah v izbranem mesecu v dveh namišljenih občinah.

Občina A		Občina B	
Oseba	Neto plača	Oseba	Neto plača
Andrej	1680 €	Bojan	1760 €
Ana	1920 €	Barbara	1848 €
		Branko	2464 €
Povprečje	1800 €	Povprečje	2024 €

Kako izračunamo povprečno mesečno neto plačo v regiji? Za pravo vrednost seštejemo plače vseh 5 zaposlenih in vsoto delimo s 5, torej $\mu = \frac{1680 + \dots + 2464}{5} = 1934,40$ €. Ekipa tako podrobnih podatkov nimajo na voljo, zato velikokrat izračunajo povprečje

objavljenih občinskih povprečij $\tilde{\mu} = \frac{1800 + 2024}{2} = 1912$ €, kar je narobe. Do pravilnega rezultata lahko pridejo, če objavljeni povprečji ustrezno **utežijo** sorazmerno s številom zaposlenih v posamezni občini $\mu = \frac{2 \cdot 1800 + 3 \cdot 2024}{2+3} = 1934,40$ €. Enaka zakonitost velja tudi pri združevanju deležev. V občini A delež žensk med zaposlenimi znaša 50 %, v občini B pa 33,3 %. S preštevanjem hitro ugotovimo, da je delež žensk med zaposlenimi v regiji enak $\pi = \frac{2}{5} = 40$ %. Če izračunamo le povprečje občinskih deležev $\tilde{\pi} = \frac{50 + 33,3}{2} = 41,7$ %, dobimo napačen rezultat. Do prave vrednosti lahko pridemo z uporabo uteženega povprečja $\pi = \frac{2 \cdot 50 + 3 \cdot 33,3}{2+3} = 40$ %. Zaradi pogostosti te napake so v zadnjih letih nekatera regionalna povprečja ali deleži na voljo v podatkovnem nizu in jih ekipam ni treba izračunati (ali pa za ustrezen izračun niti nimajo na voljo dovolj podatkov).

Izračunane vrednosti morajo ekipe tudi ustrezno interpretirati in zapisati ugotovitve. **Interpretacija** ni zgolj ubesedenje števil.

Nato smo preverili, kako sta povezana poročenost in imetje otrok in ugotovili, da je pri tipih družine brez otrok procentualno več zakonskih zvez, kot je zakonskih zvez pri tipih družine, kjer sta prisotna starša z otroki. Partnerji brez otrok so torej v večji meri poročeni kot partnerji z otroki.

Slika 12: Ekipa na zapleten način sporoča, kar enostavno povzame zadnja zapisana poved. Več ugotovitev ekipe je nakazovalo na težave pri razumevanju partnerjev brez otrok. V ta tip družine spadajo tudi pari, katerih otroci so se že osamosvojili. Vir: Yobama (2020).

Čeprav so opisi spremenljivk pogosto dolgi in birokratski, morajo biti ugotovitve zapisane v obliki preprostih in pravilnih povedi (Slika 12). Izrazi, kot so *odpadna regija*, *obrobna občina* ..., so lahko za prebivalce žaljivi. Tudi zapisom *najbolj izobražena regija* ali pa *največ odpadkov proizvede regija* se izogibamo. Izobražujejo se in odpadke proizvajajo ljudje v regijah in ne regije.

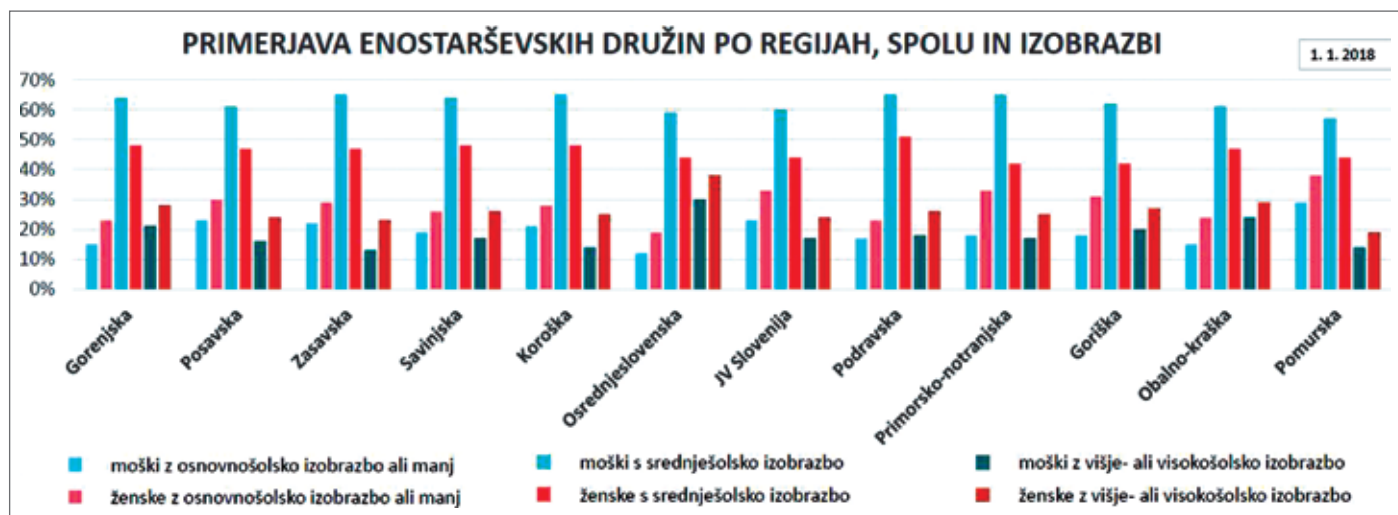
Zaključimo še s opozorilom pri opisovanju **sprememb**. Če količina odpadkov naraste z 10 ton na 11 ton, rečemo, da se je povišala za 10 %. Če stopnja brezposelnosti naraste z 10 % na 11 %, rečemo, da se je povišala za 1 odstotno točko.

3.6 Grafični prikazi

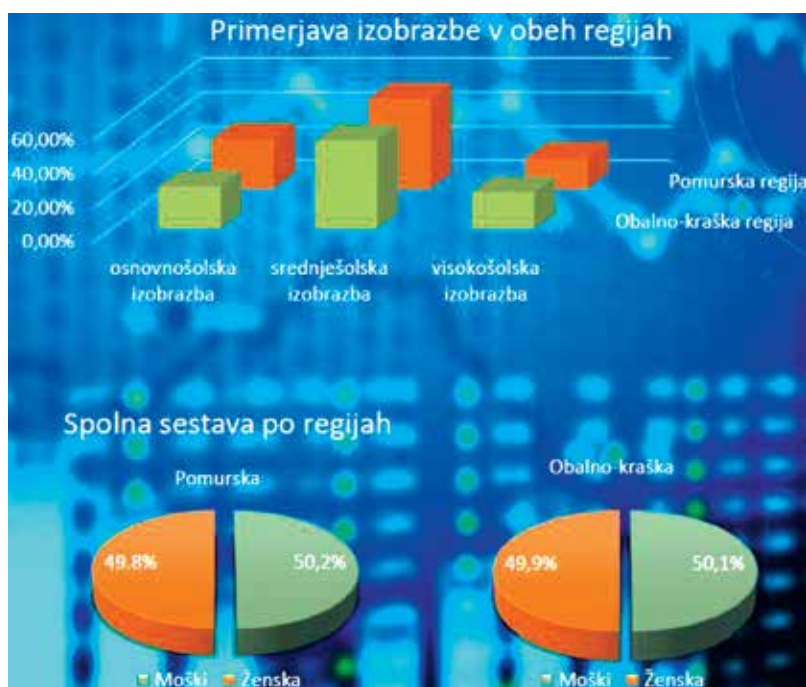
Tudi o prikazovanju podatkov sem že pisal (Toman, 2022). Ekipa z grafikoni primerjajo regije (ali občine) ali pa prikažejo dina-

miko pojava v času. Kadar prikazujejo vrednosti številske spremenljivke po regijah, najpogosteje izberejo **stolpčni ali vrstični grafikon**. Možna nerodnost pri takšnem grafikonu je izbira vrstnega reda regij. Abecedni vrstni red ali pa urejenost po kodah, ki jih uporablja SURS, običajno nista primerna (Slika 5 in Slika 17). Ker želijo primerjati regije med seboj, je grafikon nazornejši, če so regije urejene po naraščajočem ali padajočem vrstnem redu prikazane spremenljivke. Odločitev o vrstnem redu je težja, če ekipa s stolpci prikazuje več spremenljivk hkrati (Slika 13). Urejenost naj bo tedaj taka, da čim bolj podpre trditve, ki so zapisane v ugotovitvah. Stolpčni grafikon morajo biti dvorazsežni (Slika 14).

Pri prikazovanju prostorske razporeditve vrednosti spremenljivk ekipe pogosto uporabijo **tematske karte**. Pri tem je vsaka geografska enota na zemljevidu pobarvana v skladu s položajem



Slika 13: Kompleksen prikaz s stolpčnim grafikonom omogoča primerjavo enostarševskih družin po regijah (skupki stolpcev), spolu starša (barva) in izobrazbi starša (zasičenost barve). Izobrazba je urejenostna (ordinalna) spremenljivka – višja zasičenost barve ustreza višji stopnji izobrazbi. Spol je imenska (nominalna) spremenljivka – ker vrednosti nista urejeni, je uporabljena enaka zasičenost barve. Vir: Frekvenca9 (2020).



Slika 14: Stolpčni in tortna grafikona imajo odvečno tretjo dimenzijo. Pri stolpcih je zato težko določiti njihovo višino, tortna grafikona pa poudarita (a ne tu) tiste vrednosti, ki so prikazane v ospredju. Na sliki je moteče tudi ozadje, ki oteži branje in stilsko posnema grafične prikaze podatkov. Vir: Številopi (2018).



Slika 15: Primerjava geografskih enot s pomočjo tematske karte je zahtevnejša, kadar je geografskih enot veliko (zgoraj) ali pa je na isti karti prikazanih več spremenljivk (spodaj). Vir: zgoraj No1liga (2022), spodaj Amaterji (2020).



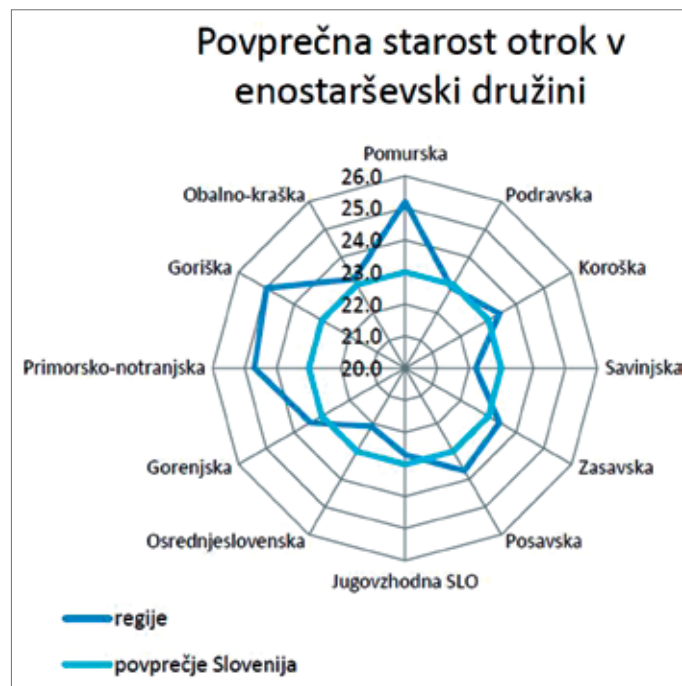
Slika 16: Dinamika pojava je lahko prikazana s stolpcnim ali črtnim grafikonom. Kombinacija obeh je koristna, kadar spremenljivki nimata podobnih vrednosti ali istih enot. Vir: Kalamari (2021).

njene vrednosti spremenljivke v barvni lestvici. Na eni karti je možno s kombinacijo barv in drugih elementov prikazati več spremenljivk, a je tak prikaz za bralca zahteven (Slika 15).

Kadar ekipe prikazujejo spreminjanje vrednosti ene ali več spremenljivk v času, najpogosteje izberejo črtni grafikonom. Pri tem je



Slika 17: Primerjava regij je lažja, če so urejene padajoče glede na prikazano spremenljivko. Ker med regijami ni logičnih povezav, uporaba črtnega grafikona ni primerna. Grafikonom uporablja dve navpični osi, pri tem oznake na desni osi niso povsem usklajene z vodilnimi črtami. Vir: Ekotistiki (2019).

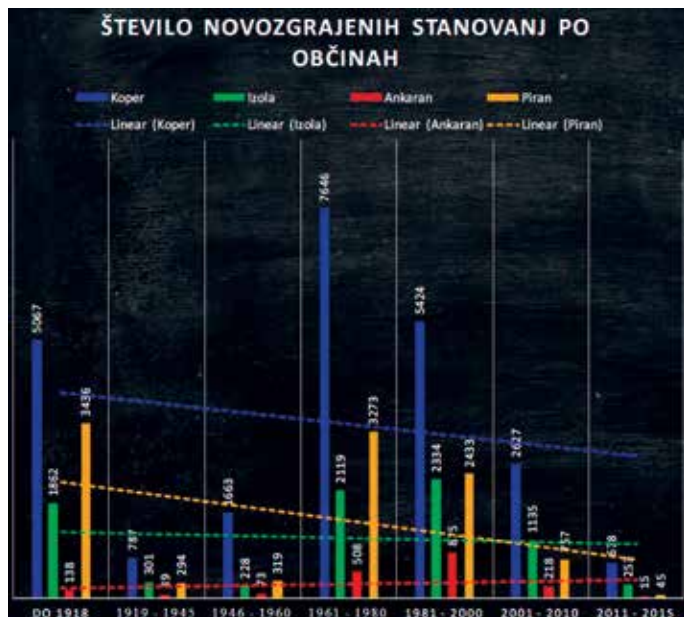


Slika 18: Polarni grafikonom je poseben primer črtnega grafikona, pri katerem mora biti smiselna tudi povezava prve in zadnje vrednosti. Pri regijah to ne drži. Vir: Llana (2020).

pomembno, da so razmiki na vodoravni osi sorazmerni dolžinam časovnih intervalov (soroden primer je prikazan na Slika 19). Opomnimo, da lahko ekipe časovno dinamiko pojava prikazujejo tudi s stolpci (Slika 16), vrednosti po regijah pa nikoli s črtnim grafikonom, saj črte, ki povezujejo zaporedne regije, nimajo pomena (in Slika 18).

Za konec omenimo še razliko med stolpčnim grafikonom in histogramom. Na vodoravni osi stolpčnega grafikona so nanešene vrednosti (ali kategorije), ki niso povezane. Stolpci so zato ločeni

s praznim prostorom. Na vodoravni osi **histograma** pa so enako široki razredi številske (numerične) spremenljivke. Višine stolpcov so frekvence razredov; te povedo število enot v posameznem razredu (Slika 20 in Slika 21).



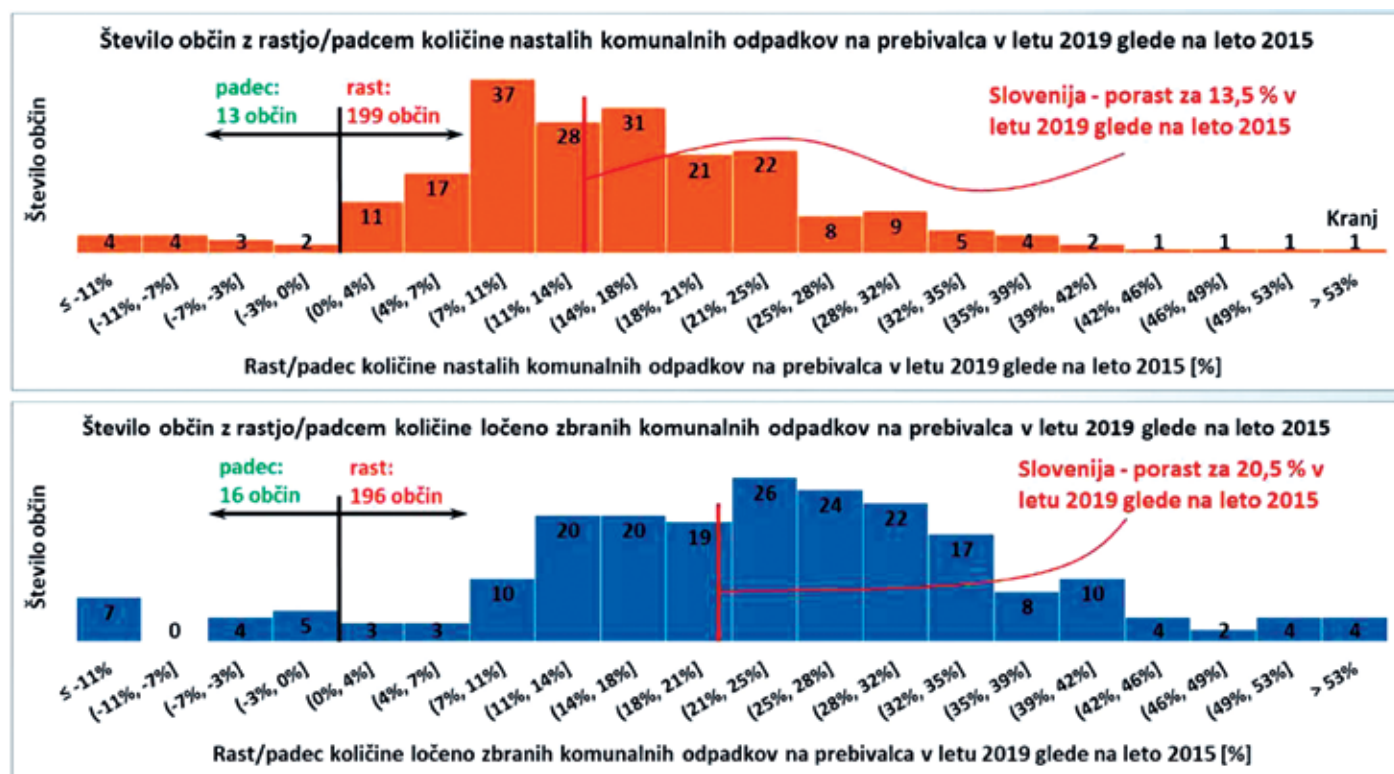
Slika 19: Pregled vodoravne osi razkrije, da časovna obdobja niso enako dolga. Višine stolpcov v različnih obdobjih zato niso primerljive, linearni trend je neustrezen. V prikazu najdemo še vsebinsko napako, saj grafikon v resnici prikazuje razvrstitev trenutnih stanovanj po letu izgradnje. Vir: Možganci (2019).

3.7 Ugotovitve in zaključki

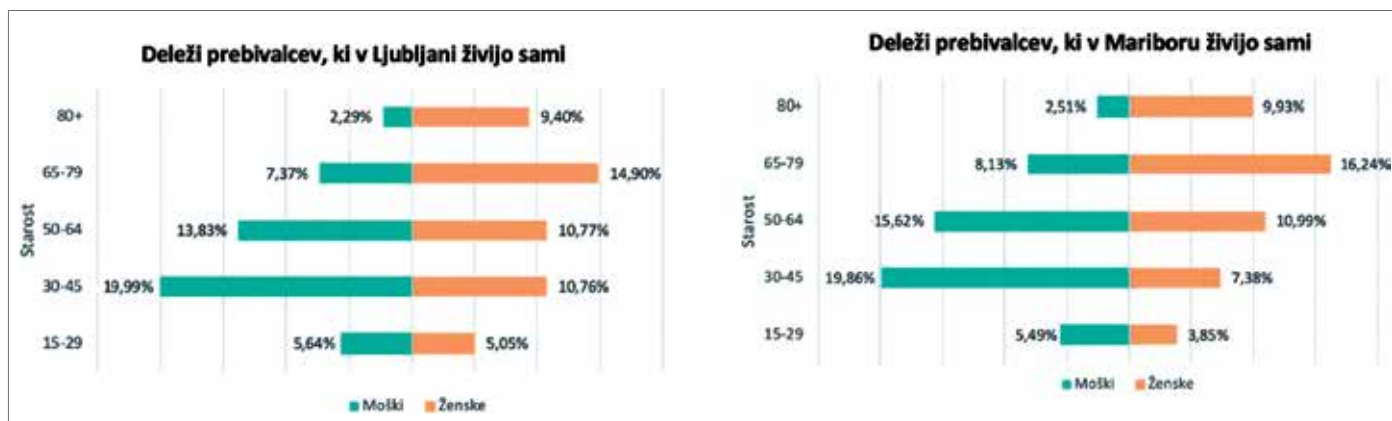
Podrobne ugotovitve je najbolje zapisati na prosojnici, na kateri je prikazana ustrezna analiza. Na zadnji vsebinski prosojnici lahko ekipe glavne ugotovitve povzamejo in jih interpretirajo v kontekstu zastavljenih ciljev. Če ekipa na začetku postavi hipoteze, je na koncu treba vsako hipotezo ovrednotiti. Kriteriji ocenjevanja spodbujajo razmislek o omejitvah podatkovnega niza. O tem razmišlja in poroča le malo ekip. Pri pregledovanju ugotovitev in zaključkov žirija ponovno pregleda celotno nalogo in preveri, ali so bili vsi zastavljeni cilji doseženi (vsak cilj mora imeti svoj zaključek) in ali je vsaka ugotovitev podprta z vmesnimi izračuni.

3.8 Prikazi izračunov

Raziskovalne naloge morajo biti **ponovljive**. Avtorji nalog morajo zato navesti dovolj informacij, da lahko bralci, ki razpolagajo s podatki, ponovijo vse izračune in preverijo vse trditve, ki so zapisane v nalogi. Da tehnične podrobnosti ne zameglijo vsebine analize, sta ekipam na voljo dodatni prosojnici za predstavitev izračunov. Natančnost in izčrpnost sta tu pomembnejši od estetike. Zaželeno je, da ekipe v opisih uporabljajo imena spremenljivk, kot so na voljo v podatkovnem nizu (Slika 22).



Slika 20: Histogram prikazuje porazdelitev občin glede na odstotno spremembo količine odpadkov. Razredi so matematično označeni, vendar niso enako široki. V večini občin se izboljšuje odnos do okolja, ker pa se občine razlikujejo po številu prebivalcev, je preštevanje občin lahko zavajajoče. Vir: Gimvel (2021).



Slika 21: Prebivalstvena piramida je posplošitev histograma. Ker so starostni razredi stični, prostor med vrsticami ni potreben. Različno široki razredi so bili določeni v podatkovnem nizu, razred 30–49 je na sliki neustrezno označen. Vir: 20in20 (2020).

IZRAČUNI

- 1. Komunalni odpadki/prebivalca:
 - Uporabila sva podatke *Odpadki_1*, *Odpadki_33_1* ki sva jih delila s podatki *Preb_tot* v letu 2019. Od skupne količine komunalnih odpadkov na prebivalca sva odštela količino nevarnih komunalnih. Rezultate sva za vsako regijo posebej prikazala s stolpčnim diagramom.
- 2. Komunalni odpadki med 2015-2019
 - Uporabila sva podatke *Odpadki_1* in jih delila s podatki *Preb_tot* v letih 2015 in 2019. Štiriletno spremembo sva za vsako regijo posebej predstavila v odstotkih in jo prikazala na zemljevidu

Prikaz izračunov:

Delež tipov družin v Sloveniji/ JV Sloveniji:

- $\frac{(Druz_tip1 + Druz_tip5)}{Druz_tot} \cdot 100$
- $\frac{(Druz_tip2 + Druz_tip6)}{Druz_tot} \cdot 100$
- $\frac{(Druz_tip3 + Druz_tip4)}{Druz_tot} \cdot 100$

Odstotek prebivalcev, ki živijo sami (Slovenija/JV Slovenija):

Izobrazba:

- družine brez otrok: $\frac{Druz_11_izb \cdot n}{Druz_11_tot} \cdot 100$ *n=1, 12, 21, 13, 31, 2, 2, 32, 23, 3, 3
- enostarševske družine:
 - mati: $\frac{Z_druz_33_izb \cdot m}{Z_druz_33_tot} \cdot 100$ *m=1, 2, 3

Slika 22: Opis izračunov v opisni (zgoraj) in matematični (spodaj) obliki. Oba opisa omogočata ponovitev opravljene analize. Vir: zgoraj Koale (2021), spodaj Lnmstorm (2020).

Zaključek

V šolskem letu 2021/22 so se uspešno zaključile 5. Evropske statistične igre. Kljub mladosti je tekmovalstvo popularno, znano in priznано v Sloveniji in Evropi in je uspešno premagalo vse ovire, ki jih je prinesla epidemija. Slovenske ekipe se skoraj vedno uvrstijo med najboljše v Evropi, kar utrjuje prepričanje, da smo na pravi poti, in nam daje neizmerno motivacijo za nadaljnje korake. Poleg tekmovalnih uspehov pa Evropske statistične igre prinašajo še spodbudo mladim in njihovim mentoricam in mentorjem k pridobivanju novih statističnih (in interdisciplinarnih) znanj in preučevanju naravnih in družbenih pojavov. Tekmovalstvo s tem vzgaja odgovorne in mednarodno vpete državljanke in državljane, ki znajo svoje odločitve in mnenja oblikovati in podpreti s pomočjo podatkov uradne statistike.

Več informacij o aktualni izvedbi iger je na voljo na naslovih <https://www.stat.si/igre> in statigre.surs@gov.si.

Viri in literatura

- Rauter Repija, I. in Godec, M.; Karas, T., in Štampar, N. (2021). Evropske statistične igre. *Matematika v šoli* 27(2), str. 45–51.
- Toman, A. (2022). Evropske statistične igre in spodbujanje statistične pismenosti pri matematiki. *Zbornik 1. mednarodne konference o poučevanju matematike, fizike in astronomije*, str. 72–78. Dostopno na https://www.dmfa.si/_CmsFiles/2022/08/zbornik-1-konferenca-dmfa-2022.pdf
- Making Data Meaningful (2009). Part 2: A Guide to Presenting Statistics. United Nations Economic Commission for Europe. Dostopno na http://www.unecp.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf
- 20in20 (2020). Ali se Slovenci odločamo za družinsko življenje? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/10877/20IN20.pdf>
- Amaterji (2020). Izobrazba v slovenskih družinah. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/10879>
- Avengers (2022). Trg dela, Slovenija 2020. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11827>
- Deadsirius (2021). Voda. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11341>
- Ekotistiki (2019). Analiza stanovanjske problematike v Sloveniji.
- Frekvenca9 (2020). Slovenija, dežela ločitev in visoke izobrazbe? Dostopno na <http://www.stat.si/StatWeb/File/DocSysFile/10883>
- Gimvel (2021). Ali je Slovenija res tako čista država? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11346>
- Kalamari (2021). Odpadki v Sloveniji. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11347>
- Kamni (2021). Uporaba in pridobivanje voda v letu 2018. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11348>
- Kapljice (2022). Ali na zaposlitev v gospodarskih dejavnostih vplivajo migracije ter višina plače ali je ravno obratno? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11834>
- Koale (2021). Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11349>
- Latte (2022). Ali je Slovincem pomembnejši čas ali denar? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11837>
- Llama (2020). Družine v Sloveniji. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/10886>
- Lnmstorm (2020). Družine v Jugovzhodni Sloveniji 2018. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/10887>
- Marreb (2021). Primerjava statističnih regij glede na odnos do okolja. Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11350>
- Možganci (2019). Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/10382>
- NoIliga (2022). Neenakost med delovno aktivnim prebivalstvom v Sloveniji. Kje v Sloveniji je najbolje delati? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11840>
- Plan_xy (2021). Kako Slovenci ravnajo z vodo? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11353>
- Statisti (2018). Analiza statističnih podatkov po občinah v Sloveniji za Evropske statistične igre
- Statisti1 (2021). Slovenija = Ekološka? Dostopno na <https://www.stat.si/StatWeb/File/DocSysFile/11356>
- Številopi (2018). Evropske statistične igre 2018.
- Teglarože (2020). Slovenske družine. Dostopno na <http://www.stat.si/StatWeb/File/DocSysFile/10893>
- Yobama (2020). (Zunaj)zakonski partnerji, njihova izobrazba ter otroci. Dostopno na <http://www.stat.si/StatWeb/File/DocSysFile/10896>