



STATISTIČNA NALOGA

Študijsko leto 2011/2012

Nosilka: doc. dr. Katja Lozar Manfreda

P., G.

novinarstvo

redni 1P

OCENJEVANJE NALOGE (vpisujejo ocenjevalci)

ODDAJA Datum: _____ Točke: _____

Opombe

POPRAVA Datum: _____ Točke: _____

Opombe

Datum zagovora: _____ Podpis nosilke: _____

KONČNO ŠTEVILO TOČK: _____

IZVAJA O AVTORSTVU

statistične naloge

Spodaj podpisani/-a _____, z vpisno številko _____, sem avtor/-ica pričujoče statistične naloge pri predmetu Statistika v študijskem letu 2009/10.

S svojim podpisom zagotavljam, da:

- je predložena statistična naloga izključno rezultat mojega lastnega dela;
- sem poskrbel/-a, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem besedilu, navedena oz. citirana v skladu fakultetnimi navodili ter navedena v seznamu virov, ki je sestavni element tega besedila;
- se zavedam, da je plagiatorstvo – predstavljanje tujih del, v katerem so tuji izračuni, interpretacije oz. ideje predstavljane kot moje lastne – kaznivo po zakonu (Zakon o avtorstvu in sorodnih pravicah, Uradni list RS št. 21/95), prekršek pa podleže tudi ukrepom Fakultete za družbene vede v skladu z njenimi pravili;
- se zavedam posledic, ki jih dokazano plagiatorstvo lahko predstavlja za predloženo delo in za moj status na Fakulteti za družbene vede.

V Ljubljani, dne _____ Podpis avtorja/-ice _____

ODDAJNI LIST

G. P. - 05.6.2012

ENV.tax.2006: (Total environmental tax revenues as a share of GDP (%))

LabourGDPinPPS.2006: (Labour productivity per person employed, GDP in Purchasing Power Standards (PPS) per person employed relative to EU-27 (EU-27 = 100))

1. podnaloga:
Opisna analiza podatkov

Vprašanje	Odgovor študenta	Pravilno?
Aritmetična sredina prve spremenljivke	2,7356	✓
Standardni odklon prve spremenljivke	0,8350	✓
Rang Slovenije	22	✓
Prvi kvartil	2,265	✓
Drugi kvartil	2,58	✓
Tretji kvartil	2,9825	✓

2. podnaloga:

Ocenjevanje populacijskih parametrov in preverjanje domnev na osnovi vzorca

Vprašanje	Odgovor študenta	Pravilno?
Vzorčna aritmetična sredina	3,018	✓
Vzorčna varianca	1,4071	✓
Vzorčni standardni odklon	1,1862	✓
Spodnja meja intervala zaupanja za aritmetično sredino	2,1702	✓
Zgornja meja intervala zaupanja za aritmetično sredino	3,8657	✓
Vzorčna aritmetična sredina za stare članice	2,9933	✓
Vzorčni standardni odklon za stare članice	1,5837	✓
Vzorčna aritmetična sredina za nove članice	3,055	✓
Vzorčni standardni odklon za nove članice	0,19527	✓
Skupni standardni odklon	1,2577	✓
Ekperimentalna vrednost testne statistike t	-0,0760	✓

3. podnaloga:

Analiza povezanosti nominalnih spremenljivk

Vprašanje	Odgovor študenta	Pravilno?
stare članice EU / enako ali manjše od mediane	8	✓
nove članice EU / enako ali manjše od mediane	6	✓

stare članice EU / večje od mediane	7	✓
nove članice EU / večje od mediane	6	✓
stare članice EU / enako ali manjše od mediane	7,7777	✓
nove članice EU / enako ali manjše od mediane	6,2222	✓
stare članice EU / večje od mediane	7,2222	✓
nove članice EU / večje od mediane	5,7777	✓
Pearsonov hi-kvadrat	0,0296	✓
Cramerjev alpha	0,0331	✓
Pearsonov koeficient kontingence C	0,0331	✓
Popravljeni Pearsonov koeficient kontingence	0,0468	✓
stare članice EU / enako ali manjše od mediane	53,33	✓
nove članice EU / enako ali manjše od mediane	50	✓
stare članice EU / večje od mediane	46,66	✓
nove članice EU / večje od mediane	50	✓

4. podnaloga:

Analiza povezanosti in odvisnosti dveh razmernostnih spremenljivk

Vprašanje	Odgovor študenta	Pravilno?
aritmetična sredina prve spremenljivke (Y)	3,018	✓
aritmetična sredina druge spremenljivke (X)	95,81	✓
r_{xy}	-0,0942	✓
eksperimentalna vrednost t	-0,2676	✓
konstanta a	3,3917	✓
konstanta b	-0,0039	✓
determinacijski koeficient (R)	0,00887	✓
standardna napaka ocene	1,1809	✓

1. PODNALOGA: OPISNA ANALIZA PODATKOV

1.1 Uvod

Cilj: Cilj prve podnaloge je analizirati spremenljivko celotnih okoljskih davčnih prihodkov, izraženo v odstotku BDP. Za spremenljivko bom izračunal aritmetično sredino ter standardno odklon, nato pa bom razporedili države v ranžirno vrsto, ter izračunal prvi, drugi in tretji kvartil. Cilj pa je tudi, da analizo tudi vsebinsko in pomensko predstavim, torej da ne podam zgolj števil, temveč pojasnim, kaj te številke pomenijo v praksi in kaj lahko iz njih sklepam. Cilj pa je tudi to, da bolje spoznam osnovne statistične pojme in izraze.

Vir podatkov: Eurostat. Dostopno prek: <http://epp.eurostat.ec.europa.eu/> (24. maj 2012).

Prva spremenljivka: ENV.tax.2006 (celotne pobrane okoljske dajatve (davki), kot delež BDP, izražene v odstotkih).

Opredelitev: Ta spremenljivka nam pove, koliko okoljskih dajatev je posamezna država pobrala, te pa so izražene kot BDP te države in sicer v odstotkih. Preprosteje povedano: koliko odstotkov so pobrani okoljski davki prispevali k celotnemu BDP-ju posamezne države. Okoljska dajatev je sicer davek, katerega davčna osnova je fizična enota za nekaj, kar ima dokazan poseben negativni vpliv na okolje. Skupni prihodki za okoljske dajatve v EU so davki na promet, energijo, onesnaževanje in raziskovanje.

Vir: Eurostat. Dostopno prek: <http://epp.eurostat.ec.europa.eu/> (24.maj 2012).

Populacija in enote: Populacija je 27 držav članic Evropske unije v letu 2006, enote pa so posamezne države.

Prikaz podatkov:

Tabela 1.1: Pobrane okoljske dajatve v letu 2006, izražene v odstotku BDP posamezne države EU.

Država	ENV.tax.2006
Avstrija	2,5
Belgija	2,15
Bolgarija	3,1
Ciper	3,29
Češka republika	2,58

Danska	6,17
Estonija	2,22
Finska	2,99
Francija	2,26
Grčija	1,99
Irska	2,5
Italija	2,7
Latvija	2,4
Litva	1,81
Luksemburg	2,61
Madžarska	2,82
Malta	3,36
Nemčija	2,4
Nizozemska	4,03
Poljska	2,76
Portugalska	2,96
Romunija	1,94
Slovaška	2,28
Slovenija	3,01
Španija	1,87
Švedska	2,76
Velika Britanija	2,4

Vir: Eurostat 2012.

1.2 Analiza populacije držav EU

Ranžirna vrsta in kvartili

Definicije: Namen razporeditve vrednosti v ranžirno vrsto je, da uredimo enote po spremenljivkah, od tiste z najmanjšo, do tiste z največjo vrednostjo. Na ta način dobimo boljši pregled in lažje urejamo enote (glede na spremenljivke). Lahko pa računamo tudi kvantile (mediano, kvartile, decile, centile) in na ta način izvemo, koliko enot ima večje in koliko enot manjše vrednosti od določene.

- *Ranžirna vrsta* so enote z ustreznimi vrednostmi spremenljivke, ki so urejene od tiste z najmanjšo vrednostjo, do tiste z največjo vrednostjo. (Ferligoj 1995).
- *Rang (R)* je zaporedno mesto, ki ga priredimo vsaki enoti v ranžirni vrsti. Če hočemo, da nam rang pokaže mesto enote v populaciji, ga moramo primerjati s celotno populacijo.

(Ferligoj 1995).

- *Kvantilni rang (P)* nam pove, na katerem delu celotnega razmika leži določena enota, oziroma koliki del celotnega razmika ima manjše vrednosti, od dane vrednosti (Ferligoj 1995).
- *Kvantil* je vrednost spremenljivke, ki pripada določenemu kvantilnemu rangju (Ferligoj 1995).
- *Kvartili* so kvantili, ki razdelijo ranžirno vrsto na četrtine. Prvi kvartil je vrednost, ki pripada kvantilnemu rangju $P=0,25$. prvi kvartil je tudi vrednost, od katere ima $\frac{1}{4}$ enot manjšo, $\frac{3}{4}$ enot pa večjo vrednost (Ferligoj 1995).

Ranžirna vrsta:

Tabela 1.2: Ranžirna vrsta za spremenljivko ENV.tax.2006 (pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države).

Rang R_i	Država	ENV.tax.2006 x_i
1	Litva	1,81
2	Španija	1,87
3	Romunija	1,94
4	Grčija	1,99
5	Belgija	2,15
6	Estonija	2,22
7	Francija	2,26
8	Slovaška	2,28
9	Velika Britanija	2,4
10	Nemčija	2,4
11	Latvija	2,4
12	Irska	2,5
13	Avstrija	2,5
14	Češka republika	2,58
15	Luksemburg	2,61
16	Italija	2,7
17	Švedska	2,76
18	Poljska	2,76
19	Madžarska	2,82
20	Portugalska	2,96
21	Finska	2,99

22	Slovenija	3,01
23	Bolgarija	3,1
24	Ciper	3,29
25	Malta	3,36
26	Nizozemska	4,03
27	Danska	6,17

Interpretacija: Ta ranžirna vrsta razvrsti države članice EU od tiste, v kateri pobrani okoljski davki predstavljajo najmanjši delež celotnega BDP, do tiste, ki s pobranimi davki najbolj „obogati“ svoj BDP. Najmanjši odstotek ima Litva, zgolj 1,18, največjega pa Danska in sicer 6,17. Vidna je velika razlika med pobranimi davki. Slovenija se nahaja zelo pri vrhu, njen rang je 22 od 27. Z dobrimi tremi odstotki, ki jih v BDP-ju naše države predstavljajo pobrane okoljske dajatve, se uvrščamo med boljše države na tem področju.

Prvi kvartil

$$R = P * N + 0,5 = 0,25 * 27 + 0,5 = 7,25$$

$$R = 7,25$$

$$R_0 = 7 \quad x_0 = 2,26$$

$$R_1 = 8 \quad x_1 = 2,28$$

$$\frac{R - R_0}{R_1 - R_0} = \frac{x - x_0}{x_1 - x_0}$$

$$\frac{7,25 - 7}{8 - 7} = \frac{x - 2,26}{2,28 - 2,26} = 2,265$$

$$x = 2,265$$

$$\underline{Q_1 = 2,265}$$

Drugi kvartil

$$R = P * N + 0,5 = 0,5 * 27 + 0,5 = 14$$

$$R = 14$$

$$x = 2,58$$

$$\underline{Me = Q_2 = x_{14} = 2,58}$$

Tretji kvartil

$$R = P * N + 0,5 = 0,75 * 27 + 0,5 = 20,75$$

$$R = 20,75$$

$$R_0 = 20 \quad x_0 = 2,96$$

$$R_1 = 21 \quad x_1 = 2,99$$

$$\frac{R - R_0}{R_1 - R_0} = \frac{x - x_0}{x_1 - x_0}$$

$$\frac{20,75 - 20}{21 - 20} = \frac{x - 2,96}{2,99 - 2,96} = 2,9825$$

$$x = 2,9825$$

$$\underline{Q_3 = 2,9825}$$

Interpretacija: Prvi kvartil ima vrednost 2,265, to pa pomeni, da pri četrtini držav članic EU pobrane okoljske dajatve predstavljajo manj kot 2,265-odstotni delež njihovega BDP v letu 2006. To je četrtina držav, pri katerih je ta odstotek najnižji. Razlogi so seveda lahko različni. Od tega, da njihova davčna zakonodaja ne predvideva veliko okoljskih dajatev, ali pa da pri pobiranju teh davkov niso dovolj uspešni. Ta kvartil nam razdeli države na tiste, v katerih je prihodek od pobranega okoljskega davka v BDP najnižji na ostale. Drugi kvartil, imenovan tudi mediana, je najbolj pomemben, saj razdeli ranžirno vrsto na dva dela in s tem pove, od katere vrednosti ima polovica enot manjšo, oziroma večjo vrednost. V mojem primeru je vrednost drugega kvartila 2,58. To pa natančneje pomeni, da kar pri polovici držav članic EU pobran okoljski davek predstavlja manj kot 2,58 njihovega BDP, pri drugi polovici pa je odstotek večji od 2,58. Ta kvartil nam lahko razdeli države na tiste v katerih pobran okoljski davek predstavlja večji odstotek BDB in na tiste, kjer ta davek predstavlja manjši odstotek BDP. Tretji kvartil pa ima vrednost 2,9825. To pomeni, da pri četrtini držav članic EU pobrani okoljski davek predstavlja več kot 2,9825% njihovega BDP v letu 2006. To so države z največ pobranega okoljskega davka. Sklepamo lahko, da imajo ali visok okoljski davek in tudi, da so pri pobiranju le-tega uspešne. Pri vrhu sta Danska in Nizozemska, kjer so okoljski davki visoki in tudi uspešno pobrani, sledijo pa Malta, Ciper, Bolgarija in takoj za tem Slovenija.

Aritmetična sredina in standardni odklon populacije

Definicije: Aritmetična sredina je vsota vseh vrednosti deljena s številom enot v populaciji in je primerna za številske, normalno porazdeljene spremenljivke (Ferligoj 1995, 36). Standardni odklon, ki ga označimo z σ , je odklon vrednosti od aritmetične sredine, ki meri razpršenost okoli aritmetične sredine. Je kvadratni koren variance. Večja kot je razpršenost, večji je standardni odklon, tem bolj se posamezne vrednosti odklanjajo od aritmetične sredine. Aritmetična sredina in standardni odklon sta primerna za normalno frekvenčno porazdelitev vrednosti.

Izračun aritmetične sredine in standardnega odklona za odstotek pobranih okoljskih dajatev, ki je izražen v deležu BDP posamezne države za leto 2006.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{27} \sum_{i=1}^{27} x_i = \frac{1}{27} * 73,86 = 2,7356$$

$$\underline{\underline{\mu = 2,7356}}$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{27} \sum_{i=1}^{27} (x_i - 2,7356)^2 = \frac{1}{27} * 18,825075 = 0,697225$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,697225} = 0,8350$$

$$\underline{\underline{\sigma = 0,8350}}$$

Interpretacija: Aritmetična sredina spremenljivke je 2,7356. To pomeni, da je povprečen delež pobranih okoljskih dajatev v državah EU 2,7356% njihovega BDP. Standardni odklon spremenljivke pa je 0,8350. To pomeni, da se v povprečju vrednosti odstotkov od aritmetične sredine (2,7356) odklanjajo za 0,8350 odstotka. To nam pove, da se države članice EU ne razlikujejo bistveno v pobranih okoljskih dajatvah.

1.3 Zaključek

V tem delu naloge sem obravnaval spremenljivko ENV.tax.2006 za EU-27. Populacija so bile vse države članice EU v letu 2006, enota pa je bila posamezna država. Najprej sem razporedil vrednosti pobranih okoljskih dajatev, izraženih v odstotku BDP posamezne članice EU, v ranžirno vrsto. S tem sem dobil boljši pregled, saj sem lažje primerjal posamezne države. Vsaki državi je pripadal tudi posamezen rang (rang številka 1 pripada Litvi, kjer odstotek pobranih okoljskih dajatev predstavlja le 1,81% celotnega BDP, zadnji, 27. rang, pa pripada Danski, kje pobrane okoljske dajatve predstavljajo kar 6,17 odstotka celotnega BDP). Sloveniji pripada 22. rang in se s 3,01% uvršča v zgornjo tretjino držav članic EU. S pomočjo ranžirne vrste sem nato izračunal prvi, drugi (ki je hkrati tudi mediana) in tretji kvartil. Prvi kvartil je imel vrednost 2,265, to pa pomeni, da pri četrtini držav članic EU pobrane okoljske dajatve predstavljajo manj kot 2,265-odstotni delež njihovega BDP v letu 2006. Drugi kvartil (mediana) ima vrednost 2,58. To pa natančneje pomeni, da kar pri polovici držav članic EU pobran okoljski davek predstavlja manj kot 2,58% njihovega BDP, pri drugi polovici pa je odstotek večji od 2,58. Tretji kvartil pa ima vrednost 2,9825. To pomeni, da pri četrtini držav članic EU pobrani okoljski davek predstavlja več kot 2,9825% njihovega BDP v letu 2006. To so države z največ pobranega okoljskega davka.

2. PODNALOGA: OCENA POPULACIJSKIH PARAMETROV IN PREVERJANJE DOMNEV NA OSNOVI VZORCA

2.1 Uvod

Cilj naloge: Cilj druge podnaloge je na podlagi podatkov o prvi spremenljivki (pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države) in drugi spremenljivki (BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006) za računalniško generiran slučajni vzorec desetih držav, oceniti aritmetično sredino za populacijo vseh držav Evropske Unije in glede na znane podatke iz 1. podnaloge oceniti tudi kvaliteto te vzorčne ocene. Vse naštetu sem ocenil in izračunal samo za prvo spremenljivko. Prikazane so tudi ugotovitve o razliki aritmetičnih sredin spremenljivk med starimi in novimi članicami EU (nove članice so tiste, pridružene v letu 2004 ali kasneje). Na koncu so podane ugotovitve o možnosti sklepanja, da so nove članice EU drugačne od starih.

Prikaz podatkov:

Tabela 2.1: Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006 in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006. Podatki za slučajni vzorec 10 držav Evropske unije.

Država	ENV.tax.2006	LabourGDPinPPS.2006
Belgija	3,1	36,2
Ciper	3,29	83,7
Danska	6,17	106,4
Francija	2,26	121,1
Irska	2,5	134,7
Madžarska	2,82	67,9
Nemčija	2,4	109,1
Slovenija	3,01	83,9
Španija	1,87	102,6
Švedska	2,76	112,5

2.2 Intervalna ocena populacijske aritmetične sredine

Definicija: Interval zaupanja je interval, v katerem se z dano gotovostjo $(1-\alpha)$ nahaja ocenjevani parameter (γ) , pri čemer poskušamo najti nepristransko statistiko g , ki se na vseh možnih vzorcih približno normalno porazdeljuje s standardno napako $SE(g)$. Z verjetnostjo tveganja α se parameter γ nahaja v intervalu zaupanja $P(a < \gamma < b) = 1 - \alpha$, pri čemer je a spodnja meja zaupanja, b zgornja meja zaupanja in α verjetnost tveganja (Ferligoj 1995, 120). 2,5 % stopnja tveganja nam npr. pove, da pri 100 slučajnih vzorcih (za vsakega smo izračunali interval zaupanja za parameter γ) lahko pričakujemo, da 2,5 intervalov zaupanja od 100 ne bo pokrilo iskanega parametra γ (Ferligoj 1995, 122). Intervala zaupanja za aritmetično sredino/standardni odklon v spodnjem primeru nam podata interval, v katerem se bosta s 95% verjetnostjo nahajali vrednosti populacijske aritmetične sredine in standardnega odklona.

Interval zaupanja za populacijsko aritmetično sredino

Vzorčna aritmetična sredina, varianca in standardni odklon:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{1}{10} * 30,18 = 3,018$$

$$\underline{\underline{X=3,018}}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 = \frac{1}{10-1} \sum_{i=1}^{10} (x_i - 3,018)^2 = \frac{1}{9} * 12,66436 = 1,4071$$

$$\underline{\underline{s^2 = 1,4071}}$$

$$s = \sqrt{s^2} = \sqrt{1,4071} = 1,1862$$

$$\underline{\underline{s = 1,1862}}$$

t vrednost

$$t_{\frac{\alpha}{2}}(n-1) = t_{\frac{0,05}{2}}(10-1) = t_{0,025}(9) = 2,26$$

Interval zaupanja

$$P = \left[\bar{X} - t_{\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \right] = 0,95$$

$$3,018 - 2,26 \frac{1,1862}{\sqrt{10}} < \mu < 3,018 + 2,26 \frac{1,1862}{\sqrt{10}}$$

$$\underline{\underline{2,1702 < \mu < 3,8657}}$$

Interpretacija: S 95% gotovostjo lahko trdimo, da povprečen odstotek pobranega okoljskega davka (izraženega v odstotku BDP) za leto 2006 v državah članicah EU znaša med 2,1702% in

3,8657%. Interval zaupanja vsebuje pravo populacijsko aritmetično sredino (2,7356%), saj je le-ta nad spodnjo in pod zgornjo mejo intervala zaupanja. To pomeni, da izbrani vzorec spada med 95% tistih vzorcev, ki pravilno pokrijejo aritmetično sredino.

2.3 Preverjanje domneve o enakosti oz. razliki populacijskih aritmetičnih sredin starih in novih članic EU

Tabela 2.2: Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006 in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006. Podatki za nove članice EU, vključene v vzorec desetih držav.

Država	ENV.tax.2006	LabourGDPinPPS.2006
Bolgarija	3,1	36,2
Ciper	3,29	83,7
Madžarska	2,82	67,9
Slovenija	3,01	83,9

Tabela 2.3: Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006 in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006. Podatki za stare članice EU, vključene v vzorec desetih držav.

Država	ENV.tax.2006	LabourGDPinPPS.2006
Danska	6,17	106,4
Francija	2,26	121,1
Irska	2,5	134,7
Nemčija	2,4	109,1
Španija	1,87	102,6
Švedska	2,76	112,5

Aritmetični sredini in standardna odklona držav v posameznih skupinah

Stare članice

$$X_s = \frac{1}{n_s} \sum_{i=1}^{n_s} x_i = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{1}{6} 17,96 = 2,9933$$

$$\underline{\underline{X_s = 2,9933}}$$

$$s_S^2 = \frac{1}{n_S - 1} \sum_{i=1}^{n_S} (x_i - \bar{X}_S)^2 = \frac{1}{6-1} \sum_{i=1}^6 (x_i - 2,9933)^2 = \frac{1}{5} 12,5405 = 2,5081$$

$$s_S = \sqrt{s_S^2} = \sqrt{2,5081} = 1,5837$$

$$\underline{\underline{S_S = 1,5837}}$$

Nove članice

$$\bar{X}_N = \frac{1}{n_N} \sum_{i=1}^{n_N} x_i = \frac{1}{4} \sum_{i=1}^4 x_i = \frac{1}{4} 12,22 = 3,055$$

$$\underline{\underline{X_n = 3,055}}$$

$$s_N^2 = \frac{1}{n_N - 1} \sum_{i=1}^{n_N} (x_i - \bar{X}_N)^2 = \frac{1}{4-1} \sum_{i=1}^4 (x_i - 3,055)^2 = \frac{1}{3} 0,1144 = 0,03813$$

$$s_N = \sqrt{s_N^2} = \sqrt{0,03813} = 0,19527$$

$$\underline{\underline{S_N = 0,19527}}$$

Preverjanje domneve enakosti aritmetičnih sredin

Ničelna domneva: $\mu_S - \mu_N = 0$

Na populaciji držav EU ni razlik med aritmetičnima sredinama pobranih okoljskih davkov, izraženih v odstotku BDP posamezne države za leto 2006, za stare in nove (pridružene leta 2004 ali kasneje) članice.

Alternativna domneva: $\mu_S - \mu_N \neq 0$

Na populaciji so razlike med aritmetičnima sredinama pobranih okoljskih davkov, izraženih v odstotku BDP posamezne države za leto 2006, za stare in nove (pridružene leta 2004 ali kasneje) članice.

Ker predpostavljamo, da sta varianci na obeh populacijah enaki, iz dveh vzorčnih varianc izračunamo (uteženo) povprečno vzorčno varianco in iz nje standardni odklon, ki bo ocena populacijskega standardnega odklona:

$$s^2 = \frac{(n_S - 1)s_S^2 + (n_N - 1)s_N^2}{n_S + n_N - 2} = \frac{(6-1)2,5081 + (4-1)0,03813}{6 + 4 - 2} = 1,58186$$

$$s = \sqrt{s^2} = \sqrt{1,58186} = 1,2577$$

$$\underline{\underline{s = 1,2577}}$$

Testa statistika. Njena eksperimentalna vrednost je v primeru preverjanja domneve o razliki populacijskih aritmetičnih sredin na malih vzorcih:

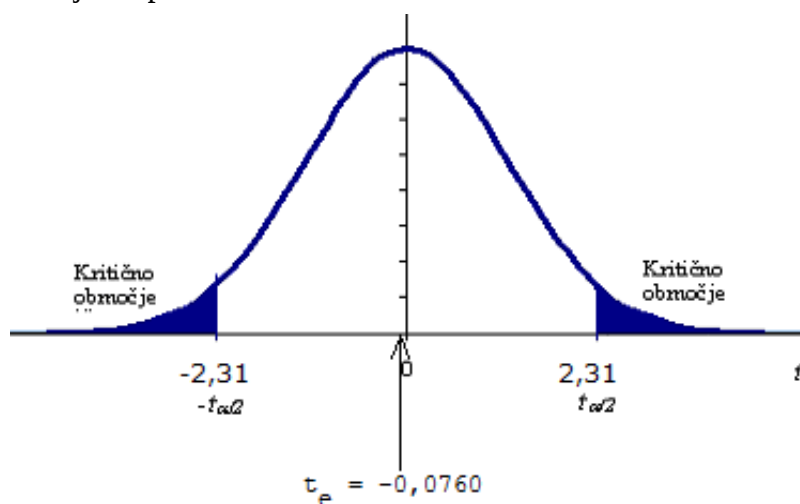
$$t_e = \frac{\bar{X}_S - \bar{X}_N - (\mu_S - \mu_N)_H}{s \sqrt{\frac{n_S + n_N}{n_S * n_N}}} = \frac{(2,9933 - 3,055) - 0}{1,2577 \sqrt{\frac{6 + 4}{6 * 4}}} = -0,0760$$

$$t_e = -0,0760$$

Kritično območje je določeno z vrednostma testne statistike t pri 5% stopnji značilnosti in 8 prostostnih stopnjah:

$$t_{\frac{\alpha}{2}}(n_S + n_N - 2) = t_{\frac{0,005}{2}}(6 + 4 - 2) = t_{0,025}(8) = \pm 2,31$$

Slika 2.1: Prikaz položaja eksperimentalne vrednosti testne statistike in kritičnih območij.



Sklep: Ker je $-0,0760 >$ od $-2,31$ oz. $|t_e| > |t_{\alpha/2}|$, ničelno domnevo lahko zavrnamo.

Interpretacija: Eksperimentalna vrednost pade v kritično območje, zato ničelno domnevo lahko zavrnamo. Povprečna vrednost pobranih okoljskih dajatev, izraženih v odstotku BDP za leto 2006 v državah članicah EU pri starih in novih članicah, ki so bile izbrane v vzorec, ob 5% stopnji značilnosti statistično ni različna. Hipotezo o pobranih okoljskih dajatvah, izraženih v odstotku BDP za leto 2006 v državah članicah EU pri starih in novih članicah torej lahko posplošimo na celotno populacijo.

2.5 Zaključek

V tem delu naloge sem ugotovil, da povprečen odstotek pobranih okoljskih dajatev, izraženih v odstotku BDP za leto 2006 v državah članicah EU na populaciji ni precej drugačen kot tisti na vzorcu. Populacijska in vzorčna aritmetična sredina se namreč razlikujeta za manj kot pol odstotka.

Tudi standardna odklona na populaciji ter na vzorcu nista zelo različna - razlika med njima je le okoli 0,20%. Pri izračunu intervala zaupanja za aritmetično sem ugotovil, da le-ta vsebuje pravo populacijsko aritmetično sredino, ki se nahaja v sprejemljivem območju. To pomeni, da izbrani vzorec spada med 95% tistih vzorcev, ki pravilno pokrijejo aritmetično sredino. V nadaljevanju sem ugotovil, da je odstotek pobranih okoljskih dajatev, izraženih v odstotku BDP za leto 2006 v državah članicah EU v izbranem vzorcu pri starih in novih članicah podobna. Razlika ni velika, zato lahko govorimo o statistično značilni podobnosti med skupinama držav, eksperimentalna vrednost namreč ne pade v kritično območje zavračanja domneve o enakosti aritmetičnih sredin ampak je nasprotno v sami sredini sprejemljivega območja.

3. PODNALOGA: ANALIZA POVEZANOSTI ZA NOMINALNI TIP PARA SPREMENLJIVK

3.1 Uvod

Cilj naloge: Cilj tretje podnaloge je ugotoviti, ali obstaja povezanost med tem, kdaj je država postala članica EU (pred ali v letu 2004) ter višino vrednosti prve spremenljivke. V ta namen je obravnavana celotna populacija držav EU, pri čemer sta analizirani naslednji dve dihotomni spremenljivki:

- pripadnost starim ali novim članicam EU (vrednost 1 - stare članice EU, pridružene pred letom 2004; vrednost 2 - nove članice EU, pridružene leta 2004 ali kasneje)
- vrednost prve spremenljivke (pobrane okoljske dajatve v letu 2006, izražene v odstotku BDP posamezne države) glede na mediano (vrednost 1 – manjša ali enaka mediani; vrednost 2 - večja od mediane)

Na osnovi omenjenih dihotomnih spremenljivk so države razporejene v kontingenčno tabelo; dodani so:

- opis razlik med starimi in novimi članicami, ki jih je mogoče razbrati iz strukturnih odstotkov;
- ustrezna ponazoritev s tabelo in strukturnimi stolpci;
- ugotovitve o povezanosti med spremenljivkama na podlagi kontingenčnih koeficientov

Obravnavani spremenljivki:

Prva spremenljivka: Pripadnost starim ali novim članicam EU z vrednostima 1 – stara članica, pridružena pred 2004, in 2 – nova članica, pridružena leta 2004 ali kasneje. Druge spremenljivka: pobrane okoljske dajatve v letu 2006, izražene v odstotku BDP posamezne države z vrednostma 1 – tiste države, kjer pobrani okoljski davki predstavljajo manj kot 2,58% BDP in vrednost 2 – tiste države, kjer pobrani okoljski davki predstavljajo več kot 2,58% njihovega BDP.

Tabela 3.1: Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006. Podatki glede na čas pridružitve EU in glede na vrednost mediane.

Rang R_i	Država	ENV.tax.2006 x_i	Čas pridružitve	Vrednost glede na mediano
1	Litva	1,81	2	1
2	Španija	1,87	1	1

3	Romunija	1,94	2	1
4	Grčija	1,99	1	1
5	Belgija	2,15	1	1
6	Estonija	2,22	2	1
7	Francija	2,26	1	1
8	Slovaška	2,28	2	1
9	Velika Britanija	2,4	1	1
10	Nemčija	2,4	1	1
11	Latvija	2,4	2	1
12	Irska	2,5	1	1
13	Avstrija	2,5	1	1
14	Češka republika	2,58	2	1
15	Luksemburg	2,61	1	2
16	Italija	2,7	1	2
17	Švedska	2,76	1	2
18	Poljska	2,76	2	2
19	Madžarska	2,82	2	2
20	Portugalska	2,96	1	2
21	Finska	2,99	1	2
22	Slovenija	3,01	2	2
23	Bolgarija	3,1	2	2
24	Ciper	3,29	2	2
25	Malta	3,36	2	2
26	Nizozemska	4,03	1	2
27	Danska	6,17	1	2

Tabela 3.2: Kontingenčna tabela: Stare in nove članice EU glede na pobrane okoljske dajatve, izražane v odstotku BDP za leto 2006.

Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006.	Stare članice EU (pridružene pred 2004)	Nove članice EU (pridružene 2004 ali kasneje)	Skupaj
Enaki ali manjši od 2,58%	8	6	13
Večji od 2,58%	7	6	14
Skupaj	15	12	27

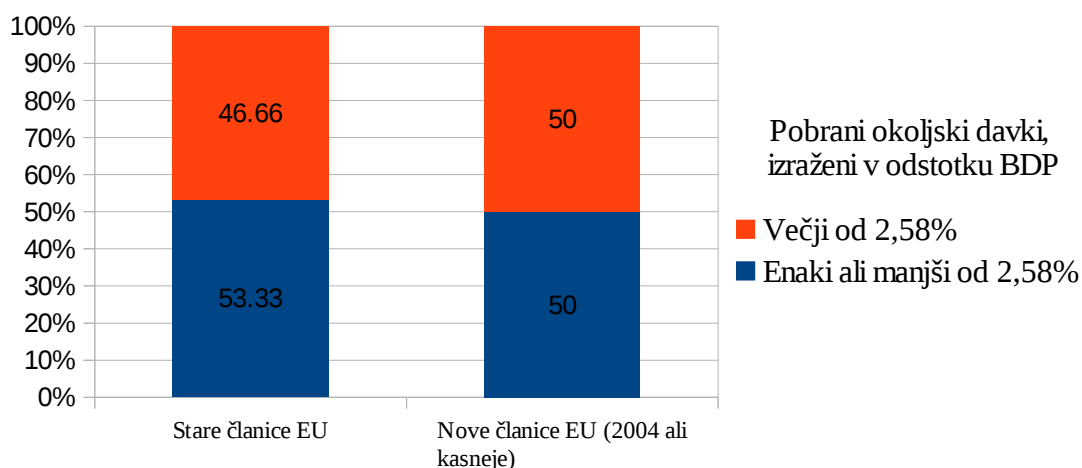
3.2 Analiza povezanosti za nominalni tip spremenljivk

Strukturni odstotki in strukturni stolpci

Tabela 3.3: Razporeditev starih in novih članic EU glede na pobrane okoljske dajatve, izražane v odstotku BDP za leto 2006 – **strukturni odstotki**.

Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006.	Stare članice EU (pridružene pred 2004)	Nove članice EU (pridružene 2004 ali kasneje)	Skupaj
Enaki ali manjši od 2,58%	53,33	50	48,15
Večji od 2,58%	46,66	50	51,85
Skupaj	100	100	100

Slika 3.1: Razporeditev starih in novih članic glede na pobrane okoljske dajatve, izražane v odstotku BDP za leto 2006.



Interpretacija: Pri starih članicah EU je 53,33% držav imelo vrednost pobranih okoljskih davkov, izraženih v odstotku BDP za leto 2006, enako manjšo od mediane, (ki je 2,58%), 46,66% držav pa je imelo vrednost pobranih okoljskih davkov, izraženih v odstotku BDP za leto 2006, večjo od mediane. Pri novih članicah je rezultat neodločen, enako ali manjšo vrednost ima namreč 50% držav, pri preostalih 50% držav pa je ta vrednost večja od mediane.

Sklepna ugotovitev: Pri starih članicah EU predstavljajo pobrani okoljski davki manjši odstotek v njihovem BDP-ju, čeprav je razlika minimalna (je praktično ni).

χ^2 , Cramerjev in Pearsonov koeficient kontingence

Definicije: Tako statistika χ^2 , kot tudi Cramerjev in kontingenčni koeficient so primerni za nominalni tip para spremenljivk (Ferligoj 1995, 159). S χ^2 koeficientom primerjamo dejanske in teoretične frekvence. Statistika χ^2 se porazdeljuje po χ^2 porazdelitvi s $(s - 1)(v - 1)$ prostostnimi stopnjami, kjer je v število vrstic v kontingenčni tabeli in s število stolpcev. Statistika χ^2 je lahko le pozitivna in zavzame vrednosti v intervalu $[0, \chi^2_{\max}]$, kjer je $\chi^2_{\max} = n(k - 1)$, če je $k = \min(v, s)$ (Ferligoj 1995, 166). Ker pa statistika χ^2 v splošnem ni primerljiva, je definiranih več kontingenčnih koeficientov, ki so bolj ali manj primerni; dva izmed njih sta kontingenčni in Cramerjev koeficient. Prvi je definiran na intervalu $[0, C_{\max}]$ (kjer je $C_{\max} = \sqrt{(k-1)/k}$) in izraža samo stopnjo (jakost) povezanosti (kontingence) med spremenljivkama, ne pove pa, kakšna je smer te povezanosti (pozitivna ali negativna). Cramerjev koeficient je definiran na intervalu $[0, 1]$ in variira med 0 in 1 ter ni odvisen od velikosti kontingenčne tabele. Cramerjeve koeficiente lahko zato neposredno primerjamo za različno velike tabele (Ferligoj 1995, 166).

Teoretične frekvence in χ^2

Tabela 3.4: Kontingenčna tabela: Stare in nove članice EU glede na pobrane okoljske dajatve, izražene v odstotku BDP, za leto 2006

Pobrani okoljski davki, izraženi kot odstotni delež BDP-ja posamezne države za leto 2006.	Stare članice EU (pridružene pred 2004)	Nove članice EU (pridružene 2004 ali kasneje)	Skupaj
Enaki ali manjši od 2,58%	7,7777	6,2222	13
Večji od 2,58%	7,2222	5,7777	14
Skupaj	15	12	27

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - f'_i)^2}{f'_i} = \frac{(8 - 7,7777)^2}{7,7777} + \frac{(6 - 6,2222)^2}{6,2222} + \frac{(7 - 7,2222)^2}{7,2222} + \frac{(6 - 5,7777)^2}{5,7777} = 0,02965$$

$$\underline{\underline{\chi^2 = 0,02965}}$$

Cramerjev koeficient in Pearsonov koeficient kontingence

$$\alpha = \sqrt{\frac{\chi^2}{n*(k-1)}} = \sqrt{\frac{0,02965}{27*(2-1)}} = 0,0331$$

$$\underline{\underline{\alpha = 0,0331}}$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0,02965}{0,02965 + 27}} = 0,0331$$

$$\underline{\underline{C = 0,0331}}$$

$$C_{max} = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{2-1}{2}} = 0,70711$$

$$\underline{\underline{C_{max} = 0,70711}}$$

$$C_{pop} = \frac{C}{C_{max}} = \frac{0,0331}{0,70711} = 0,0468$$

$$\underline{\underline{C_{pop} = 0,0468}}$$

Interpretacija: Tako Cramerjev (0,0331) kot kontingenčni (0,0331; popravljen kontingenčni 0,0468) koeficient kažeta na dokaj slabo povezanost med spremenljivkama. Spremenljivki čas pridružitve EU in pobrani okoljski davki, izraženi v odstotku BDP za leto 2006 torej nista povezani v močni meri. To sem ugotovili tudi s strukturnimi odstotki, saj je bila razlika v odstotkih vrednosti, ki so večje od mediane, pri starih in novih članicah skoraj enaka.

3.3 Zaključek

Ugotovili sem, da bistvene razlike med starimi in novimi članicami EU pri pobranih okoljskih dajatvah, izraženih v odstotku BDP-ja posamezne države, ni. Obe skupini imata namreč skoraj enak odstotek pobranih okoljskih davkov, ki je višji od mediane – stare članice 46,66%, nove (pridružene leta 2004 ali kasneje) pa 50%. Ugotovil sem, da čas pridružitve EU ter odstotek pobranih okoljskih dajatev, izraženih v odstotku BDP, med seboj nista povezana. Te ugotovitve niti niso presenetljive, saj prav vse članice EU veže ista okoljska zakonodaja. Kljub majhni odstotkovni razliki pa je zanimivo, da nove članice vseeno poberejo več z okoljem povezanih davkov, kot stare.

4. PODNALOGA: ANALIZA POVEZANOSTI IN ODVISNOSTI ZA RAZMERNOSTNI TIP PARA SPREMENLJIVK

4.1 Uvod

Cilj naloge: Cilj četrtega dela statistične naloge je na podlagi računalniško ustvarjenega slučajnega vzorca desetih držav ugotoviti, ali na populaciji držav EU obstaja linearna povezanost med spremenljivkama ter kako je prva spremenljivka odvisna od druge spremenljivke. V ta namen so analizirani:

- obstoj linearne povezanosti med spremenljivkama na populaciji držav EU;
- linearna odvisnost prve (odvisne) spremenljivke (pobrani okoljski davki, izraženi v odstotku BDP posamezne države za leto 2006) od druge (neodvisne) spremenljivke (BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006);
- grafična ponazoritev z grafikonom, iz katerega so jasno razvidne posamezne države iz vzorca;
- višina spremembe odvisne spremenljivke v primeru spremembe neodvisne spremenljivke;
- delež variabilnosti odvisne spremenljivke v navezavi s pojasnjevanjem vpliva neodvisne spremenljivke in variabilnosti držav okoli regresijske premice

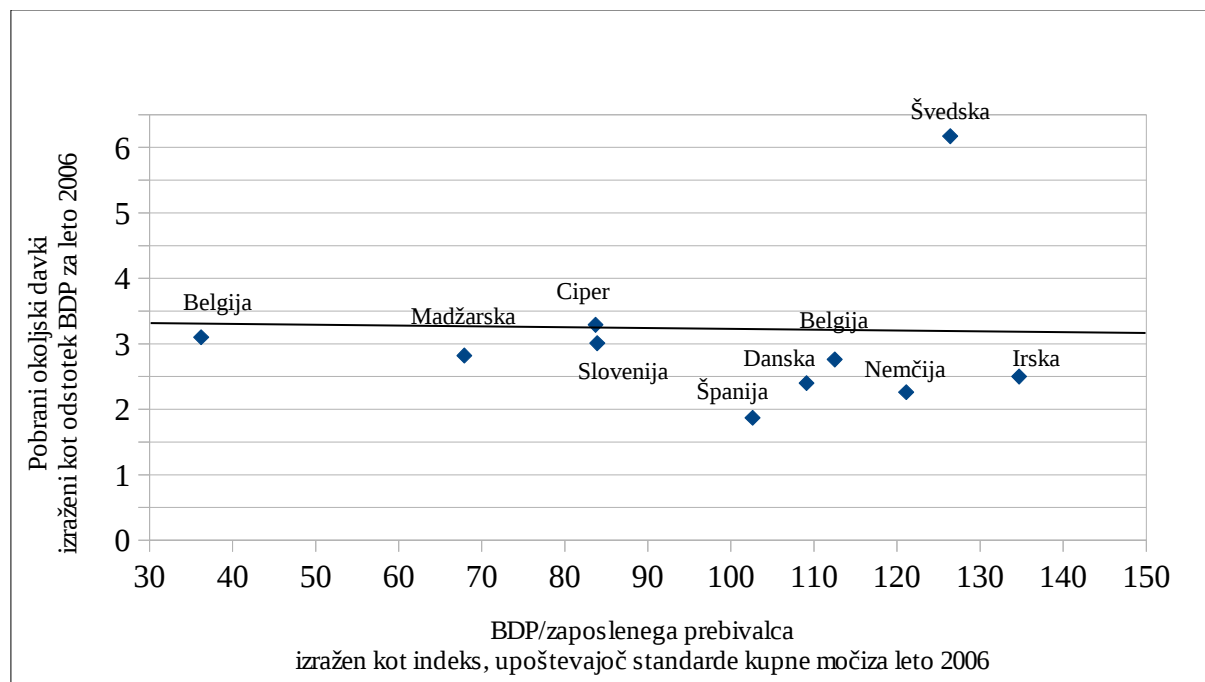
Tabela 4.1: Pobrani okoljski davki, izraženi v odstotku BDP posamezne države za leto 2006 (ENV.tax.2006) in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006 (LabourGDPinPPS.2006).

Država	ENV.tax.2006 Y	LabourGDPinPPS.2006 X
Belgija	3,1	36,2
Ciper	3,29	83,7
Danska	6,17	106,4
Francija	2,26	121,1
Irska	2,5	134,7
Madžarska	2,82	67,9
Nemčija	2,4	109,1
Slovenija	3,01	83,9
Španija	1,87	102,6
Švedska	2,76	112,5

4.2 Grafična predstavitev odnosa med spremenljivkama

Razsevni grafikon

Slika 4.2: Razsevni grafikon: Vzorec držav EU glede na odstotek pobranih okoljskih dajatev, izraženih kot odstotek BDP posamezne države za leto 2006 in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006.



Interpretacija: Razsevni grafikon je nazorna grafična predstavitev regresijske odvisnosti med spremenljivkama (Ferligoj 1995, 183). V njem na ordinatno os nanašamo vrednosti prve, na abscisno pa vrednosti druge spremenljivke. Dobljene točke predstavljajo položaj posameznih enot na grafu.

4.3 Analiza povezanosti med spremenljivkama

Pearsonov koeficient korelacije

Definicije: Pearsonov koeficient korelacije je primeren za številski tip para spremenljivk in se uporablja za ugotavljanje linearne povezanosti med spremenljivkama. Zavzame lahko vrednosti v intervalu $[-1,1]$; če se z večanjem vrednosti prve spremenljivke večajo vrednosti druge spremenljivke je koeficient pozitiven, v kolikor se vrednosti druge spremenljivke nižajo je negativen (Ferligoj 1995, 177). Pearsonov koeficient je primeren za dvorazsežno normalno porazdelitev; slabost Pearsonovega koeficienta je, da z njim ne moremo ugotavljati povezanosti spremenljivk, če je ta nelinearna.

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n (x_i - 95,81)(y_i - 3,018)}{\sqrt{\sum_{i=1}^n (x_i - 95,81)^2 \sum_{i=1}^n (y_i - 3,018)^2}} = \frac{-28,8098}{305,95047} = -0,0942$$

$$\underline{r_{XY} = -0,0942}$$

Interpretacija: Vrednost Pearsonovega koeficienta (-0,0942) kaže na šibko povezanost med spremenljivkama, saj je koeficient bližje 0, kot pa -1. Hkrati Pearsonov koeficient kaže, da sta spremenljivki povezani negativno. V mojem primeru to pomeni, da je povezava pobrani okoljski davki, izraženimi v odstotku BPD za leto 2006 in BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006, izredno šibka.

Preverjanje domneve o linearni povezanosti na populaciji

Ničelna domneva: $\rho_{XY} = 0$

Na populaciji držav EU ni linearne povezanosti med spremenljivko pobrani okoljski davki, izraženi v odstotku BPD za leto 2006 in spremenljivko BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006.

Alternativna domneva: $\rho_{XY} \neq 0$

Na populaciji je linearna povezanost med spremenljivko pobrani okoljski davki, izraženi v odstotku BPD za leto 2006 in spremenljivko BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006.

Za ugotavljanje, ali obstaja linearna povezanost na populaciji pri 10% stopnji značilnosti, je primerna testna statistika t .

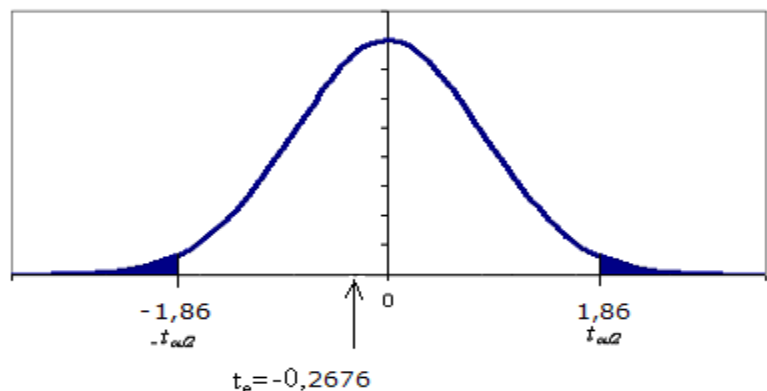
$$t_e = \frac{r_{XY} \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} = \frac{-0,0942 \sqrt{10-2}}{\sqrt{1-(-0,0942)^2}} = -0,2676$$

$$\underline{t_e = -0,2676}$$

Kritično območje je določeno z naslednjima dvema vrednostma testne statistike t pri 10% stopnji značilnosti in 8 prostostnih stopnjah:

$$t_{\frac{\alpha}{2}}(n-2) = t_{\frac{0,10}{2}}(10-2) = t_{0,05}(8) = \pm 1,86$$

Slika 4.1: Prikaz položaja eksperimentalne vrednosti testne statistike in kritičnih območij.



Sklep: Ker je $-0,2676 >$ od $-2,31$ oz. $|t_e| < |t_{\alpha/2}|$, ničelne domneve ne moremo zavrniti.

Interpretacija: Ker eksperimentalna vrednost testne statistike ne pade v območje zavračanja, ničelne hipoteze ob 10 % stopnji značilnosti ne moremo zavrniti. Torej ne moremo zavrniti domneve, da na populaciji držav EU ni linearne povezanosti med spremenljivko pobrani okoljski davki, izraženi kot odstotek BDP v letu 2006 in spremenljivko BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006.

4.4 Analiza odvisnosti med spremenljivkama

Regresijska premica

Definicija: Linearno odvisnost med spremenljivkama proučujemo s pomočjo prve regresijske funkcije, ki kaže, kakšen bi bil vpliv spremenljivke X na spremenljivko Y, če razen vpliva spremenljivke X ne bi bilo drugih vplivov na spremenljivko Y (Ferligoj 1995, 182). V spodnjem primeru BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006 predstavlja neodvisno (X), pobrani okoljski davki, izraženi kot odstotek BDP posamezne države za leto 2006 pa odvisno spremenljivko (Y).

$$Y' = \bar{Y} + \frac{C_{XY}}{S_X^2} (X - \bar{X}) = \bar{Y} + \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} (X - \bar{X})$$

$$Y' = 3,018 + \frac{\sum_{i=1}^n (x_i - 95,81)(y_i - 3,018)}{\sum_{i=1}^n (x_i - 95,81)^2} (X - 95,81) = 3,018 + \frac{(-28,8098)}{7391,269} (X - 95,81)$$

$$Y' = 3,018 - 0,003897 (X - 95,81)$$

$$Y' = -0,003897 X + 3,391702466$$

$$\underline{Y' = 3,3917 - 0,0039X}$$

Regressijski koeficient

Vrednost regresijskega koeficienta je -0,0039.

Interpretacija: Koeficient je precej nizek, je pa tudi negativen, zato premica zelo počasi pada. V primeru, da se bo indeks BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, povečal za npr. 100, gre pričakovati, da se bodo pobrane okoljske dajatve, izražene v odstotku BDP, znižale za 0,39%.

Determinacijski koeficient

Definicija: Pri napovedovanju odvisnosti pri vsaki od enot delamo napako, ki jo imenujemo varianca napake ali nepojasnjena varianca; pravilno napovedan del variance predstavlja pojasnjena varianca ali determinacijski koeficient. Ta kaže delež celotne variance odvisne spremenljivke (Y) ki je pojasnjen z variiranjem neodvisne spremenljivke (X) (Leskošek 2006).

$$R = r_{XY}^2 = -0,0942^2 = 0,00887$$

$$\underline{R = 0,00887}$$

Interpretacija: Z vplivom neodvisne spremenljivke (BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, izražen kot indeks za leto 2006) je mogoče pojasniti 0,887% delež variabilnosti odvisne spremenljivke (pobrani okoljski davki, izraženi v odstotku BDP posamezne države za leto 2006).

Standardna napaka ocene

Definicija: Kvadratni koren iz nepojasnjene variance imenujemo standardna napaka regresijske ocene in meri razpršenost točk okoli regresijske krivulje (v obravnavanem primeru premice). Ocenjuje kakovost ocenjevanja vrednosti odvisne spremenljivke (delež proračunskega primanjkljaja) z regresijsko funkcijo (Ferligoj 1995, 191).

$$s_e = s_y \sqrt{1 - r_{XY}^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2} \sqrt{1 - r_{XY}^2} = \sqrt{\frac{1}{10-1} \sum_{i=1}^n (y_i - 3,018)^2} \sqrt{1 - (-0,0942)^2} = 1,1809$$

$$\underline{s_e = 1,1809}$$

Interpretacija: Variabilnost držav okoli regresijske premice znaša 1,1809% pobranih okoljskih davkov, izraženih v odstotku BDP za leto 2006. Torej se pri napovedi vrednosti odvisne spremenljivke na osnovi regresijske premice standardno zmotimo za 1,1809%.

4.4. Zaključek

V tem delu sem na podlagi računalniško ustvarjenega slučajnega vzorca desetih držav poskušal ugotoviti, ali na populaciji držav EU obstaja linearna povezanost med spremenljivkama ter kako je prva spremenljivka odvisna od druge spremenljivke. Ugotovil sem, da sta spremenljivki šibko povezani (oz. praktično nista povezani), ter da sta povezani negativno (z večanjem vrednosti neodvisne spremenljivke se vrednost odvisne spremenljivke manjša). Ugotovil sem tudi, da regresijska premica pada in sicer zelo počasi, velika pa je tudi variabilnost okoli nje, zato lahko opazim, da BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči, praktično ne vpliva pobrane okoljske dajatve, izražene v odstotku BDP za leto 2006. Statistične povezanosti pobranih okoljskih dajatev, izraženih v odstotku BDP posamezne države z višino BDP na zaposlenega prebivalca, upoštevajoč standarde kupne moči ob 5% stopnji značilnosti ne moremo potrditi za celotno populacijo. Ugotovitve so pričakovane, saj ima BDP na zaposlenega prebivalca upoštevajoč standardne kupne moči v praksi zelo malo vpliva na to, koliko okoljskih dajatev pobere država. Morda bi lahko trdil, da višina BDP na zaposlenega prebivalca upoštevajoč standarde kupne moči kaže na višji standard in s tem posledično na višjo stopnjo davčne obremenitve, vendar so ugotovitve drugačne. Povezanost med spremenljivkam je zelo zelo šibka (je praktično ni), zato ne vplivata ena na drugo.

Viri in literatura

- *Eurostat*. Dostopno prek: <http://epp.eurostat.ec.europa.eu/> (24. maj 2012)
- Ferligoj, Anuška. 1995. *Osnove statistike na prosojnicah*. Ljubljana: samozaložba.
- Leskošek, Bojan. 2006. *Korelacija in regresija*. Dostopno prek: <http://www2.fsp.uni-lj.si/AKIS/Statistika/KorelacijaRegresija.pdf> (24. maj 2012)
- *Statistika@fdv*. Dostopno prek: <http://statistika.fdvinfo.net/> (24. maj 2012)