

Univerza v Ljubljani
Fakulteta *za upravo*



SEMINARSKA NALOGA

STATISTIKA

Statistična populacija: **Gorenjska, Preddvor**
Številka naloge: **O-10**

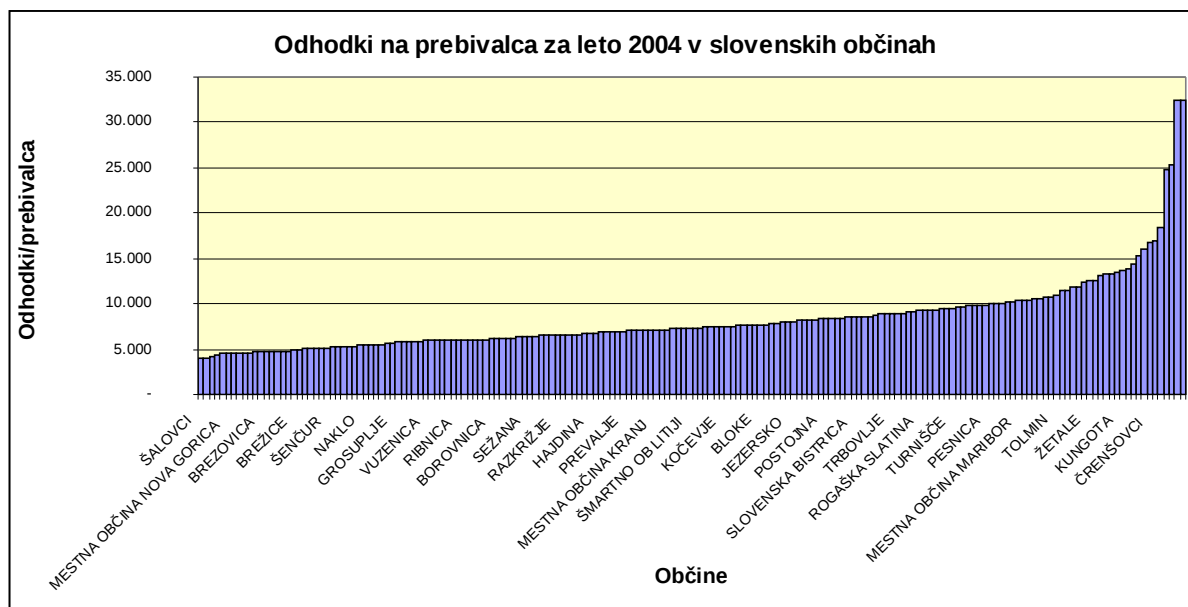
Predmet: **Statistika**
Mentor:
Študent:
Skupina vaj: **četrta**

Ljubljana, 13. januar 2006

1. VAJA

POSAMIČNI PODATKI

Grafikon 1: Odhodki za socialno varnost na prebivalca za leto 2004 v slovenskih občinah



KOMENTAR:

Zgornji grafikon prikazuje odhodke za socialno varnost na prebivalca po posameznih občinah za leto 2004. Iz grafa je lepo razvidno, da imajo skoraj vse občine približno enako odhodkov na prebivalca ali pa minimalno naraščajo. Izstopajo samo zadnje štiri oz. zadnjih pet občin. To so občina Krško, Kostel, Ormož, Ruše in Pivka (prikazane od desne proti levi po vrsti). Te občine imajo zelo visoke odhodke za socialno varnost na prebivalca v letu 2004. Vrednost koeficienta občine Krško, ki ima največji koeficient odhodkov na prebivalca, znaša kar 32 525. Koeficient občine Starše pa znaša 2560 in je tudi najmanjši med vsemi slovenskimi občinami v letu 2004. Za občino Kanal v letu 2004 pa ni podanih podatkov, zato sem jo izključila iz nadaljnje obravnave in je tudi ni prikazane na grafikonu.

Tabela 1: Pridobljeni rezultati z obdelavo posamičnih podatkov za proučevano spremenljivko

vsota	1.541.776
število enot	192
Mediana	7166,031691
prvi kvartil	5741,856072
tretji kvartil	9247,067159
deveti decil	11793,31055
prvi decil	4572,015093
kvantilni rang za aritmetično sredino	0,702
koeficient asimetrije	3,01

koeficient sploščenosti	13,33
aritmetična sredina	8.635
Povprečni absolutni odklon od AS	2863,300785
Varianca	17.775.751
standardni odklon	4216,129824

KOMENTAR:

Vsota odhodkov vseh občin na prebivalca v letu 2004 je **1.541.776**, kar pove, da je vseh 192 občin v letu 2004 skupno porabilo 1.541.776 odhodkov za socialno varnost, če bi v vsaki občini vsak od prebivalcev porabil enako odhodkov in če bi imela vsaka občina samo 1 prebivalca, ki bi predstavljal vse občane z enakimi vrednostmi spremenljivke.

Število vseh občin v letu 2004 je 192, ker za občino Kanal ni danih podatkov in sem jo zato izločila iz nadaljnje analize.

Mediana je **7166,03**. Mediana pa je tista vrednost številske spremenljivke, za katero velja, da ima $\frac{1}{2}$ enot spremenljivke manjšo ali večjemu enako medialni vrednosti. Iz tega lahko rečem, da ima **polovica slovenskih občin v letu 2004 manjšo ali večjemu enako 7166,03 odhodkov na prebivalca, ostala polovica občin pa ima večjo porabo odhodkov na prebivalca.**

Prvi kvartil proučevane spremenljivke znaša **5741,86** in nam pove, da ima **25% slovenskih občin 5741,86 ali manj odhodkov na prebivalca v letu 2004.**

Tretji kvartil nam pove, da ima **75% vseh slovenskih občin v letu 2004 9247,07 ali manj odhodkov na prebivalca.**

Deveti decil pa nam pove, da ima v letu 2004 **90% vseh slovenskih občin 11793,31 ali manj odhodkov na prebivalca.**

Prvi decil nam kaže, da je v letu 2004 imelo **10% vseh proučevanih občin 4572,02 ali manj odhodkov na prebivalca.**

Aritmetična sredina koeficientov vseh 192 proučevanih občin je 8.635. Srednja vrednost je predstavnik »parameter« populacije in na ta način omogoča primerjavo med posameznimi populacijami. Srednje vrednosti pokažejo lastnost statistične populacije in njene porazdelitve. Kvantilni rang za aritmetično sredino nam pove v deležu, na katerem delu ranžirnega razmika leži določena enota, ki ji pripada rang R. Torej lahko rečem, da je **70,2% vseh proučevanih občin v letu 2004 imelo večjemu enako 8635 odhodkov na prebivalca v letu 2004, ostalih 29,8% občin pa je imelo več kot 8635 odhodkov na prebivalca.** Iz teh podatkov lahko sklepam, da imajo nekatere občine ekstremno velike odhodke na prebivalca, tako da se aritmetična sredina temu primerno toliko zviša, da zavzame skoraj tri četrtine vseh proučevanih občin. Ta sklep lahko tudi preverimo v podatkih in ugotovimo, da ima 5 občin ekstremne koeficiente.

Pri odklonih opazujemo za koliko se posamične vrednosti razlikujejo od določene srednje vrednosti. Povprečni absolutni odklon je povprečje absolutnih odklonov vseh individualnih vrednosti od neke srednje vrednosti. Pri aritmetični sredini pa je pomembno to, da so odkloni individualnih vrednosti od aritmetične sredine po definiciji enaki nič. Zato se odkloni individualni vrednosti od srednjih vrednosti računajo v absolutni vrednosti. S tem postopkom dobimo novo srednjo vrednost, in sicer srednjo vrednosti med odkloni. **Povprečni absolutni odklon od aritmetične sredine znaša v mojem primeru 2863,30. To pa pomeni, da je bilo odklanjanje koeficientov od aritmetične sredine v letu 2004 v povprečju 2863,30 odhodkov na prebivalca navzgor in navzdol.**

Varianca ali povprečni kvadratni odklon temelji na odklonih individualnih vrednosti od srednjih vrednosti (v mojem primeru samo od aritmetične sredine). Pri varianci odklone kvadiramo, medtem, ko pri povprečnem absolutnem odklonu upoštevamo njihove absolutne vrednosti. Kvadirane vrednosti nato seštejemo in delimo s številom proučevanih občin. Tako dobimo vrednost variance, ki znaša **17.775.751**, ki pa je izražena v kvadratu osnovne enote. Zaradi boljše primerljivosti pa tudi lažje razlage rezultata iz variance izračunamo kvadratni koren in tako dobimo **standardni odklon**, ki v mojem primeru znaša **4216,13**. To pa pomeni, da v primeru, če bi imeli normalno porzdelitev podatkov (pri nas je asimetrična v desno), bi lahko vrednost 4216,13 prišteli in odšteli od povprečja in bi tako dobili interval, v katerega bi padlo 2/3 enot statistične množice..

Koeficient asimetrije je pozitiven in znaša **3,01**. **Koeficient nam pove, da je bila oblika posamičnih podatkov za občine v letu 2004 asimetrična v desno**. Se pravi, če povem drugače, je bilo v letu 2004 zelo veliko občin, ki so imele nizke dohodke na prebivalca in malo občin, ki so imele visoke dohodke na prebivalca.

Simetrične in asimetrične porazdelitve pa so lahko tudi bolj ali manj sploščene od normalne. To ugotavljamo s **koeficientom sploščenosti**, ki pri moji analizi občin znaša **13,33**. Pri tem je pomembna ugotovitev, da je koeficient večji od ena. Iz te ugotovitve sledi trditev, da **porazdelitev** proučevanih občin bolj **sploščena** od normalne.

GRUPIRANI PODATKI

Tabela 2: Pomembni podatki pri določevanju razredov

maksimalna vrednost	25.256
minimalna vrednost	2.560
število razredov	9

KOMENTAR:

Maksimalna vrednost odhodkov na prebivalca v letu 2004 izmed proučevanih občin je znašala **25.256** odhodkov na prebivalca, in sicer je bila to občina Ormož. Občini Kostel in Krško sta sicer imeli še večjo vrednost, toda zaradi praznih razredov pri določevanju neenako širokih razredov sem ekstremni vrednosti za obe omenjeni občini zanemarila in privzela, da ima največjo vrednost občina Ormož. Tako sem se izognila praznim razredom. Pri nadaljnji obravnavi in analizi pa sem upoštevala samo 190 občin (za občino Kanal ni bilo podatkov).

Minimalna vrednost v letu 2004 izmed 190 občin je znašala **2560** odhodkov na prebivalca in to vrednost je imela občina Starše.

Posamične podatke sem razporedila v **9 neenako širokih razredov**.

ENAKO ŠIROKI RAZREDI

Tabela 3: Določitev enako širokih razredov

širina razreda		2521,7448
spodnja	zgornja	f_j
2559,87	5081,6	35
5081,6	7603,4	77
7603,4	10125,1	47
10125,1	12646,8	17
12646,8	15168,6	7
15168,6	17690,3	4
17690,3	20212,1	1
20212,1	22733,8	0
22733,8	25255,6	1

KOMENTAR: Razdelitev v enako široke razrede ni v redu, ker dobimo prazen razred in frekvence preveč nihajo.

NEENAKO ŠIROKI RAZREDI

Tabela 4: Določitev neenako širokih razredov in njihova frekvenca

količnik razredov		1,289615
spodnja	zgornja	f_j
2559,9	3301,2	3
3301,2	4257,3	12
4257,3	5490,3	29
5490,3	7080,4	48
7080,4	9131,0	51
9131,0	11775,4	28
11775,4	15185,8	12
15185,8	19583,8	5
19583,8	25255,6	2
vsota		190 ¹

KOMENTAR:

Absolutna frekvenca pove, koliko proučevanih občin je razporejenih v določen velikostni razred. Ker so v mojem primeru različne širine razredov, pravzaprav ne morem primerjati frekvence med seboj in nato podati trditev, kje je največ enot. Če bi hotela to povedati, bi morala izračunati gostoto posameznega razreda.

Neenako široki razredi se tudi grafično drugače prikazujejo. Npr. pri stolpcih bi morala risati različno širino, glede na to, kolikšna je dolžina posameznega razreda.

Lahko pa rečem, da je bilo v letu 2004, glede na zgornje širine razredov, največ občin, ki je imelo odhodke za socialno varnost na prebivalca med 4257,3 in 11775,4. Toda s to

¹ Pri razvrščanju v neenako široke razrede je bil še vedno razred s frekvenco 0. Zato sem dve ekstremno visoki vrednosti ignorirala. Ena občino pa sem izločila že takoj na začetku, ker ni imela podane vrednosti kazalnika.

trditvijo ne povem veliko. Za bolj natančno opredelitev, bi morala izračunati gostoto, kot sem to že omenila zgoraj.

Vseh občin je skupno 190 in so porazdeljene glede na njihove koeficiente v 9 neenako širokih razredov.

Tabela 5: Aritmetična sredina, povprečni absolutni odklon, varianca in standardni odklon za grupirane podatke

aritmetična sredina	7877,86
povprečni absolutni odklon	2395,592
varianca	386459,9
standardni odklon	621,659

KOMENTAR:

Tehtana aritmetična sredina je **7877,86** in le ocena prave vrednosti, ki sem jo izračunala iz posamičnih podatkov. Aritmetična sredina pove, da bi **ob domnevi, da bi imele vse proučevane občine v letu 2004 enako odhodkov na prebivalca, bi bila ta vrednost enaka 7877,86 odhodkom na prebivalca.**

Povprečni absolutni odklon za grupirane podatke znaša **2395,59**, to pa pomeni, da bi se **odhodki na prebivalca za proučevane občine v letu 2004 od aritmetične sredine v povprečju odklanjali 2395,59 navzgor ali navzdol.**

Varianca je **386459,9**, vendar je to le približek prave vrednosti variance, ker imamo podatke razporejene v razrede. Pravo vrednost variance namreč dobimo samo, če jo izračunamo iz posamičnih podatkov. Pri računanju pa bi morali upoštevati tudi Sheppardov popravek, toda zaradi neenako širokih razredov to ni mogoče. Tako da je napaka pri varianci še nekoliko večja.

Standardni odklon je znašal v letu 2004 **621,66** in je približek, ker ga izračunamo iz približne variance kot sem omenila zgoraj.

2. VAJA

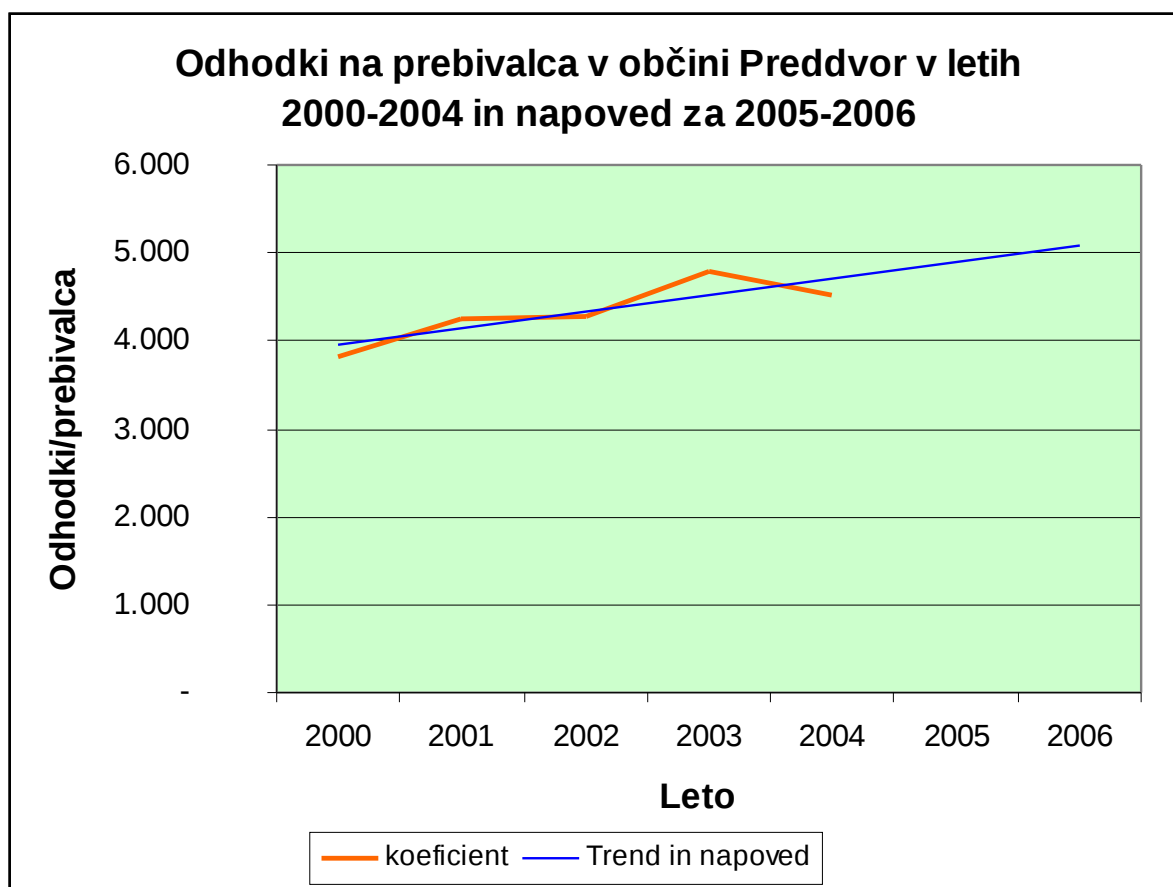
Tabela 6: Odhodki za socialno varnost v letih 2000-2004 ter trend in napoved za občino Preddvor

leto	koeficient	trend in napoved
2000	3.829	3946,7
2001	4.241	4136,2
2002	4.272	4325,6
2003	4.779	4515,1
2004	4.507	4704,6
2005		4894,0
2006		4906,6

KOMENTAR:

Koeficienti med leti 2000-2003 lepo naraščajo. Samo leta 2002 se vrednost odhodkov na prebivalca za občino Preddvor spremeni minimalno. Leta 2004 pa vrednost koeficienta nekoliko pade. Iz danih podatkov sem izračunala linearni trend. Nato pa sem z metodo ekstrapolacije izračunala še napoved gibanja koeficienta v letih 2005 in 2006.

Grafikon 2: Odhodki za socialno varnost na prebivalca v občini Preddvor v letih 2000-2004 in napoved za 2005-2006



SPLOŠNO:

Pri časovnih vrstah je vnašanje podatkov v linijski grafikon odvisno od vrste podatkov, ki jih prikazujemo. **Trenutne podatke** vnašamo v grafikon **točno ob času**, na katerega se nanašajo. Pri **razmičnih podatkih** pa vnašamo vrednosti na **sredino obdobja**.

KOMENTAR:

Pri zgornjem grafikonu so podatki vneseni na sredino obdobja, kar ni pravilno. V mojem primeru bi pri letih morale biti točke narisane na koncu obdobja, ker predstavljajo interval in gledamo preko celega leta. Rdeča črta prikazuje gibanje odhodkov za socialno varnost na prebivalca v obdobju 2000 do 2004 za mojo občino Preddvor. Opazimo lahko, da koeficient narašča do leta 2003, nato pa v obdobju 2004 nekoliko pada. Leta 2002 koeficient skoraj nič ne naraste v primerjavi z letom 2001.

Pri določanju trenda sem uporabila metodo najmanjših kvadratov. Pri tej metodi so odkloni od trenda v posameznih časovnih enotah lahko pozitivni ali negativni, njihova vsota pa je po definiciji enaka nič. Vsota kvadratov odklonov mora biti minimalna. Šele tista funkcija, ki ustreza omenjenim pogojem, je funkcija trenda. Ker so naši podatki časovni, je pomembno tudi to, da je časovno obdobje med vsakim koeficientom enako; v našem primeru je to 1 leto. Linearni trend, ki sem ga izračunala je opredeljen z linearno funkcijo $T_t = \alpha + \beta * x_t$. Linearna funkcija se upošteva samo v primerih, ko pojav enakomerno narašča ali pada (kar pa v mojem primeru velja).

Z metodo ekstrapolacije na osnovi enačbe linearnega trenda $T = 189,46x + 3757,3$ sem napovedala gibanje v letu 2005 in 2006. Ta metoda temelji na domnevi, da bo razvoj v prihodnosti enak razvoju v preteklosti. Zato sem lahko enačbo trenda, ki sem jo izračunala za določeno časovno obdobje v preteklosti, uporabila za napovedovanje prihodnjega razvoja opazovanega pojava. Pri sami enačbi trenda pa je pomemben predvsem parameter β , ki pove za koliko se funkcija spremeni v časovni enoti. V mojem primeru sem β izračunala s pomočjo programa Microsoft Excel in znaša **189,46**, kar pomeni, da se povečajo odhodki na prebivalca izraženi z y letno za **189,46**.

V letu 2005 naj bi torej, glede na našo napoved, občina Preddvor imela približno **4700 odhodkov za socialno varnost na prebivalca**.