

Moguće odrednice dohotka stanovnika lokalnih jedinica u RH

Krištofić, Tina

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:741873>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-08-17**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Tina Krištofić

MOGUĆE ODREDNICE DOHOTKA
STANOVNIKA LOKALNIH JEDINICA U
REPUBLICI HRVATSKOJ

Diplomski rad

Voditeljica rada:
prof. dr. sc. Katarina Ott

Zagreb, veljača, 2023.

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

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Mojoj obitelji, bez čije podrške ovo ne bi bilo moguće

Sadržaj

Sadržaj	iv
Uvod	2
1 Podaci i pregled literature	3
2 Metodologija	7
2.1 Generalizirani linearni model	7
2.2 Procjena parametara	10
2.3 Odabir optimalnog modela	11
2.4 Analiza reziduala	13
2.5 GLM u R-u	14
3 Empirijska analiza	15
3.1 Deskriptivna analiza	16
3.2 Generalizirani linearni model	25
4 Zaključak	67
Bibliografija	69

Uvod

Ovaj rad ispituje utjecaj ekonomskih i demografskih varijabli na prosječni dohodak po stanovniku u jedinicama lokalne i područne samouprave Republike Hrvatske (RH). Jedinice lokalne samouprave sastoje se od 428 općina i 128 gradova, a regionalne jedinice od 21 županije. Grad Zagreb ima status grada i županije. Cilj je rada odrediti i opisati utjecaj odabranih varijabli na prosječni dohodak po stanovniku. Značajnost neke od varijabli može pomoći u objašnjavanju razlika u dohotku među gradovima, općinama ili županijama.

U radu se prvo predstavlja literatura na kojoj je temeljen odabir varijabli kojima će se baviti u naknadnim poglavljima. U mnogim je zemljama svijeta nejednakost u dohotku veliki problem pa je zato i veoma bitno područje za istraživanje. U prvom poglavlju pregledavaju se neke od najčešće korištenih varijabli za pokušaj opisivanja dohotka kao što su porez na dohodak, stopa nezaposlenosti i stupanj obrazovanja. Također, predstavljaju se radovi u kojima su korištene te varijable i dobiveni rezultati. Na temelju tih informacija kreirane su hipoteze o utjecaju varijabli na prosječni dohodak po stanovniku koje se testiraju u daljnjem radu. Provjerit će se značajnost utjecaja prosječnog poreza po stanovniku i stope nezaposlenosti, pozitivan utjecaj postotka visoko obrazovanog stanovništva i pristupa moru, negativan utjecaj postotka stanovništva mlađeg od 14 i starijeg od 65 godina, značajan i pozitivan utjecaj broja stanovnika i gustoće naseljenosti te značajan i negativan utjecaj pristupa poljoprivrednom zemljištu.

Drugo poglavlje posvećuje se opisu korištene metodologije. Prvo se predstavlja generalizirani linearni model i dodatno objašnjava tip modela kojim se ovaj rad bavi. Na više modela dobivene su, preko programskog jezika i okruženja R, procjene utjecaja izabranih varijabli na prosječni dohodak po stanovniku na svakoj skupini podataka. Potom je izabran optimalniji model od predstavljenih pomoću statistike omjera vjerodostojnosti, te je provjerena njegova kvaliteta analizom reziduala. Izdvojena je i interpretacija informacija dobivenih preko programskog jezika R koji je korišten za analizu podataka.

U zadnjem je poglavlju provedena analiza podataka. Prvo su predstavljene sve varijable i obavljena je deskriptivna analiza podataka na razini gradova, općina i županija. Koris-

tili su se ekonomski podaci (kao što su dohodak, porez, stopa nezaposlenosti) iz 2019. te demografski podaci iz popisa stanovništva 2011. Potom su provedeni metodološki postupci opisani u drugom poglavlju te su predstavljeni dobiveni rezultati i uspoređeni s početnim hipotezama. Za svaku skupinu podataka, tj. za općine, gradove i županije, među obrađivanim modelima pronađen je onaj koji najbolje opisuje dohodak u toj skupini. Hipoteza o značajnom utjecaju prosječnog poreza na dohodak djelomično je potvrđena. Prosječni porez po stanovniku značajno utječe na dohodak u gradovima, ali se isto nije moglo potvrditi za općine i županije. Negativan i značajan utjecaj stope nezaposlenosti te pozitivan i značajan utjecaj postotka stanovništva sa završenim visokim obrazovanjem potvrđeni su za sve skupine podataka. Također su potvrđene hipoteze da su postotak stanovnika mlađih od 14 i onih starijih od 65 negativnog utjecaja. Zbog različitog utjecaja na modele u svim skupinama podataka nisu potvrđene ni opovrgnute hipoteze povezane s brojem stanovnika, pristupom moru, pristupom poljoprivrednom zemljištu te gustoćom naseljenosti.

Poglavlje 1

Podaci i pregled literature

Mnogi radovi i istraživanja (npr. Auten i Carroll (1999) [6]; Holcombe i Lacombe (2004) [17]; Blejer i Guerrero (1990) [9]) bave se pitanjem dohotka. Znanje o tome koje varijable najviše utječu na sami dohodak, prosječni dohodak stanovnika ili nejednakosti u raspodjeli dohotka važno je za daljnju analizu kako bi se smanjila nejednakost među stanovništvom, poboljšao položaj stanovnika, smanjilo siromaštvo ili povećala zaposlenost, u različitim regijama u državi ili između više država.

Hipoteza 1: Prosječni porez na dohodak po stanovniku značajno utječe na dohodak.

Jedna od najčešćih varijabli usko povezanih uz dohodak koja se spominje kroz razne radove porez je na dohodak. Porezna uprava RH [14] definira osnovicu poreza na dohodak kao ukupni iznos dohotka od nesamostalnog rada, dohotka od samostalne djelatnosti, dohotka od imovine i imovinskih prava, dohotka od kapitala i drugog dohotka, koji se ostvari u tuzemstvu i u inozemstvu, a umanjjen za osobni odbitak.

Auten i Carroll (1999) [6] u svom su radu o rastu dohotka tijekom 1980-ih zaključili da promjena u stopi poreza na dohodak ima značajan utjecaj na dohodak po stanovniku i da se porez ne smije zanemariti kod rasprava o dohotku, pogotovo oko nejednakosti u dohotku. U svojim istraživanju utjecaja poreza na dohodak na prosječni dohodak po stanovniku, Holcombe i Lacombe (2004) [17] zaključili su da rast poreza utječe negativno na rast dohotka.

U navedenim radovima uspoređuju se promjene u stopama poreza ili se uspoređuju teritorijalne jedinice s različitim poreznim stopama kao što su američke države i okruzi.

U RH je stopa poreza jednaka u svim teritorijalnim jedinicama. Ono što se razlikuje je stopa prireza na porez na dohodak o kojoj odlučuje predstavničko tijelo svake lokalne jedinice zasebno, no taj se podatak nije koristio zbog malog broja jedinica koje ga naplaćuju. Također je očit i utjecaj dohotka na porez na dohodak, a ne samo obrnuta veza. Unatoč toj problematici varijable poreza na dohodak, navedeni radovi su zaključili značajan utjecaj poreza na dohodak na raspodjelu dohotka i sami dohodak, pa se u ovom radu ta varijabla koristi, makar se rezultate treba uzeti s dozom opreza i zahtijevaju daljnju analizu na većem uzorku podataka kroz više vremenskih razdoblja.

Hipoteza 2: Stopa nezaposlenosti ima značajan utjecaj na dohodak.

Stopa nezaposlenost još je jedna od varijabli koju mnogi koriste u analizi dohotka. Kod velike nezaposlenosti manji broj stanovnika ostvaruje dohodak, pa bi se moglo pretpostaviti da stopa nezaposlenosti utječe na prosjek dohotka. Blejer i Guerrero (1990) [9] u radu o utjecaju makroekonomskih politika na dohodak u Filipinima zaključili su da jaka nezaposlenost negativno utječe na raspodjelu dohotka. Do istog su došli i u radu Mocan (1999) [7] gdje su još detaljnije podijelili samu nezaposlenost. Još jedan od razloga zašto bi stopa nezaposlenosti mogla imati značajan utjecaj na dohodak je činjenica da su u studiji evaluacije postojećeg modela za izračun indeksa razvijenosti Bogović, Drezgić, Čegar et. al. (2017) [15] zaključili da je stopa nezaposlenosti po jedinicama lokalne i područne (regionalne) samouprave bitan dio novog modela kojeg su predstavili za izračun danog indeksa, ali da zahtijeva dodatnu analizu. Preporučuju praćenje stope zaposlenosti i kretanja stanovništva kako bi se provjerilo što uzrokuje smanjenje stope nezaposlenosti - smanjenje broja radno sposobnog stanovništva ili povećanje stope zaposlenosti.

Hipoteza 3: Postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj na dohodak.

Može se pretpostaviti i da osoba sa završenim višim obrazovanjem prima veću plaću pa je potrebno ispitati utjecaj broja stanovnika sa završenim visokim obrazovanjem na prosječni dohodak. Su i Heshmati (2013) [2] u analizi odrednica dohotka u urbanoj i ruralnoj Kini zaključili su da obrazovanje ima pozitivan utjecaj na dohodak stanovnika. Povezanost stanovništva s višim obrazovanjem i većim rastom dohotka dokazali su i Auten i Carroll (1999) [6]. Mapa i Briones (2006) [4] su u svom istraživanju o varijablama koje utječu na dohodak u Filipinima pokazali da stupanj obrazovanja ima pozitivan utjecaj na dohodak, ali neznačajan u njegovom opisu.

Hipoteza 4: Broj stanovnika značajno i pozitivno utječe na rast dohotka.

Hipoteza 5: Postotak stanovnika mlađih od 14 godina negativno utječe na dohodak.

Hipoteza 6: Postotak stanovnika starijih od 65 godina negativno utječe na dohodak.

Veći broj stanovnika mogao bi biti povezan s većom raznolikosti u stanovništvu, npr. moglo bi biti više visoko obrazovanih stanovnika koji mogu ostvarivati velike dohotke, mogla bi biti veća šansa da pojedinci postignu veliki uspjeh u svojem polju zaposlenja i slično. S druge strane, što je veći broj stanovnika to se ukupan dohodak u danj jedinici dijeli s većim brojem kako bi se dobio prosječan dohodak po stanovniku. Dob i zastupljenost različitih dobnih skupina u stanovništvu također bi mogli imati utjecaj na prosječni dohodak. Očekuje se da veliki postotak mladog stanovništva ili onog u starijoj životnoj dobi ima utjecaj na prosječni dohodak po stanovniku jer najčešće osobe iz tih dobnih skupina nisu zaposlene, a dohodak koji eventualno ostvaruju je mnogo manji od radnog stanovništva. U RH raditi mogu osobe s navršениh 15 godina, uz suglasnost roditelja ili zakonskog skrbnika te taj posao mora biti u skladu s Pravilnikom o poslovima na kojima se ne smije zaposliti maloljetnik [19]. Iznos prosječne mirovine u RH [22] značajno je manji od prosječnog dohotka zaposlene osobe.

Monchuk et. al. (2007) [3] su u analizi ekonomskog rasta u Americi koristili različite dobne skupine i ispitivali njihov utjecaj na dohodak. Njihov rad pokazuje da broj stanovnika pozitivno i značajno utječe na rast dohotka i da postotak mlađih od 20 godina ima pozitivan, ali neznačajan utjecaj na dohodak. Također su pokazali da postotak stanovništva preko 65 godina te postotak stanovništva između 20 i 34 negativno i značajno utječe na dohodak. Nadalje, Mapa i Briones (2006) [4] su proučavali utjecaj raznih dobnih skupina na dohodak. Zaključili su da udio najmlađeg stanovništva, onog mlađeg od 14 godina, ima negativan i značajan utjecaj na rast dohotka. Značajnost dobi stanovnika u opisu dohotka istakli su i Auten i Carroll (1999) [6] gdje su pokazali da rast dohotka opada s godinama.

Hipoteza 7: Pristup moru ima pozitivan utjecaj na dohodak.

U RH turizam je važan izvor prihoda, a jedan od često hvaljenih ljepota zemlje su njezini primorski dijelovi. Iako je značajan izvor dohotka stanovnicima koji se njime bave, dio dohotka od priobalnog turizma nije prijavljen. Na primjer, dio dohotka naplaćen u gotovini ili u naturi ne prijavljuje se državi kako bi se izbjeglo plaćanje poreza. Literatura nudi

neujednačene rezultate po pitanju značajnosti utjecaja. Démurger et. al. (2002) [20] zaključili su da je u Kini pristup moru imao pozitivan utjecaj na rast dohotka. Mapa i Briones (2006) [4] dobili su da pristup moru ima neznačajan utjecaj na dohodak.

Hipoteza 8: Broj kućanstava s pristupom poljoprivrednom zemljištu negativno i značajno utječe na dohodak.

Gledajući zaposlenost ljudi u različitim sektorima, Monchuk et. al.(2007) [3] ispitali su utjecaj postotka dohotka dobivenog iz poljoprivrede i dobili da negativno i značajno utječe na dohodak. Zbog značajnosti poljoprivrede kao djelatnosti u RH, unatoč nedostupnosti tih podataka za Hrvatsku, u radu će se koristiti podatak o broju kućanstava s pristupom poljoprivrednom zemljištu dobiven iz popisa stanovništva iz 2011. godine [23]. Pretpostavka je da je dohodak od poljoprivrede niži od prosjeka pa što više stanovništva ostvaruje dohodak od poljoprivrede to će biti manji prosječni dohodak.

Hipoteza 9: Gustoća naseljenosti značajno i pozitivno utječe na dohodak.

Monchuk et. al. (2007) [3] bavili su se i utjecajem gustoće naseljenosti i zaključili da značajno i pozitivno utječe na dohodak. Još jedan razlog za uzimanje te varijable u ovom radu je činjenica da je to jedan od predloženih novih pokazatelja za indeks razvijenosti u studiji evaluacije postojećeg modela za izračun indeksa razvijenosti (Bogović, Drezgić, Čegar et. al. (2017) [15]). Predložili su uvođenje gustoće naseljenosti zbog izrazito negativnih demografskih kretanja na području RH, ali su je izbacili iz konačnog prijedloga jer se nije moglo jednostavno povezati s pozitivnim ili negativnim razvojem u teritorijalnim jedinicama.

Poglavlje 2

Metodologija

Ovaj rad koristi generalizirani linearni model. Ovo poglavlje temelji se na [1], [16], [18] i [8]. Prvo se predstavlja generalizirani linearni model i model izabran u ovom radu, potom se opisuje način na koji se procjenjuju parametri u modelu te na kraju način provjere učinkovitosti izabranog modela. Poglavlje završava kratkim objašnjenem interpretacije rezultata dobivenih programskim jezikom R.

2.1 Generalizirani linearni model

Generalizirani linearni model (GLM) poopćenje je modela linearne regresije koji pretpostavlja da na varijablu odziva Y (eng. *response*) utječu različite, međusobno nezavisne, kovarijate x_1, \dots, x_r (eng. *predictors*). Sastoji se od dvije komponente, slučajne i sistematične. Slučajne komponente generaliziranog linearnog modela su realizacije y_i odziva Y , koji se modelira kao slučajna varijabla, u $i = 1, \dots, n$ opservacija. Tada podatke možemo zapisati kao nizove parova: $(y_i, x_{ji}), i = 1, \dots, n, j = 1, \dots, r$. Kovarijate utječu na varijablu odziva preko sistematske komponente GLM-a, linearnog predviditelja η , za koju vrijedi $\eta := \beta_1 x_1 + \dots + \beta_r x_r$.

Generalizirani linearni model smanjuje ograničenja modela linearne regresije i omogućuje modeliranje ovisnosti varijable odziva i kovarijata s numeričkim i kategorijalnim vrijednostima. Također, varijabla odziva ne mora imati normalnu razdiobu, već razdiobu iz eksponencijalne familije razdioba.

Slučajna varijabla Y koja pripada eksponencijalnoj familiji ima gustoću oblika:

$$f(y; \theta, \varphi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right\} \quad (2.1)$$

za neke funkcije a , b i c te ovisi o prirodnom parametru θ i parametru disperzije ili skaliranja φ . Za gornje funkcije vrijedi:

- Funkcija a zove se funkcija disperzije i omogućuje veću fleksibilnost u modelu. Često se pretpostavlja $\varphi > 0$.
- Funkcija b dvaput je neprekidno diferencijabilna i b' je invertibilna.
- Funkcija c nema utjecaja na procjenu parametra u GLM pa se najčešće ignorira.
- Parametar θ ima vrijednosti u otvorenom skupu.

Najčešći oblik funkcije disperzije je $a(\varphi) = \varphi/w$, gdje je $w = w_i$ za opservaciju i , a w_i su već znane nenegativne težine (eng. *prior weights*) koje daju težinu svakoj opservaciji. Te težine najčešće su jednake jedan.

Za Y koja pripada eksponencijalnoj familiji razdioba vrijedi da očekivanje ovisi isključivo o parametru θ , tj. vrijedi $EY = b'(\theta)$. Također vrijedi $VarY = a(\varphi)b''(\theta)$. (Vidi [1].) Varijanca podataka ovisi o dvije komponente, parametru skaliranja i očekivanju. Zato se može zapisati u obliku $VarY = a(\varphi)V(\mu)$ kako bi se bolje prikazala ta ovisnost. Funkcija $\mu \mapsto V(\mu) = b''(\theta) = b''(b'^{-1}(\mu))$ zove se funkcija varijance.

Opći linearni modeli pretpostavljaju da postoji linearna veza između očekivanja varijable odziva i kovarijata tj. $EY_i = \sum_{j=1}^r \beta_j x_{j,i}$, $i = 1, \dots, n$. Kako u GLM kovarijate utječu na varijablu odziva preko linearnog predviditelja, očekivanje od varijable odziva glatka je i invertibilna funkcija linearnog predviditelja η oblika $b' \circ h$ za neku funkciju h tj.

$$\theta = h(\eta) \quad i \quad \mu = EY = b'(\theta) = b'(h(\eta))$$

pa je $\mu = g^{-1}(\eta)$ za funkciju veze $g := h^{-1} \circ b'^{-1}$. Dakle, kod generaliziranih linearnih modela pretpostavlja se $EY_i = g^{-1}\left(\sum_{j=1}^r \beta_j x_{j,i}\right)$, $i = 1, \dots, n$.

Prirodni parametar θ također je glatka funkcija od η jer vrijedi $\theta = b'^{-1}(\mu) = h(\eta)$. U posebnom slučaju kada vrijedi $\theta \equiv \eta$, funkcija $g = b'^{-1}$ naziva se kanonska funkcija veze. Neke od važnijih kanonskih veza su logit funkcija $\log(\mu/(1 - \mu))$ za binomnu razdiobu, logaritam $\log \mu$ za Poissonovu razdiobu, identiteta μ za normalnu, te inverz μ^{-1} za gama razdiobu. Kanonske funkcije veze su jednostavnije, ali nije ih nužno koristiti.

Izabrani model

U ovom radu koristi se normalna distribucija iz eksponencijalne familije za GLM tj. funkcija distribucije je oblika:

$$f(y; \theta, \varphi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\}. \quad (2.2)$$

Ako se zapiše kao:

$$f(y; \theta, \varphi) = \exp\left\{\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log 2\pi\sigma^2\right)\right\},$$

onda uz oznake:

$$\theta := \mu, \quad \varphi := \sigma^2,$$

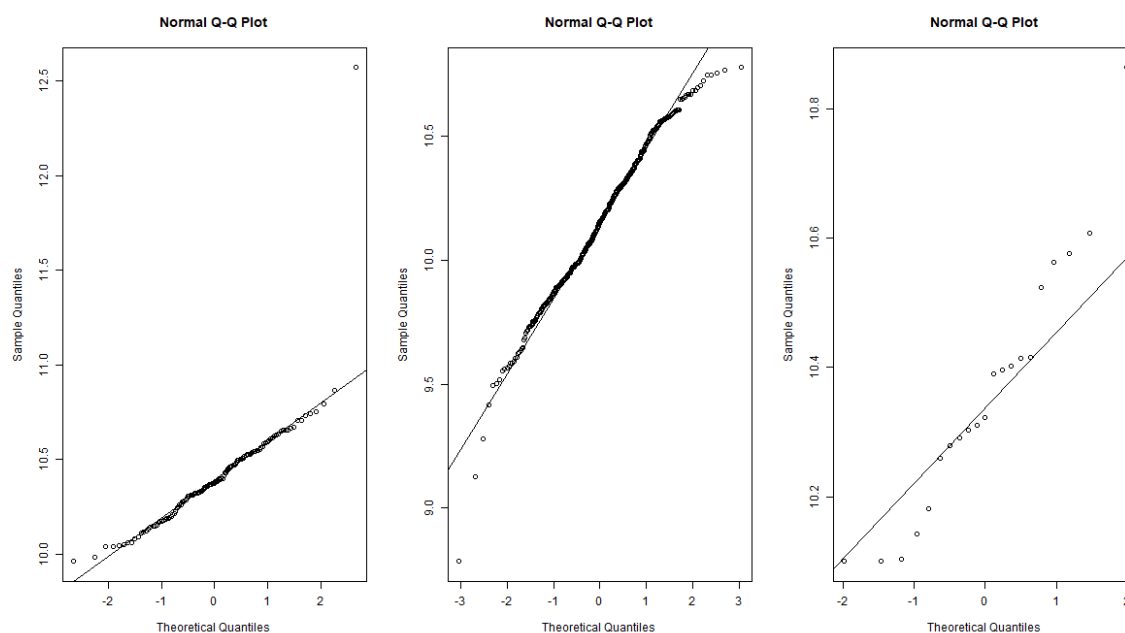
$$a(\varphi) := \varphi, \quad b(\theta) := \theta^2/2,$$

$$c(y, \varphi) := -\frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log 2\pi\sigma^2\right)$$

se jednadžba (2.2) može zapisati u obliku (2.1). Vidi se da je prirodni parametar normalne distribucije μ , a parametar skaliranja σ^2 . Također vrijedi $EY = b'(\theta) = \theta = \mu$ te $VarY = a(\varphi)b''(\theta) = \varphi = \sigma^2$ što se i očekuje za normalnu distribuciju.

Lilliefors test, koji se bazira na Kolmogorov-Smirnovljevom testu, testira nul hipotezu da podaci dolaze iz normalne razdiobe. Pri provjeri normalnosti podataka varijable odziva, po Lilliefors testu u R-u dobiveno je da se može pretpostaviti da logaritmirani podaci prosječnog dohotka po stanovniku u općinama, gradovima i županijama dolaze iz približno normalne razdiobe. Isto se može i vidjeti iz grafova na Slici 2.1.

Zbog toga umjesto podataka prosječnog dohotka po stanovniku koristit će se logaritam tih podataka, a za funkciju veze koristi se kanonska veza - identiteta.



Izvor: Izrada autorice

Slika 2.1: Provjera normalnosti varijable odziva

2.2 Procjena parametara

Standardna procedura procjene parametara u generaliziranim linearnim modelima korištenje je maksimuma funkcije vjerodostojnosti (MLE metoda).

Za funkciju gustoće $f(y, \theta)$ gdje je y opažanje uz dani parametar θ , funkcija log-vjerodostojnosti je $l(\theta; y) := \log f(y; \theta)$. Ako ima $n \in \mathbb{N}$ opažanja, onda je log-vjerodostojnost za nezavisna opažanja y_1, \dots, y_n uz svoje respektivne parametre $\theta_1, \dots, \theta_n$ jednaka

$$l(\theta_1, \dots, \theta_n; y_1, \dots, y_n) = \sum_{i=1}^n \log(f_i(y_i; \theta_i)).$$

Parametri se najčešće procjenjuju traženjem lokalnog maksimuma funkcije vjerodostojnosti.

Osim za jednostavne modele pronaći te maksimume je složeno te se koriste numerički algoritmi. U R-u naredba *glm* koristi Fisherov algoritam bodovanja (eng. *Fisher scoring*

algorithm) za izračun maksimuma funkcije vjerodostojnosti za parametre β_j , u oznaci $\hat{\beta}_j$, za $j = 1, \dots, r$.

Procijenjeni parametri $\hat{\beta}_j$ rješenja su izjednačenja funkcije rezultata (eng. *score function*) $U(\beta_j)$ s nulom, tj.

$$U(\beta_j) := \sum_{i=1}^n (y_i - \mu_i)x_{j,i} = 0, \quad j = 1, \dots, r.$$

Algoritam računa $\hat{\beta}_j$ tako da iterativno doraduje radne procjene (eng. *working estimates*) dok ne dođe do konvergencije. Svaka iteracija sastoji se od rješavanja jednadžbe koja koristi funkciju rezultata i Fisherovu informaciju $I_{j,k}(\beta) := \sum_{i=1}^n \mu_i x_{j,i} x_{k,i}$, $j, k = 1, \dots, r$.

Svaka iteracija algoritma je ekvivalentna regresiji najmanjih kvadrata radnih odgovora (eng. *working responses*) $z_i := \eta_i + \frac{d\eta_i}{d\mu_i}(y_i - \mu_i)$ na kovarijate $x_{j,i}$ koristeći radne težine (eng. *working weights*) $W_i := \frac{w_i}{V(\mu)(d\eta_i/d\mu_i)^2}$. U svakoj iteraciji z_i i W_i dobiju nove vrijednosti i regresija se ponovi na novim radnim koeficijentima $\hat{\beta}_j^k$ (procjena od β_j u iteraciji k). Iz radnih koeficijenata izračuna se novi η_i i vrijednosti $\mu_i = g^{-1}(\eta_i)$ te se iteracija ponovi.

Ovaj algoritam najčešće konvergira poprilično brzo (manje od 10 iteracija) te ne koristi φ .

2.3 Odabir optimalnog modela

Potencijalno, model sa proizvoljno mnogo kovarijata mogao bi postići savršenu predikciju, tj. riješio bi se sustav:

$$y_i = g^{-1}(\eta_i) = g^{-1}(\beta_1 x_{i,1} + \dots + \beta_r x_{i,r}).$$

Takav model zove se zasićeni i nije koristan sam po sebi. On se koristi tako da ga se uspoređuje s drugim modelima kako bi se ispitali koliko su oni zaista dobri.

Ako je sa \hat{l} označena vrijednost maksimuma funkcija log-vjerodostojnosti za model koji se koristi, a sa \hat{l}_F maksimiziranu log-vjerodostojnost u zasićenom modelu, onda se može definirati skalirana devijanca modela kao

$$d_M = 2(\hat{l}_F - \hat{l}).$$

Neskalirana ili rezidualna devijanca za odabrani model definira se kao

$$D_M := d_M \varphi.$$

Ona je općenito oblika

$$D_M = \sum_{i=1}^n w_i d(y_i, \mu_i)$$

gdje funkcija d ovisi o korištenoj eksponencijalnoj familiji razdioba. Za GLM gdje se koristi normalna razdioba vrijedi $D_M = \sum_{i=1}^n (y_i - \mu_i)^2$ što je kvadrirana euklidska udaljenost između vrijednosti y_i i μ_i .

Ako je u modelu r kovarijata, linearni procjenitelj je $\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$ te su se maksimiziranjem funkcije log-vjerodostojnosti dobili procjenitelji $\hat{\beta}_i$, $i = 1, \dots, r$ i dosegnuo se maksimum \hat{l} , tada je moguće testirati koliko su bitne pojedinačne kovarijate u modeliranju varijable odziva. Nul hipoteza koja testira nebitnost zadnjih $r - q$ kovarijata je $H_0 : \beta_{q+1} = \dots = \beta_r = 0$, za $q < r$. Testiranje se provodi u tri koraka:

1. Koristeći MLE procijene se parametri u punom i restringiranom modelu (modelu koji sadrži kovarijate x_1, \dots, x_q).
2. Pronađe se maksimalna log-vjerodostojnost u punom i restringiranom modelu, \hat{l} i \hat{l}_0 respektivno.
3. Izračuna se statistika omjera vjerodostojnosti (eng. *log-likelihood ratio statistics*) $2(\hat{l} - \hat{l}_0)$ te ako je ona iznad kritične vrijednosti za izabrani nivo značajnosti i razdiobu χ^2 sa $r - q$ stupnjeva slobode, odbacuje se nul hipoteza. To znači da je model značajno različit od zasićenog modela te da su dodatni parametri potrebni.

Testna statistika je ustvari razlika devijanci, tj. $2(\hat{l} - \hat{l}_0) = 2(\hat{l} - \hat{l}_F + \hat{l}_F - \hat{l}_0) = d_{M_0} - d_M$ što je u slučaju normalne razdiobe egzaktna jednakost, dok je za ostale test točan samo asimptotski.

Još jedan način usporedbe modela je Akaike informacijski kriterij (AIC) koji služi za usporedbu prilagođenosti različitih modela koji su dobiveni istom metodom. Računa se po formuli $AIC := 2r - 2l$, gdje je r broj procijenjenih parametara, a l maksimalna vrijednost funkcije vjerodostojnosti. Ta vrijednost sama ne znači ništa, ali kada se uspoređuju modeli, bolji je onaj sa manjom AIC vrijednošću.

2.4 Analiza reziduala

Reziduali se mogu koristiti za provjeru kvalitete prilagodbe modela te predstavljaju odmak predviđenih vrijednosti od opaženih.

Baza za reziduale u linearnoj regresiji su reziduali odgovora (eng. *response residuals*) koji se računaju kao razlika između opaženih vrijednosti i predviđenih, tj. $y_i - \hat{\mu}_i$. Takvi reziduali ne odgovaraju za GLM jer se varijanca eksponencijalnih familija distribucije bazira na srednjoj vrijednosti. Zato se definiraju Pearsonovi ostaci koji podijele reziduale odgovora sa procijenjenom standardnom devijacijom varijable odziva, tj.

$$r_i^P := \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(\hat{\mu}_i)}}.$$

Zajedno sa Pearsonovim ostacima, ostaci u devijanci najčešće se koriste u softverskim paketima. Ostaci u devijanci su definirani kao

$$r_i^D := \text{sign}(y_i - \hat{\mu}_i) \sqrt{w_i d(y_i - \hat{\mu}_i)}.$$

Za normalnu distribuciju su ostaci devijance $r_i^D = (y_i - \hat{\mu}_i) \sqrt{w_i}$.

Ako je model adekvatan, ostaci imaju aproksimativno normalnu razdiobu. To se provjerava crtanjem histograma koji se uspoređuje sa funkcijom gustoće normalne varijable ili grafom kvantila. Graf kvantila za normalnu razdiobu sastoji se od uređenih parova $(x_i, y_{(i)})$, $i = 1, \dots$, gdje su $y_{(i)}$ sortirana opažanja, a $x_i = \phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$ za ϕ funkciju distribucije jedinične normalne razdiobe. Ako su ostaci aproksimativno na pravcu tada to podupire hipotezu da dolaze iz normalne razdiobe.

U R-u se graf kvantila crta naredbom *qqnorm* te je moguće raznim testovima utvrditi normalnost podataka, kao što je na primjer Liliefors test.

Crtanjem Pearsonovih modela naspram predviđenih vrijednosti treba se provjeriti postoji li neki neobjašnjeni sistematski uzorak na grafu. Ako model dovoljno dobro opisuje varijablu odziva, na grafu ne bi trebalo biti očitog uzorka. Bilokakav trend na grafu ukazuje na to da se treba uključiti dodatan član u model.

2.5 GLM u R-u

U programskom jeziku i okruženju R dostupna je naredba *glm*. Specifično, u ovom radu koristi se naredba

```
glm(varijabla_odziva ~ kovarijata_1+...+kovarijata_r, family=gaussian)
```

gdje je r broj kovarijata u modelu. *Glm* pomoću Fisherovih iteracija procjenjuje parametre te daje njihove standardne greške i odgovarajuće p-vrijednosti. Standardna greška govori o varijabilnosti procjene parametra, a t vrijednost dobije se kao procjena koeficijenta podijeljena sa standardnom greškom. $Pr(> |t|)$ ili p-vrijednost daje vjerojatnost povezanu sa danom t vrijednošću. Ta vrijednost govori koliko svaka od kovarijata dobro predviđa vrijednost varijable odziva jer testira nul hipotezu da kovarijata nije značajna. Ako je p-vrijednost manja od odabrane razine značajnosti, odbacuje se nul hipoteza te je kovarijata značajna. Kao i u mnogim radovima, u ovom se koristi razina značajnosti od 5%.

Naredba *glm* također daje nul devijancu i rezidualnu devijancu te *AIC* vrijednost. Nul devijanca nam govori koliko dobro varijabla odziva može biti previđena samo sa slobodnim članom β_0 , a rezidualna devijanca koliko dobro može biti predviđena sa tim specifičnim modelom. Što je vrijednost niža, to model bolje predviđa vrijednosti varijable odziva. Te vrijednosti služe za provjeru korisnosti modela koristeći χ^2 statistiku na razlici nul devijance i rezidualne devijance sa stupnjevima slobode jednakim razlici stupnjevima slobode nul devijance i stupnjevima slobode rezidualne devijance. Ako je p-vrijednost niža od razine značajnosti, puni model je bolji nego model koji se sastoji samo od slobodnog člana.

U R-u se koristi naredba *anova(model)* za provjeru značajnosti kovarijata u modelu u smislu smanjenja rezidualne devijance u modelu dodavanjem varijable u model. Ona uzima u obzir redosljed dodavanja kovarijata u model i omogućuje lakšu analizu p-vrijednosti ako se provede naredba *anova(model, test = "Chisq")*. U slučaju usporedbe dva modela, može se koristiti naredba *anova(Model1, Model2)* koji provjerava koliko su dva modela dobro prilagođena podacima u slučaju da je jedan model restringirana verzija drugog modela. Nulta hipoteza koja se testira je da je *Model1* bolji od *Modela2*, pa ako je dobivena razlika u devijancama veća od kritične vrijednosti, tada je *Model2* bolji i dobacujemo nul hipotezu.

Poglavlje 3

Empirijska analiza

Ovo poglavlje nudi deskriptivnu analizu varijable odziva i kovarijata za općine, gradove i županije RH. Za svaku varijablu prikazana je najmanja, najveća i srednja vrijednost varijable zajedno s medijanom i standardnom devijacijom. Srednja vrijednost je zbroj svih podataka podijeljen s njihovim brojem, a medijan je vrijednost za koju vrijedi da je 50% podataka manje od ili jednako toj vrijednosti i 50% podataka je veće ili jednako njoj [10]. Standardna devijacija je prosječno srednje kvadratno odstupanje podataka od njihove srednje vrijednosti. Za te skupine podataka se također analizira korelacija varijable odziva i kovarijata te kovarijata međusobno. Koeficijent korelacije mjeri snagu linearne veze između dvije varijable i poprima vrijednosti između -1 i 1 te što je koeficijent bliži tim vrijednostima, linearna je veza jača. Veća korelacija kovarijata sa varijablom odziva mogla bi dati naslutiti značajan utjecaj kovarijate, dok veća korelacija među kovarijatama može narušiti točnost procijenjenog modela. Izrazito visoko korelirane kovarijate objašnjavaju istu dimenziju varijable odziva te rezultati mogu biti netočno promijenjeni zbog te veze.

Nakon toga se provodi metoda predstavljena u poglavlju 2 za istu podjelu podataka. Prvo će se predstaviti dva tipa generaliziranih linearnih modela u svakoj skupini podataka - model u kojem se ispituje utjecaj svake varijable zasebno na dohodak te puni model koji sadrži sve varijable. Potom se pomoću programskog okruženja R procijene parametri danih modela i analiziraju dobiveni rezultati. Zatim se analizira točnost punog modela te provjerava postoji li restringirani model, tj. model sa manjim brojem varijabli, koji jednako dobro ili bolje opisuje dane podatke. Restringirani model uspoređuje se s punim pomoću statistike omjera vjerodostojnosti te se izabere onaj optimalniji. Na kraju se provjerava kvaliteta odabranog optimalnog modela pomoću analize reziduala.

3.1 Deskriptivna analiza

Varijabla odziva u ovom modelu prosječan je dohodak po stanovniku u godini dana. Dohodak je razlika između primitaka priteklih u poreznom razdoblju i izdataka nastalih u istom razdoblju i utvrđuje se kao godišnji ili konačan dohodak. Izvori dohotka primici su ostvareni od nesamostalnog rada, samostalne djelatnosti, imovine i imovinskih prava, kapitala i drugih primitaka. (Vidi Porezna uprava RH [13].)

Podaci dohotka u godini dana dobiveni su od Ministarstva regionalnog razvoja i fondova Europske unije (MRRFEU) za potrebe ovog rada. Potom su podaci podijeljeni brojem stanovnika preuzetim iz popisa stanovništva [23] kako bi se dobio prosječan dohodak po stanovniku koji će se u daljnjim potpoglavljima analizirati za općine, gradove i županije zasebno.

Oznaka varijable	Opis varijable	Izvor podataka
Doh	Dohodak po stanovniku u kunama u 2019.	Ministarstvo regionalnog razvoja i fondova Europske unije

Izvor: Izrada autorice

Tablica 3.1: Varijabla odziva

U Tablici 3.2 prikazane su sve varijable kojima se bavi ovaj rad, korištene oznake te izvor otkuda su podaci dobiveni.

Porez na dohodak u godini dana preuzet je sa stranice Ministarstva financija [5] za 2019. godinu. Osnovica poreza na dohodak ukupni je iznos dohotka od nesamostalnog rada, dohotka od samostalne djelatnosti, dohotka od imovine i imovinskih prava, dohotka od kapitala i drugog dohotka, koji se ostvari u tuzemstvu i u inozemstvu, a umanjeno za osobni odbitak. (Vidi Porezna uprava RH [14].) Također, koriste se podaci o stopi nezaposlenosti iz 2019. dobiveni od MRRFEU za potrebe ovog rada. Stopa nezaposlenosti je udio nezaposlenog stanovništva u ukupnom radno sposobnom stanovništvu.

Podaci o stanovništvu dobiveni su iz zadnjeg popisa stanovništva dostupnog za vrijeme pisanja ovog rada, popisa iz 2011. [23] Iz njega su preuzeti podaci o broju stanovnika u jedinicama lokalne i područne (regionalne) samouprave, broju stanovnika do 14 godina te onih starijih od 65, broju kućanstava s pristupom poljoprivrednom zemljištu, broju stanovnika sa završenim visokim obrazovanjem (sa završenim stručnim studijem, sveučilišnim studijem ili s doktoratom znanosti). Gustoća naseljenosti, izračunata kao broj stanovnika

po kvadratnom kilometru, bila je dostupna samo za županije.

Kategorijalna varijabla pristupa moru, koja poprima vrijednost 1 ako jedinica samouprave ima pristup moru, a 0 ako nema, dobivena je od Instituta za javne financije (IJF) za korištenje u ovom radu.

Podaci su prilagođeni kako bi odgovarali kovarijatima tj. porez na dohodak podijeljen je s brojem stanovnika i tako dobiven prosječan porez na dohodak po stanovniku. Broj stanovnika mlađih od 14 godina i broj stanovnika starijih od 65 godina podijeljen je s ukupnim brojem stanovnika i pomnožen sa 100, kako bi se dobio postotak.

Oznaka varijable	Opis varijable	Izvor podataka
Por	Prosječni porez po stanovniku u kunama u 2019.	Ministarstvo financija [5]
St_nez	Stopa nezaposlenosti u 2019.	Ministarstvo regionalnog razvoja i fondova Europske unije
Obr	Postotak stanovništva sa završenim tercijarnim obrazovanjem u 2011.	Popis stanovništva 2011. [23]
Br_stan	Broj stanovnika u 2011.	Popis stanovništva 2011. [23]
ML_stan	Postotak mladog stanovništva (do 14 godina) u 2011.	Popis stanovništva 2011. [23]
St_stan	Postotak stanovništva starijeg od 65 godina u 2011.	Popis stanovništva 2011. [23]
More	Kategorijalna varijabla pristupa moru	Institut za javne financije
Polj	Broj kućanstava s pristupom poljoprivrednom zemljištu u 2011.	Popis stanovništva 2011. [23]
Gus	Gustoća naseljenosti u 2011. (Br_stan/km^2)	Popis stanovništva 2011. [23]

Izvor: Izrada autorice

Tablica 3.2: Popis kovarijata

Općine

U Tablici 3.3 vidimo analizu prosječnog godišnjeg dohotka po stanovniku u općinama RH te njegove logaritmirane vrijednosti koje koristimo u modelu.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Doh	6 523	48 049	26 714	25 767	7 698.927
log(Doh)	8.78	10.78	10.15	10.16	0.30

Izvor: Izrada autorice

Tablica 3.3: Deskriptivna analiza varijable odziva za općine

Najmanji prosječan godišnji dohodak po stanovniku iznosa 6 523 kn pripada općini Svetoj Nedelji u Istarskoj županiji, dok je najveći u općini Omišalj u Primorsko-goranskoj županiji i iznosi 48 049 kn. Prosječna vrijednost varijable odziva za općine iznosi 26 714 kn. Dodatnom analizom prvog i trećeg kvantila podataka dobije se da prosječan dohodak po stanovniku manji od 20 859 kn imaju stanovnici u 25% općina, dok 25% ima više od 31 428 kn.

U Tablici 3.4 prikazana je deskriptivna statistika za kovarijate za općine RH.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Por	269.20	8 717.50	2 237.00	1 972.70	967.68
St_nez	0.59	36.37	8.67	6.44	6.60
Obr	0.84	25.65	6.36	5.32	3.69
Br_stan	239	18 059	2 974	2 412	2 074.90
MI_stan	4.16	26.28	15.15	15.16	3.01
St_stan	10.40	61.09	20.11	18.88	5.55
More	0	1	0.19	0	0.39
Polj	104	5 576	995	823	670.73

Izvor: Izrada autorice

Tablica 3.4: Deskriptivna analiza kovarijata za općine

Najmanji prosječni porez na dohodak u godini dana od 269.2 kn imaju Lokvičići u Splitsko-dalmatinskoj županiji, dok je najveći porez od 8 717.50 kn iz općine Vir u Zadar-

skoj županiji. Najniža stopa nezaposlenosti od 0.59% je u općini Hrašćina u Krapinsko-zagorskoj županiji, a najveća, koja iznosi 36.37%, je u općini Kistanje u Šibensko-kninskoj županiji. Najmanji postotak visoko obrazovanog stanovništva, 0.84%, nalazi se u općini Civljane u Šibensko-kninskoj županiji, a 25.65% visoko obrazovanog stanovništva nalazi se u Kostreni u Primorsko-goranskoj županiji. 239 stanovnika ima općina Civljane, dok maksimum od 18 059 stanovnika ima općina Sveta Nedelja u Istarskoj županiji. Najmanji postotak stanovništva mlađeg od 14 godina, 4.16%, ima općina Ervenik u Šibensko-kninskoj županiji, a 26.28% ih ima u općini Voćin u Virovitičko-podravskoj županiji. 10.4% stanovništva starijeg od 65 godina ima općina Viškovo u Primorsko-goranskoj županiji, a maksimum od 61.09% ostvaruje se u općini Civljane. 104 kućanstava u općini Zadvarje u Splitsko-dalmatinskoj županiji ima pristup poljoprivrednom zemljištu, dok je u općini Sveta Nedelja u Istarskoj županiji taj broj 5 576.

Korelacije zaokružene na dvije decimale između varijable odziva i kovarijata te kovarijata međusobno prikazane su u Tablici 3.5. Veća korelacija s varijablom logaritmiranog dohotka mogla bi dati naslutiti značajan utjecaj kovarijata na nju. Najveću korelaciju s varijablom odziva imaju varijabla poreza po stanovniku, stopa nezaposlenosti, postotak visoko obrazovanog stanovništva te varijabla pristupa moru. Među kovarijatima je najveća korelacija između varijable broja stanovnika i broja kućanstava s pristupom poljoprivrednom zemljištu. Točnost procijenjenog modela može se narušiti velikom korelacijom među varijablama, pa će se provesti procjena i na smanjenom modelu bez barem jedne od tih varijabli.

	log(Doh)	Por	St_nez	Obr	Br_stan	MI_stan	St_stan	More	Polj
log(Doh)	1	0.42	-0.66	0.61	0.01	-0.20	-0.13	0.43	0.03
Por	0.42	1	-0.24	0.55	-0.14	-0.26	0.10	0.63	-0.08
St_nez	-0.66	-0.24	1	-0.33	-0.05	0.07	0.21	-0.24	-0.01
Obr	0.61	0.55	-0.33	1	0.13	-0.35	0.04	0.67	0.20
Br_stan	0.01	-0.14	-0.05	0.13	1	0.29	-0.40	-0.08	0.98
MI_stan	-0.20	-0.26	0.07	-0.35	0.29	1	-0.74	-0.27	0.19
St_stan	-0.13	0.10	0.21	0.04	-0.40	-0.74	1	0.14	-0.32
More	0.43	0.63	-0.24	0.67	-0.08	-0.27	0.14	1	-0.03
Polj	0.03	-0.08	-0.01	0.20	0.98	0.19	-0.32	-0.03	1

Izvor: Izrada autorice

Tablica 3.5: Tablica korelacija za općine

Gradovi

Sljedeća analiza obavit će se na gradovima RH. Deskriptivna analiza prosječnog godišnjeg dohotka po stanovniku te pripadne logaritmirane vrijednosti prikazane su u Tablici 3.6.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Doh	21 187	288 050	34 993	32 119	23 427.29
log(Doh)	9.96	12.57	10.40	10.38	0.27

Izvor: Izrada autorice

Tablica 3.6: Deskriptivna analiza varijable odziva za gradove

Najmanji prosječni dohodak po stanovniku od 21 187 kn pripada Pleternici u Požeško-slavonskoj županiji, a najveći gradu Zagrebu u kojem taj prosjek iznosi 288 050 kn. Dodatnom analizom prvog i trećeg kvantila podataka dobije se da prosječan dohodak po stanovniku manji od 28 470 kn ima 25% gradova. Prosječan dohodak po stanovniku između 28 470 i 37 352 kuna ima 50% stanovništva, dok 25% ima između 37 352 kn i maksimuma. Velikoj razlici između trećeg kvantila koji iznosi 37 352 i maksimuma od 288 050 najviše pridonosi uključenje grada Zagreba u analizu.

U Tablici 3.7 prikazana je deskriptivna statistika za kovarijate za gradove RH.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Por	874.50	22 966.30	2 957.90	2 522.60	1 994.67
St_nez	2.12	20.40	7.20	6.66	4.01
Obr	3.79	24.72	11.43	11.17	4.19
Br_stan	1 526	790 017	23 532	10 456	72 298.66
ML_stan	10.73	23.16	15.04	15.10	2.15
St_stan	10.52	27.40	17.89	17.14	3.07
More	0	1	0.29	0	0.46
Polj	637	303 441	8 541	3 572	27 726.46

Izvor: Izrada autorice

Tablica 3.7: Deskriptivna analiza kovarijata za gradove

Najmanji prosječni porez na dohodak u godini dana od 874.50 kn ima Skradin u Šibensko-kninskoj županiji, dok je najveći porez od 22 966.30 kn iz Svete Nedelje u Zagrebačkoj županiji. Najniža stopa nezaposlenosti od 2.12% je u Križevcima u Koprivničko-križevačkoj županiji, a najveća, koja iznosi 20.4%, je u Belom Manastiru u Osječko-baranjska županija. Najmanji postotak visoko obrazovanog stanovništva, 3.79%, nalazi se u Skradinu, a 24.72% visoko obrazovanog stanovništva nalazi se u Zagrebu. 1 526 stanovnika ima Komiža u Splitsko-dalmatinskoj županiji, dok maksimum od 790 017 stanovnika ima grad Zagreb. Najmanji postotak stanovništva mlađeg od 14 godina, 10.73%, ima Opatija u Primorsko-goranskoj županiji, a 23.16% ih ima u Imotskom u Splitsko-dalmatinskoj županiji. 10.52% stanovništva starijeg od 65 godina ima Solin u Splitsko-dalmatinskoj županiji, a maksimum od 27.40% ostvaruje se u Skradinu. 637 kućanstava u Komiži ima pristup poljoprivrednom zemljištu, dok je u Zagrebu taj broj 303 441.

U Tablici 3.7 prikazane su korelacije zaokružene na dvije decimale između varijable odziva i kovarijata te kovarijata međusobno. Najveća korelacija s varijablom logaritmiranog dohotka je s kovarijatom poreza po stanovniku, stope nezaposlenosti i postotka visoko obrazovanog stanovništva. To bi moglo pokazivati značajan utjecaj tih kovarijata na prosječni dohodak po stanovniku. Između kovarijata je najveća korelacija među brojem stanovnika i brojem kućanstava s pristupom poljoprivrednom zemljištu. Točnost procijenjenog modela može se narušiti velikom korelacijom među kovarijatama, pa će se provesti procjena i na restringiranom modelu bez jedne od tih varijabli.

	log(Doh)	Por	St_nez	Obr	Br_stan	MI_stan	St_stan	More	Polj
log(Doh)	1	0.77	-0.54	0.49	0.19	-0.35	0.02	0.24	0.19
Por	0.77	1	-0.21	0.15	0.15	-0.22	0.16	0.18	0.15
St_nez	-0.54	-0.21	1	-0.32	-0.06	0.34	0.03	-0.25	-0.06
Obr	0.49	0.15	-0.32	1	0.43	-0.29	-0.11	0.47	0.43
Br_stan	0.19	0.15	-0.06	0.43	1	-0.01	-0.06	-0.01	0.99
MI_stan	-0.35	-0.22	0.34	-0.29	-0.01	1	-0.65	-0.29	-0.02
St_stan	0.02	0.16	0.03	-0.11	-0.06	-0.65	1	0.13	-0.05
More	0.24	0.18	-0.25	0.47	-0.01	-0.29	0.13	1	-0.01
Polj	0.19	0.15	-0.06	0.43	0.99	-0.02	-0.05	-0.01	1

Izvor: Izrada autorice

Tablica 3.8: Tablica korelacija za gradove

Županije

U Tablici 3.9 vidimo analizu prosječnog dohotka po stanovniku u županijama RH te njegove logaritmirane vrijednosti koju koristimo u modelu.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Doh	24 339	52 259	32 141	30 405	66 95.29
log(Doh)	10.10	10.86	10.36	10.32	0.19

Izvor: Izrada autorice

Tablica 3.9: Deskriptivna analiza varijable odziva za županije

Najmanji dohodak po stanovniku iznosa 24 339 kn pripada Vukovarsko-srijemskoj županiji, dok najveći ponovno pripada Gradu Zagrebu i iznosi 52 259 kn. Prosječna vrijednost varijable odziva za općine iznosi 32 141 kn. Dodatnom analizom prvog i trećeg kvantila podataka dobije se da prosječan dohodak po stanovniku manji od 28 536 kn ima 25% županija, dok 25% ima više od 33 394 kn.

U Tablici 3.10 prikazana je deskriptivna statistika za kovarijate za županije RH.

Ime varijable	Minimalna vrijednost	Maksimalna vrijednost	Srednja vrijednost	Medijan	Standardna devijacija
Por	441.50	6 978.60	865.60	545.60	1 405.42
St_nez	3.00	13.00	7.24	7.00	3.39
Obr	6.93	24.72	11.22	10.08	4.28
Br_stan	50 927	790 017	204 042	158 575	165 283.80
MI_stan	12.48	17.07	15.31	15.41	1.33
St_stan	15.60	24.69	18.24	17.73	2.11
More	0	1	0.33	0	0.48
Polj	19 617	303 441	72 335	52 056	62 932.81
Gus	9.51	1232.48	126.35	73.15	255.96

Izvor: Izrada autorice

Tablica 3.10: Deskriptivna analiza kovarijata za županije

Najmanji prosječni porez na dohodak u godini dana od 441.5 kn ima Virovitičko-podravska županija, dok je najveći porez od 6 978.6 kn iz Grada Zagreba. Najniža stopa

nezaposlenosti od 3% je u Istarskoj i Varaždinskoj županiji, a najveća, koja iznosi 13%, je u Osječko-baranjska i Virovitičko-podravskoj županiji. Najmanji postotak visoko obrazovanog stanovništva, 6.93%, nalazi se u Virovitičko-podravskoj županiji, a 24.72% visoko obrazovanog stanovništva nalazi se u Gradu Zagrebu. 50 927 stanovnika ima Ličko-senjska županija, dok maksimum od 790 017 stanovnika ima Grad Zagreb. Najmanji postotak stanovništva mlađeg od 14 godina, 12.48%, ima Primorsko-goranska županija, a 17.07% ih ima u Brodsko-posavskoj županiji. 15.6% stanovništva starijeg od 65 godina ima Međimurska županija, a maksimum od 24.69% ostvaruje se u Ličko-senjskoj županiji. 19 617 kućanstava u Ličko-senjskoj županiji ima pristup poljoprivrednom zemljištu, dok je u Gradu Zagrebu taj broj 303 441. Najmanja gustoća stanovništva je u Ličko-senjskoj županiji i iznosi 9.51 *stan/km²*, a najveća u Gradu Zagrebu od 1232.48 *stan/km²*.

U Tablici 3.10 prikazane su korelacije zaokružene na dvije decimale između varijable odziva i kovarijata te kovarijata međusobno. Postoje visoke korelacije između varijable poraza po stanovniku i postotka visoko obrazovanog stanovništva, broja stanovnika te broja kućanstava s pristupom poljoprivrednom zemljištu. Također, visoka vrijednost korelacije je i između varijable postotka visoko obrazovanog stanovništva i broja stanovnika te broja kućanstava s pristupom poljoprivrednom zemljištu. Ponovila se i jaka korelacija između broja stanovnika i broja kućanstava s pristupom poljoprivrednom zemljištu. Gustoća stanovnika ima višu koreliranost s nekoliko drugih kovarijata. Provest će se i procjena na modelu bez nekih od varijabli među kojima je najveća korelacija kako se ne bi narušila točnost procijenjenog modela jakom korelacijom među kovarijatima.

	log(Doh)	Por	St_nez	Obr	Br_stan
log(Doh)	1	0.66	-0.62	0.86	0.69
Por	0.66	1	-0.27	0.76	0.83
St_nez	-0.62	-0.27	1	-0.34	-0.15
Obr	0.86	0.76	-0.34	1	0.81
Br_stan	0.69	0.83	-0.15	0.81	1
MI_stan	-0.47	-0.14	0.20	-0.37	-0.08
St_stan	-0.06	-0.11	0.14	-0.04	-0.33
More	0.33	-0.11	-0.11	0.41	-0.01
Polj	0.70	0.86	-0.15	0.84	0.99
Gus	0.63	0.99	-0.28	0.72	0.84

	MI_stan	St_stan	More	Polj	Gus
log(Doh)	-0.47	-0.06	0.33	0.70	0.63
Por	-0.14	-0.11	-0.11	0.86	0.99
St_nez	0.20	0.14	-0.11	-0.15	-0.28
Obr	-0.37	-0.04	0.41	0.84	0.72
Br_stan	-0.08	-0.33	-0.01	0.99	0.84
MI_stan	1	-0.64	-0.40	-0.14	-0.05
St_stan	-0.64	1	0.43	-0.05	-0.20
More	-0.40	0.43	1	0.01	-0.19
Polj	-0.14	-0.05	0.01	1	0.85
Gus	-0.05	-0.20	-0.19	0.85	1

Izvor: Izrada autorice

Tablica 3.11: Tablica korelacija za županije

Ukupna deskriptivna analiza

U ovom potpoglavlju predstavila se deskriptivna analiza podataka kako bi se vidjele vrijednosti s kojima analiza raspolaže te kako bi se bolje mogle naslutiti veze koje će se dalje ispitivati. Na primjer, općina Sveta Nedelja u Istarskoj županiji ima najmanji prosječan godišnji dohodak po stanovniku te najviše stanovnika i najveći broj kućanstava s pristupom poljoprivrednom zemljištu što bi odgovaralo negativnom utjecaju pristupa poljoprivrednom zemljištu, ali bi opovrgnulo pozitivan utjecaj broja stanovnika. U županijama se ističe Grad Zagreb s najvećim prosječnim dohotkom, najvećim prosječnim porezom, najviše visoko obrazovanih stanovnika, najviše stanovnika pa s time i najvećom gustoćom stanovnika.

Potom se za svaku skupinu podataka ispitala korelacija. Veći utjecaj kovarijate na varijablu odziva moglo bi se dati naslutiti iz veće korelacije. Za općine najveću korelaciju s varijablom odziva imaju varijabla prosječnog poreza po stanovniku, stopa nezaposlenosti, postotak visoko obrazovanog stanovništva te varijabla pristupa moru. Za gradove su to kovarijate: prosječan porez po stanovniku, stopa nezaposlenosti i postotak visoko obrazovanog stanovništva, a za županije sve varijable osim postotka stanovništva starijeg od 65 i varijable pristupa moru. S druge strane, veća korelacija među kovarijatama može narušiti točnost procijenjenog modela pa ih je bitno uočiti i pokušati pronaći model koji ne sadrži visoko korelirane kovarijate. U općinama visoko su korelirane varijable broja stanovnika i broja kućanstava s pristupom poljoprivrednom zemljištu te postotka stanovništva mlađeg od 14 i starijeg od 65 godina. U gradovima je najveća korelacija između broja stanovnika

i broja kućanstava s pristupom poljoprivrednom zemljištu. U županijama je bilo više parova kovarijata među kojima se javila velika korelacija, kao što su porez po stanovniku i postotak visoko obrazovanog stanovništva, porez po stanovniku i broj stanovnika, porez po stanovniku i broj kućanstava s pristupom poljoprivrednom zemljištu, postotak visoko obrazovanog stanovništva i broj stanovnika, postotak visoko obrazovanog stanovništva i broj kućanstava s pristupom poljoprivrednom zemljištu, broj stanovnika i broj kućanstava s pristupom poljoprivrednom zemljištu. Gustoća stanovnika za županije imala je višu koreliranost s nekoliko drugih kovarijata.

3.2 Generalizirani linearni model

Prvo će se provjeriti utjecaj svake varijable zasebno na varijablu odziva tako da se analiziraju modeli

$$\mu = \beta_0 + \beta_1 x_i \quad (3.1)$$

gdje je μ očekivanje varijable logaritmirane vrijednosti prosječnog dohotka po stanovniku tj. $\mu = E(\log(Doh))$ te x_i , $i = 1, \dots, r$ su varijable predstavljene u Tablici 3.2. Potom će se analizirati model oblika

$$\mu = \beta_0 + \beta_1 Por + \beta_2 St_nez + \beta_3 Obr + \beta_4 Br_stan + \beta_5 ML_stan + \beta_6 St_stan + \beta_7 More + \beta_8 Polj. \quad (3.2)$$

Za županije je dodana i varijabla gustoće stanovnika pa je tada puni model:

$$\mu = \beta_0 + \beta_1 Por + \beta_2 St_nez + \beta_3 Obr + \beta_4 Br_stan + \beta_5 ML_stan + \beta_6 St_stan + \beta_7 More + \beta_8 Polj + \beta_9 Gus. \quad (3.3)$$

Na kraju će se provjeriti adekvatnost tog modela te provjeriti postoji li neki restringirani model koji jednako dobro opisuje logaritmirani prosječni dohodak po stanovniku.

Općine

Analiza nezavisnih varijabli

Utjecaj pojedinih varijabli na varijablu odziva možemo vidjeti u Tablici 3.12 gdje su prikazani procijenjeni parametri β_0 i β_1 modela oblika (3.1), njihove standardne greške te odgovarajuće p-vrijednosti. Sve vrijednosti dobivene su naredbom *glm* u R-u.

Ime varijable		Procijenjen koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Por	β_0	9.8657	0.0328	300.72	$< 2e - 16$ ***
	β_1	0.0001	$1.3460 \cdot 10^{-5}$	9.46	$< 2e - 16$ ***
St_nez	β_0	10.4064	0.0178	585.19	$< 2e - 16$ ***
	β_1	-0.0295	0.0016	-18.08	$< 2e - 16$ ***
Obr	β_0	9.8403	0.0226	434.70	$< 2e - 16$ ***
	β_1	0.0488	0.0031	15.85	$< 2e - 16$ ***
Br_stan	β_0	10.1460	0.0250	405.27	$< 2e - 16$ ***
	β_1	$1.5546 \cdot 10^{-6}$	$6.9070 \cdot 10^{-6}$	0.23	0.8220
ML_stan	β_0	10.4553	0.0719	145.45	$< 2e - 16$ ***
	β_1	-0.0201	0.0047	-4.32	$1.93e - 05$ ***
St_stan	β_0	10.2946	0.0534	192.92	$< 2e - 16$ ***
	β_1	-0.0072	0.0026	-2.80	0.0054 **
More	β_0	10.0895	0.0144	702.31	$< 2e - 16$ ***
	β_1	0.3231	0.0330	9.78	$< 2e - 16$ ***
Polj	β_0	10.1369	0.0256	395.77	$< 2e - 16$ ***
	β_1	$1.3811 \cdot 10^{-5}$	$2.1360 \cdot 10^{-5}$	0.65	0.5180
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.12: GLM na pojedinačnim kovarijatama za općine

Iz Tablice 3.12 mogu se iščitati procijenjeni koeficijenti β_1 koji predstavljaju za koliko se μ jedinica promijeni ako se pripadna kovarijata poveća za jednu jedinicu. Ako se poveća varijabla iznosa poreza po stanovniku za jednu jedinicu, μ se poveća za 0.0001 jedinice. To također znači da porez po stanovniku ima pozitivan utjecaj na očekivanje logaritmiranog prosječnog dohotka po stanovniku pa tako i pozitivan utjecaj i na prosječni dohodak po stanovniku zbog svojstava logaritma kao rastuće funkcije. Varijable Obr i More utječu pozitivno na očekivanu vrijednost logaritmiranih vrijednosti prosječnog dohotka po stanovniku te svojim povećanjem za jednu jedinicu uzrokuju povećanje μ za 0.0488 i 0.3231 jedinica, respektivno. Varijable St_nez, ML_stan te St_stan utječu negativno na očekivanu vrijednost logaritmiranih vrijednosti prosječnog dohotka po stanovniku za općine RH. Svojim povećanjem za jednu jedinicu uzrokuju smanjenje varijable odziva za 0.0295, 0.0201 i 0.0072 jedinica, respektivno. Ovi rezultati podupiru početne hipoteze.

Po p-vrijednostima slijedi da su sve kovarijate, osim broja stanovnika te broja kućanstava s pristupom poljoprivrednom zemljištu, značajne u opisu varijable odziva na

razini značajnosti od 5%. Tu tvrdnju potvrđuju i p-vrijednosti u Tablici 3.13.

Početne hipoteze koje su završile nepotvrđene su značajnost utjecaja broja stanovnika i broja kućanstava s pristupom poljoprivrednom zemljištu na varijablu odziva te negativan utjecaj broja kućanstava s pristupom poljoprivrednom zemljištu.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	$Pr(> Chi)$
<i>NULL</i>			427	37.363	
Por	1	6.4896	426	30.873	$< 2.2e - 16$ ***
St_nez	1	16.219	426	21.143	$< 2.2e - 16$ ***
Obr	1	13.856	126	23.507	$< 2.2e - 16$ ***
Br_stan	1	0.0044	426	37.358	0.8219
MI_stan	1	1.5692	426	35.794	$1.55e - 05$ ***
St_stan	1	0.6744	426	36.688	0.0051 **
More	1	6.8539	426	30.509	$< 2.2e - 16$ ***
Polj	1	0.0366	426	37.326	0.5179

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Izvor: Izrada autorice

Tablica 3.13: Analiza varijance za sve kovarijate zasebno za općine

Pri provedbi testa na devijancama opisanog u 2.5 za modele oblika (3.1), dobivene su vrijednosti prikazane u Tablici 3.14.

Ime varijable	Rezultat χ^2 testa
Por	0.0108
St_nez	$5.6396 \cdot 10^{-5}$
Obr	0.0002
Br_stan	0.9436
MI_stan	0.2104
St_stan	0.4113
More	0.0088
Polj	0.8475

Izvor: Izrada autorice

Tablica 3.14: χ^2 test za razliku nul i rezidualne devijance te jednim stupnjem slobode

Zbog p-vrijednosti manjih od 0.05 može se zaključiti da je model koji sadrži jednu od varijabli Por, St_nez, Obr te More značajno bolji od modela koji sadrži samo slobodan član. Ostale kovarijate ne opisuju varijablu odziva značajno bolje od nul modela.

Analiza punog modela

U Tablici 3.15 prikazani su procijenjeni koeficijenti β_j , $j = 0, \dots, 8$ punog generaliziranog linearnog modela oblika (3.2).

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.5132	0.1328	79.20	$< 2e - 16$ * **
Por	$1.2971 \cdot 10^{-5}$	$1.2330 \cdot 10^{-5}$	1.05	0.2934
St_nez	-0.0228	0.0016	-13.92	$< 2e - 16$ * **
Obr	0.0300	0.0038	7.81	$4.50e - 14$ * **
Br_stan	$-1.0322 \cdot 10^{-4}$	$3.2701 \cdot 10^{-5}$	-3.16	0.0017 * *
MI_stan	-0.0087	0.0054	-1.63	0.1037
St_stan	-0.0108	0.0027	-4.01	$7.32e - 05$ * **
More	-0.0084	0.0338	-0.25	0.8042
Polj	$2.7457 \cdot 10^{-4}$	$9.8980 \cdot 10^{-5}$	2.77	0.0058 * *
Signif. codes: 0 * ** 0.001 * * 0.01 * 0.05 . 0.1 1				
Nul devijanca: 37.363 sa 427 stupnjeva slobode				
Rezidualna devijanca: 13.686 sa 419 stupnjeva slobode				
AIC: -238.9				

Izvor: Izrada autorice

Tablica 3.15: GLM na podacima za općine

Značajno i negativno na varijablu odziva za općine RH utječu varijable stope nezaposlenosti, broja stanovnika te postotak stanovnika starijeg od 65 godina. Njihovo povećanje za jednu jedinicu uzrokuje smanjenje μ za 0.0228, $1.0322 \cdot 10^{-4}$ i 0.0108, respektivno. Značajan utjecaj imaju i varijable postotka visoko obrazovanog stanovništva te broja kućanstava s pristupom poljoprivrednom zemljištu čije povećanje za jednu jedinicu uzrokuje povećanje varijable odziva za 0.03 i $2.7457 \cdot 10^{-4}$, respektivno. Neznačajan pozitivan utjecaj ima varijabla poreza po stanovniku, dok neznačajan negativan utjecaj imaju postotak stanovništva mlađeg od 14 godina i varijabla pristupa moru.

Uspoređujući rezultate s početnim hipotezama, potvrdilo se da stopa nezaposlenosti, broj stanovnika i broj kućanstava s pristupom poljoprivrednom zemljištu značajno utječu na varijablu odziva. Također, postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj, a postotci stanovništva mlađih od 14 i starijih od 65 imaju negativan utjecaj na varijablu odziva. Hipoteze koje su završile nepotvrđene su značajan utjecaj prosječnog poreza po stanovniku, pozitivan utjecaj broja stanovnika i pristupa moru na varijablu odziva te negativan utjecaj broja kućanstava s pristupom poljoprivrednom zemljištu.

Također, proveden je test za značajnost modela naspram modela koji sadrži samo slobodan član. P-vrijednost χ^2 statistike od 23.677 sa 8 stupnjeva slobode iznosi 0.0026 što omogućuje zaključak da je puni model bolji na danoj razini značajnosti.

Naredba *anova* daje rezultate u Tablici 3.16 gdje se može vidjeti da su na danoj razini značajnosti β_j koji pripadaju kovarijatama Por, St_nez, Obr, St_stan i Polj različiti od nule. Ostale kovarijate imaju p-vrijednost veću od 0.05.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
<i>NULL</i>			427	37.363	
Por	1	6.4896	426	30.87	< 2.2e - 16 * **
St_nez	1	12.4060	425	18.47	< 2.2e - 16 * **
Obr	1	3.8525	424	14.62	< 2.2e - 16 * **
Br_stan	1	0.1242	423	14.49	0.0512 .
MI_stan	1	0.0021	422	14.49	0.7979
St_stan	1	0.5496	421	13.94	4.093e - 05 * **
More	1	0.0020	420	13.94	0.8069
Polj	1	0.2512	419	13.69	0.0056 * *

Signif. codes: 0 * ** 0.001 * * 0.01 * 0.05 . 0.1 1

Izvor: Izrada autorice

Tablica 3.16: Analiza varijance za puni model za općine

Analiza restringiranog modela

Preko značajnosti procijenjenih koeficijenata i analize vrijednosti dobivenih *anovom* može se zaključiti da bi moglo biti razumno gledati i restringirani model. Varijabla prosječnog poreza po stanovniku izbačena je zbog svoje visoke korelacije s postotkom visoko obrazovanog stanovništva i s varijablom pristupa moru što možemo vidjeti u Tablici 3.5. Također,

po Tablici 3.15 vidi se da nije značajna. Broj stanovnika, iako je značajna varijabla, po naredbi *anova* ne doprinosi značajno smanjenju rezidualne devijance kada se ubaci u model te je jako korelirana s varijablom Polj što može narušiti točnost modela. Varijable postotka stanovništva mlađeg od 14 godina i varijabla pristupa moru imati veliku p-vrijednost u obje gore predstavljene tablice. Izbacuju se i zbog jake koreliranosti između Ml_stan i St_stan te More i Obr.

Gornja rasprava je također potvrđena naredbom *drop1* u R-u koja omogućuje pregled koliko svaka varijabla svojim izbacivanjem iz modela utječe na promjenu AIC vrijednosti. Izabrane varijable za micanje iz modela su one koje su najmanje utjecale na povećanje dane vrijednosti, za koju znamo da je bolje da je što manja.

U Tablici 3.17 možemo vidjeti primjenjenu metodu na restringirani model koji je oblika:

$$\mu = \beta_0 + \beta_1 St_nez + \beta_2 Obr + \beta_3 St_stan + \beta_4 Polj. \quad (3.4)$$

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.2189	0.0433	236.23	$< 2e - 16$ * **
St_nez	-0.0218	0.0015	-14.59	$< 2e - 16$ * **
Obr	0.0375	0.0027	14.01	$< 2e - 16$ * **
St_stan	-0.0044	0.0018	-2.48	0.0135 *
Polj	$-3.9547 \cdot 10^{-5}$	$1.4580 \cdot 10^{-5}$	-2.71	0.0070 **
Signif. codes: 0 * ** 0.001 ** 0.01 * 0.05 . 0.1 1				
Nul devijanca: 37.363 sa 427 stupnjeva slobode				
Rezidualna devijanca: 14.421 sa 423 stupnjeva slobode				
AIC: -224.5				

Izvor: Izrada autorice

Tablica 3.17: GLM na restringiranom modelu za općine

Po p-vrijednostima, koje su sve manje od 5%, vidi se da su sve kovarijate uključene u ovaj model značajne za opis zavisne varijable. U tablici se također mogu vidjeti procijenjeni koeficijenti koji označuju koliko će se promijeniti μ ako se kovarijata poveća za jednu jedinicu. Pozitivno na očekivanu vrijednost prosječnog dohotka po stanovniku utječe postotak visoko obrazovanog stanovništva, dok ostale utječu negativno. Također, iz nul devijance i rezidualne devijance dobije se vrijednost 0.0001 koja je manja od 0.05 te se može zaključiti da je restringirani model značajno bolji od modela koji sadrži samo slobodan

član na danoj razini značajnosti.

Uz provedenu *anova* naredbu koju vidimo u Tablici 3.18 slijedi da kovarijate *St_nez*, *Obr* i *Polj* značajno utječu na smanjenje rezidualne devijance.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
<i>NULL</i>			427	37.363	
<i>St_nez</i>	1	16.2193	426	21.143	< 2.2e - 16 * **
<i>Obr</i>	1	6.3830	425	14.760	< 2.2e - 16 * **
<i>St_stan</i>	1	0.0889	424	14.672	0.1063
<i>Polj</i>	1	0.2508	423	14.421	0.0067 * *
Signif. codes: 0 * ** * 0.001 * * 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.18: Analiza varijance za restringirani model za općine

Ostaje još provjeriti kvalitetu restringiranog modela naspram punog u smislu adekvatnog opisa varijable odziva. Test se provodi putem naredbe *anova* koja uspoređuje dva modela uz nul hipotezu da puni model (Model 0) bolje opisuje podatke od restringiranog (Model 1).

<i>Model 0 : log(Doh) ~</i>					
<i>Por + St_nez + Obr + Br_stan + Ml_stan + St_stan + as.factor(More) + Polj</i>					
<i>Model 1 : log(Doh) ~ St_nez + Obr + St_stan + Polj</i>					
	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
0	419	13.685			
1	423	14.421	-4	-0.7352	0.0002 * **
Signif. codes: 0 * ** * 0.001 * * 0.01 * 0.05 . 0.1 1					

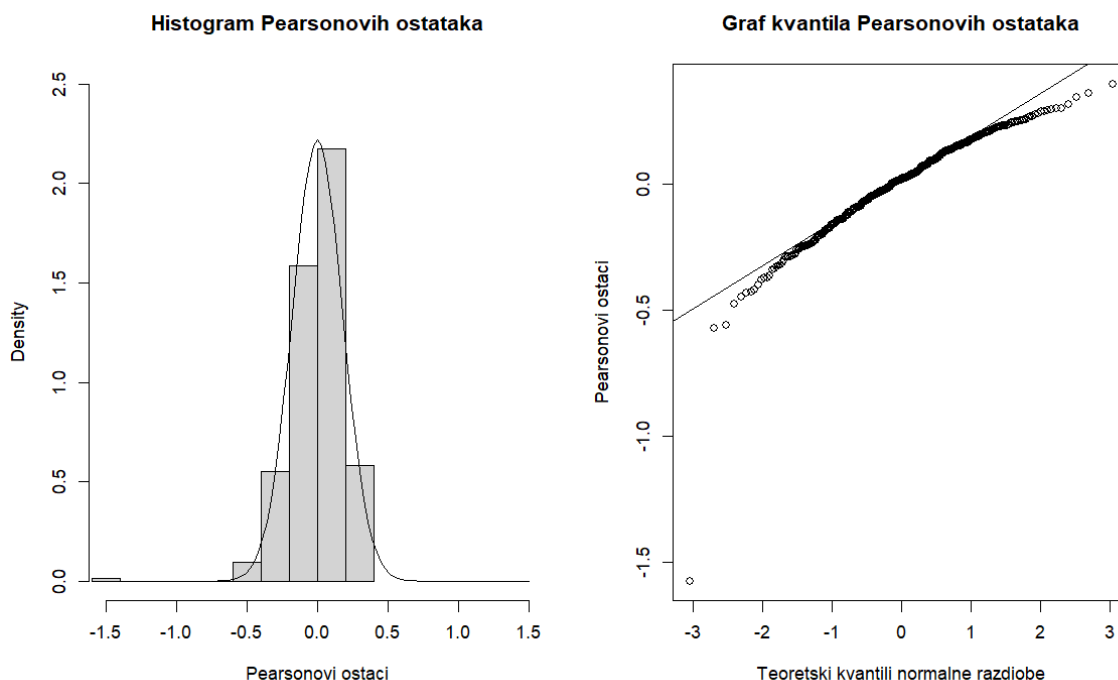
Izvor: Izrada autorice

Tablica 3.19: Analiza devijance

Zbog p-vrijednosti od 0.0002 u Tablici 3.19 može se na razini značajnosti od 5% odbaciti nul hipoteza. Također, AIC restringiranog modela je postigao sličnu vrijednost kao i

AIC punog modela, ali uz smanjenje broja kovarijata. Dakle, razumno je koristiti restringirani model.

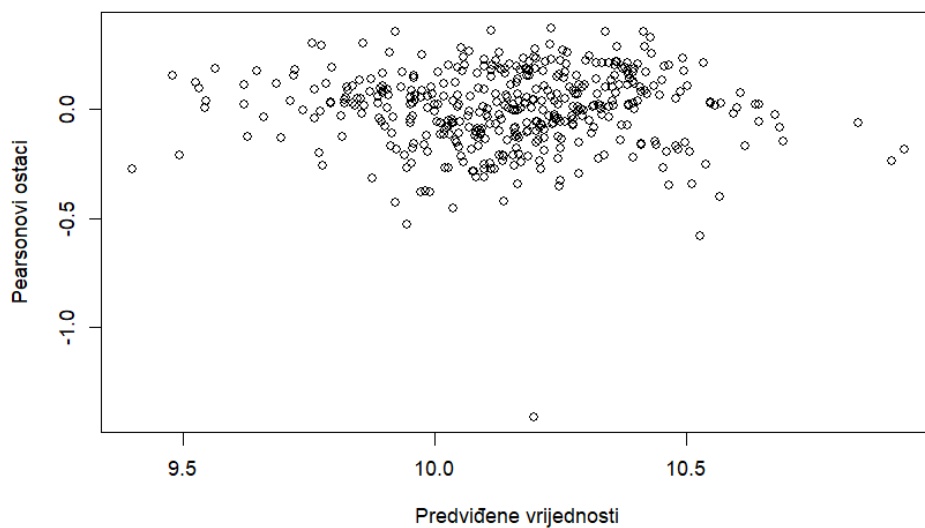
Sljedeće što treba analizirati su ostaci restringiranog modela kako bi se dodatno provjerilo koliko dobro model opisuje logaritmiranu vrijednost prosječnog dohotka po stanovniku. Na Slici 3.1 vide se histogram i graf kvantila Pearsonovih ostataka. Zbog velikih odstupanja vrijednosti od najbliže normalne distribucije, nije moguće zaključiti da ostaci dolaze iz normalne razdiobe. Liliefors test normalnosti daje p-vrijednost od 0.0008 te također odbacuje nul hipotezu da podaci dolaze iz te razdiobe.



Izvor: Izrada autorice

Slika 3.1: Analiza normalnosti Pearsonovih ostataka

Na Slici 3.2 su prikazane predviđene vrijednosti modela naspram Pearsonovih ostataka. Zbog manjka ikakvog trenda u podacima možemo zaključiti da je model adekvatno opisao podatke. Iako ostaci nisu normalno distribuirani, model je značajno bolji od nul modela i punog modela te nema neopisanih trendova u podacima.



Izvor: Izrada autorice

Slika 3.2: Analiza Pearsonovih ostataka

Analiza rezultata za općine

Ime varijable	Zasebni model	Puni model	Restringirani model
Por	✓+	+	x
St_nez	✓-	✓-	✓-
Obr	✓+	✓+	✓+
Br_stan	+	✓-	x
Ml_stan	✓-	-	x
St_stan	✓-	✓-	✓-
More	✓+	-	x
Polj	+	✓+	✓-

✓ statistički značajan utjecaj
 + pozitivan utjecaj - negativan utjecaj
 x ne nalazi se u tom modelu

Izvor: Izrada autorice

Tablica 3.20: Rezultati za općine

U Tablici 3.20 mogu se vidjeti sažeti rezultati za sva tri modela gledana u ovom poglavlju. Prikazane su varijable koje su u danim modelima ispale statistički značajne te kako one utječu - pozitivno ili negativno na varijablu odziva.

Analizom devijance pomoću naredbe *anova* dobilo se da je restringirani model oblika (3.4) bolji od punog modela. Provedbom metode opisane u Poglavlju 2 dobivena je sljedeća analiza hipoteza:

Hipoteza 1: Prosječni porez na dohodak po stanovniku značajno utječe na dohodak.

Varijabla prosječnog poreza na dohodak po stanovniku nije značajno utjecala na varijablu odziva u punom modelu pa, iako je značajno utjecala na smanjenje rezidualne devijance, nije uključena u restringirani model. Također je visoko korelirana s varijablom visoko obrazovanog stanovnika i s varijablom pristupa moru pa bi njeno uključivanje u model zajedno s navedenim varijablama moglo utjecati na točnost modela u opisu podataka.

Hipoteza 2: Stopa nezaposlenosti ima značajan utjecaj na dohodak.

Varijabla stope nezaposlenosti značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je zato ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0218 jedinica.

Hipoteza 3: Postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj na dohodak.

Varijabla postotka visoko obrazovanog stanovništva značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je zato ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje povećanje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0375 jedinica.

Hipoteza 4: Broj stanovnika značajno i pozitivno utječe na rast dohotka.

Varijabla broja stanovnika nije značajno utjecala na smanjenje rezidualne devijance pa, iako je značajno utjecala na varijablu odziva u punom modelu, nije uključena u restringirani model. Također je visoko korelirana s varijablom broja kućanstava s pristupom poljoprivrednom zemljištu pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Hipoteza 5: Postotak stanovnika mladih od 14 godina negativno utječe na dohodak.

Varijabla postotka stanovnika mlađih od 14 godina nije značajno utjecala na smanjenje rezidualne devijance ni na varijablu odziva u punom modelu pa nije uključena u restringirani model. Također je visoko korelirana s varijablom postotka stanovnika starijih od 65 godina pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Hipoteza 6: Postotak stanovnika starijih od 65 godina negativno utječe na dohodak.

Varijabla postotka stanovnika starijih od 65 godina značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu punog modela te je zato ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0044 jedinica.

Hipoteza 7: Pristup moru ima pozitivan utjecaj na dohodak.

Varijabla pristupa moru nije značajno utjecala na smanjenje rezidualne devijance ni na varijablu odziva u punom modelu pa nije uključena u restringirani model. Također je visoko korelirana s varijablom visoko obrazovanog stanovništva pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Hipoteza 8: Broj kućanstava s pristupom poljoprivrednom zemljištu negativno i značajno utječe na dohodak.

Varijabla broja kućanstava s pristupom poljoprivrednom zemljištu značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je zato ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za $3.9547 \cdot 10^{-5}$ jedinica.

Gradovi

Analiza nezavisnih varijabli

Utjecaj pojedinih varijabli na varijablu odziva možemo vidjeti u Tablici 3.21 gdje su prikazani procijenjeni parametri β_0 i β_1 modela oblika (3.1), njihove standardne greške te odgovarajuće p-vrijednosti. Sve vrijednosti dobivene su naredbom *glm* u R-u.

Ime varijable		Procjena koeficijenta	Standardna greška	t vrijednost	$Pr(> t)$
Por	β_0	10.0917	0.03	364.24	$< 2e - 16$ ***
	β_1	$1.0520 \cdot 10^{-4}$	$7.78 \cdot 10^{-6}$	13.53	$< 2e - 16$ ***
St_nez	β_0	10.6671	0.04	253.94	$< 2e - 16$ ***
	β_1	-0.0367	0.01	-7.19	$4.98e - 11$ ***
Obr	β_0	10.0348	0.06	163.47	$< 2e - 16$ ***
	β_1	0.0322	0.01	6.38	$3.04e - 09$ ***
Br_stan	β_0	10.3862	0.03	415.48	$< 2e - 16$ ***
	β_1	$7.1009 \cdot 10^{-7}$	$3.30 \cdot 10^{-7}$	2.15	0.0333 *
Ml_stan	β_0	11.0805	0.16	68.98	$< 2e - 16$ ***
	β_1	-0.0451	0.01	-4.26	$3.94e - 05$ ***
St_stan	β_0	10.3720	0.14	72.30	$< 2e - 16$ ***
	β_1	0.0017	0.01	0.22	0.83
More	β_0	10.3614	0.03	371.92	$< 2e - 16$ ***
	β_1	0.1435	0.05	2.77	0.00646 **
Polj	β_0	10.3870	0.02	417.75	$< 2e - 16$ ***
	β_1	$1.8684 \cdot 10^{-6}$	$8.60 \cdot 10^{-7}$	2.17	0.0317 *
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.21: GLM na pojedinačnim kovarijatama za gradove

Procijenjeni koeficijenti β_1 iz tablice 3.21 predstavljaju koliko se promijeni očekivanje logaritmirane vrijednosti prosječnog dohotka po stanovniku pri povećanju pripadne nezavisne varijable za jednu jedinicu. Ako se poveća varijabla iznosa poreza po stanovniku za jednu jedinicu, μ se poveća za $1.052 \cdot 10^{-4}$ jedinice. To također znači da porez po stanovniku ima pozitivan utjecaj na očekivanje logaritmiranog prosječnog dohotka po stanovniku pa tako i pozitivan utjecaj i na prosječni dohodak po stanovniku zbog svojstava logaritma kao funkcije.

Varijable stope nezaposlenosti i stanovništva mlađeg od 14 godina utječu negativno na prosječni dohodak po stanovniku te ako se povećaju za jednu jedinicu uzrokuju smanjenje očekivanja logaritmiranih vrijednosti prosječnog dohotka po stanovniku, μ , za 0.0367 i 0.0451, respektivno. Povećanje varijable visoko obrazovanih stanovnika, broja stanovnika i varijable pristupa moru za jednu jedinicu povećava μ za 0.0322, $7.1009 \cdot 10^{-7}$ i 0.1435, respektivno. Ti zaključci odgovaraju početnim hipotezama.

Pozitivan utjecaj kovarijate postotka stanovništva starijeg od 65 godina i broja kućanstava s pristupom poljoprivrednom zemljištu odstupa od pretpostavki.

Po p-vrijednostima za β_1 koeficijente slijedi da su sve kovarijate, osim postotka stanovništva starijeg od 65 godina, značajne u opisu varijable odziva na razini značajnosti od 5%. Tu tvrdnju potvrđuju i p-vrijednosti dobivene naredbom *anova* u Tablici 3.22.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
<i>NULL</i>			127	9.4410	
Por	1	5.5916	126	3.8493	$< 2.2e - 16$ ***
St_nez	1	2.7483	126	6.6926	$6.33e - 13$ ***
Obr	1	2.3065	126	7.1344	$1.74e - 10$ ***
Br_stan	1	0.3347	126	9.1062	0.0314 *
ML_stan	1	1.1892	126	8.2517	$2.03e - 05$ ***
St_stan	1	0.0036	126	9.4374	0.8275
More	1	0.5418	126	8.8992	0.0056 **
Polj	1	0.3408	126	9.1001	0.0298 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Izvor: Izrada autorice

Tablica 3.22: Analiza varijance za sve kovarijate zasebno za gradove

Pri provedbi testa nad devijancama opisanog u Potpoglavlju 2.5 za modele oblika (3.1), dobivene su vrijednosti prikazane u Tablici 3.23. Visoke p-vrijednosti govore da je model koji se sastoji samo od slobodnog člana bolji od modela koji sadrži slobodan član i danu kovarijatu. Za model koji sadrži varijablu prosječnog poreza po stanovniku se, na razini značajnosti od 5%, može zaključiti da je bolji od modela koji sadrži samo slobodan član. Za ostale, iako je većina kovarijata značajna u opisu varijable odziva, nije dovoljno imati samo jednu od kovarijata u modelu kako bi se značajno bolje opisala varijabla odziva.

Ime varijable	Rezultat χ^2 testa
Por	0.0180
St_nez	0.0974
Obr	0.1288
Br_stan	0.5628
MI_stan	0.2755
St_stan	0.9522
More	0.4617
Polj	0.5593

Izvor: Izrada autorice

Tablica 3.23: χ^2 test za razliku nul i rezidualne devijance te jednim stupnjem slobode**Analiza punog modela**

U Tablici 3.24 prikazani su procijenjeni koeficijenti β_j , $j = 0, \dots, 8$ punog generaliziranog linearnog modela oblika (3.2).

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.3971	0.1992	52.19	$< 2e - 16$ * **
Por	$9.4470 \cdot 10^{-5}$	$5.1140 \cdot 10^{-6}$	18.47	$< 2e - 16$ * **
St_nez	-0.0187	0.0028	-6.661	$8.89e - 10$ * **
Obr	0.0243	0.0034	7.09	$1.06e - 10$ * **
Br_stan	$7.5084 \cdot 10^{-6}$	$5.1400 \cdot 10^{-6}$	1.46	0.1467
MI_stan	-0.0168	0.0078	-2.14	0.0348 *
St_stan	-0.0078	0.0048	-1.62	0.1091
More	-0.0889	0.0260	-3.43	0.0008 * **
Polj	$-2.0566 \cdot 10^{-5}$	$1.3370 \cdot 10^{-5}$	-1.54	0.1267
Signif. codes: 0 * ** 0.001 ** 0.01 * 0.05 . 0.1 1				
Nul devijanca: 9.4410 sa 127 stupnjeva slobode				
Rezidualna devijanca: 1.3994 sa 119 stupnjeva slobode				
AIC: -194.8				

Izvor: Izrada autorice

Tablica 3.24: GLM na podacima za gradove

Procijenjeni parametar za kovarijatu prosječnog poreza po stanovniku je $9.4470 \cdot 10^{-5}$, što znači da ako se porez po stanovniku poveća za jednu jedinicu, očekivanje μ povećat će se za $9.4470 \cdot 10^{-5}$ jedinica. Zbog pozitivnog utjecaja na očekivanje logaritma prosječnog dohotka po stanovniku, očitava se i pozitivan utjecaj na prosječni dohodak po stanovniku. Ta veza bi se mogla objasniti činjenicom da povećanje dohotka uzrokuje povećanje plaćenog poreza. Iako je parametar male vrijednosti, zbog male p-vrijednosti vidimo da je kovarijata značajna u opisu varijable odziva.

Analognim analiziranjem, stopa nezaposlenosti, postotak stanovništva mlađeg od 14 godina te varijabla pristupa moru značajne na razini značajnosti od 5% su varijable te njihovo povećanje za jednu jedinicu uzrokuje smanjenje očekivanja μ za 0.0187, 0.0168 i 0.0889 jedinica, respektivno. Ove kovarijate negativno utječu na očekivanje prosječnog dohotka po stanovniku. Odstupanje od početnih pretpostavki očituje se u negativnom utjecaju pristupa moru na dohodak. S druge strane, značajno i pozitivno na dohodak utječe kovarijata obrazovanja, čije povećanje za jednu jedinicu uzrokuje povećanje μ za 0.0243 jedinica.

P-vrijednost veću od 0.05 imaju varijable broja stanovnika, postotka stanovništva starijih od 65 godina te broja kućanstava koja imaju pristup poljoprivrednom zemljištu. Budući da na to da su vrijednosti veće od razine značajnosti od 5% to znači da one nisu značajne u modelu pri opisu varijable odziva.

Uspoređujući rezultate s početnim hipotezama, potvrdilo se da porez na dohodak po stanovniku i stopa nezaposlenosti značajno utječu na varijablu odziva. Postotak stanovništva sa završenim visokim obrazovanjem i broj stanovnika imaju pozitivan utjecaj, dok broj kućanstava s pristupom poljoprivrednom zemljištu te postotak stanovništva mlađih od 14 i starijih od 65 godina ima negativan utjecaj na prosječni dohodak. Hipoteze koje su završile nepotvrđene su: pozitivan utjecaj mora na dohodak te značajan utjecaj broja stanovnika i broja kućanstava s pristupom poljoprivrednom zemljištu.

Također, proveden je test značajnosti modela naspram modela koji sadrži samo slobodan član opisanog u Potpoglavlju 2.5. χ^2 testom sa 8 stupnjeva slobode za 8.0416 dobije se vrijednost 0.4294 pa slijedi da puni model nije značajno bolji od nul modela na danoj razini značajnosti.

Rezultati naredbe *anova* u Tablici 3.25 daju da su kovarijate *Por*, *St_nez*, *Obr*, *St_stan* te *More* značajno smanjuju rezidualnu devijancu u modelu na razini značajnosti od 5%, tj. može se zaključiti da su pripadajući β_j različiti od nule. Ostale kovarijate imaju p-vrijednost veću od 0.05.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
<i>NULL</i>			127	9.4410	
Por	1	5.5916	126	3.8493	< 2.2e - 16 ***
St_nez	1	1.3947	125	2.4546	< 2.2e - 16 ***
Obr	1	0.7688	124	1.6858	6.185e - 16 ***
Br_stan	1	0.0396	123	1.6462	0.0664 .
MI_stan	1	0.0007	122	1.6454	0.8055
St_stan	1	0.0646	121	1.5808	0.01909 *
More	1	0.1536	120	1.4272	0.0003 ***
Polj	1	0.0278	119	1.3994	0.1240

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Izvor: Izrada autorice

Tablica 3.25: Analiza varijance za puni model za gradove

Analiza restringiranog modela

U slučaju modela s mnogo varijabli poželjno je provjeriti postoji li model s manje varijabli koji podjednako dobro opisuje dan podatke. Preko značajnosti procijenjenih koeficijenata i analize vrijednosti dobivenih *anovom* može se zaključiti da bi moglo biti razumno gledati i restringirani model koji bi sadržavao prve tri kovarijate te varijablu pristupa moru jer sve značajno utječu na varijablu odziva. Te četiri varijable imaju najniže p-vrijednosti i za procijenjene koeficijente i one dobivene *anovom*. Također, više nema varijabli između kojih je najveća korelacija kao prikazano u 3.7. Restringirani model je, dakle, oblika:

$$\mu = \beta_0 + \beta_1 Por + \beta_2 St_nez + \beta_3 Obr + \beta_4 More. \quad (3.5)$$

U Tablici 3.26 možemo vidjeti primijenjenu metodu na restringirani model.

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.0423	0.0439	228.99	$< 2e - 16$ ***
Por	$9.2083 \cdot 10^{-5}$	$5.1830 \cdot 10^{-6}$	17.77	$< 2e - 16$ ***
St_nez	-0.0215	0.0027	-8.02	$7.17e - 13$ ***
Obr	0.0233	0.0028	8.37	$1.08e - 13$ ***
More	-0.0785	0.0252	-3.11	0.0023 **
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1				
Nul devijanca: 9.4410 sa 127 stupnjeva slobode				
Rezidualna devijanca: 1.5627 sa 123 stupnjeva slobode				
AIC: -188.67				

Izvor: Izrada autorice

Tablica 3.26: GLM na restringiranom modelu za gradove

Po p-vrijednostima, koje su sve manje od 5%, vidi se da su sve kovarijate uključene u ovaj model značajne za opis zavisne varijable. U tablici se također mogu vidjeti procijenjeni koeficijenti koji označuju koliko će se promijeniti μ ako se kovarijata poveća za jednu jedinicu. Pozitivno na očekivanu vrijednost prosječnog dohotka po stanovniku utječu prosječni porez na dohodak po stanovniku te postotak visoko obrazovanog stanovništva, dok stopa nezaposlenosti i varijabla pristupa moru utječu negativno.

Analiza nul devijance i rezidualne devijance daje vrijednost od 0.0961. Iako na danoj razini značajnosti ne možemo odbaciti nul hipotezu da je nul model bolji u opisu varijable odziva, vrijednost je mnogo manja nego u punom modelu.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	$Pr(> Chi)$
<i>NULL</i>			127	9.4410	
Por	1	5.5916	126	3.8493	$< 2.2e - 16$ ***
St_nez	1	1.3947	125	2.4546	$< 2.2e - 16$ ***
Obr	1	0.7688	124	1.6858	$7.31e - 15$ ***
More	1	0.1230	123	1.5627	0.0019 **
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.27: Analiza varijance za restringirani model za gradove

Uz provedenu *anova* naredbu, koju vidimo u Tablici 3.27, slijedi da su sve nezavisne varijable značajne na razini značajnosti od 5%.

Ostaje još provjeriti kvalitetu restringiranog modela naspram punog u smislu adekvatnog opisa varijable odziva. Test se provodi putem naredbe *anova* koja uspoređuje dva modela uz nul hipotezu da puni model (Model 0) bolje opisuje podatke od restringiranog (Model 1)

<i>Model 0 : log(Doh) ~</i>					
<i>Por + St_nez + Obr + Br_stan + Ml_stan + St_stan + as.factor(More) + Polj</i>					
<i>Model 1 : log(Doh) ~ Por + St_nez + Obr + as.factor(More)</i>					
	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
0	119	1.3994			
1	123	1.5627	-4	-0.1633	0.0077 **
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

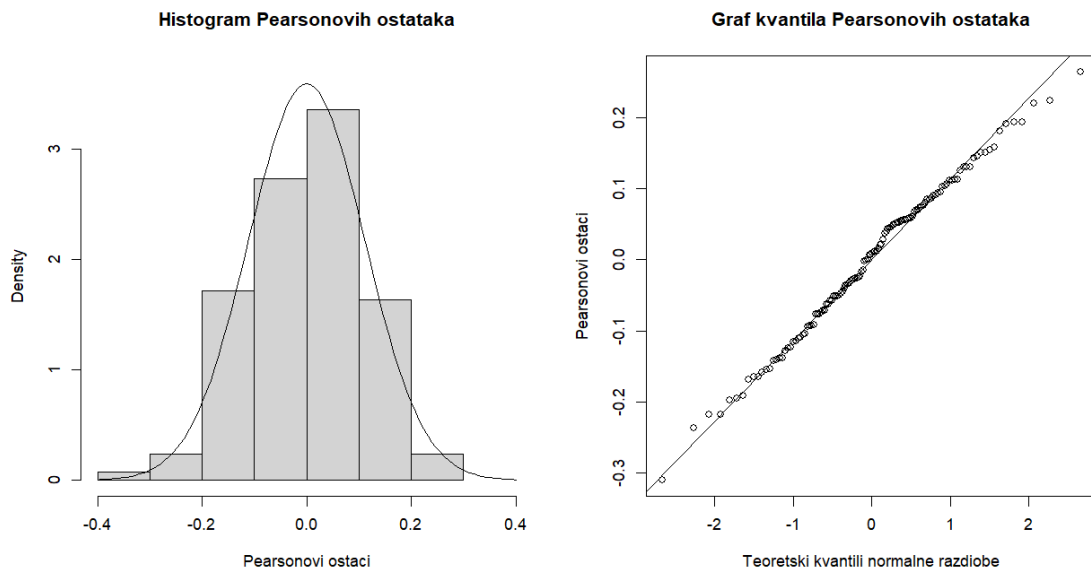
Izvor: Izrada autorice

Tablica 3.28: Analiza devijance

U Tablici 3.28 vidi se da je moguće odbaciti nul hipotezu na razini značajnosti od 5%. Ako se još uspoređi i AIC oba modela, dobije se da je AIC restringiranog modela tek malo veći, ali uz značajno smanjenje broja kovarijata. Dakle, razumno je koristiti restringirani model u daljnjoj analizi.

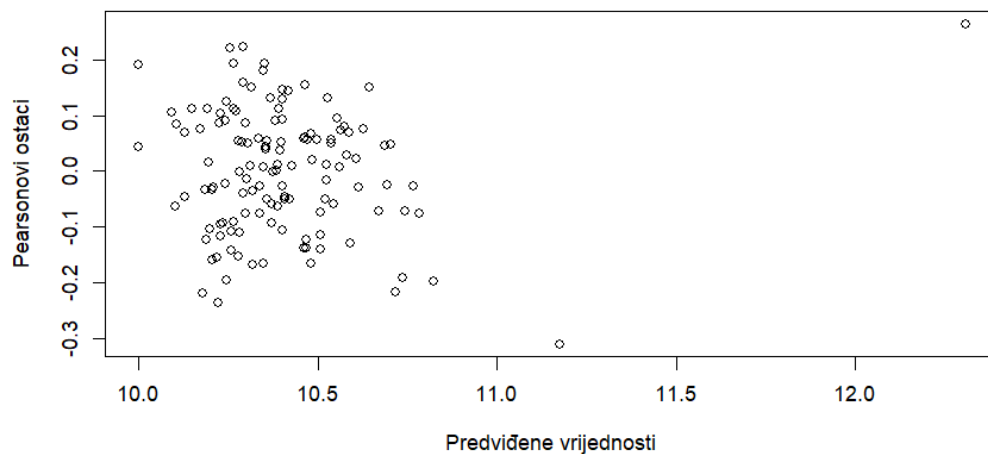
Sljedeće što treba analizirati su ostaci restringiranog modela kako bi se dodatno provjerilo koliko dobro model opisuje logaritmiranu vrijednost prosječnog dohotka po stanovniku. Na Slici 3.3 vide se histogram i graf kvantila Pearsonovih ostataka. Na oba možemo vidjeti da je razumno pretpostaviti da dolaze iz normalne razdiobe, što također potvrđuje i Liliefors test normalnosti za koji se uz p-vrijednost od 0.0698 ne može odbaciti nul hipoteza da podaci dolaze iz te razdiobe, što znači da model adekvatno opisuje podatke.

Na Slici 3.4 prikazane su predviđene vrijednosti modela naspram Pearsonovih ostataka. Na grafu nema uzorka ni trenda među podacima, dakle može se zaključiti da ne postoje neki dodatni neobjašnjeni trendovi u podacima varijable doziva.



Izvor: Izrada autorice

Slika 3.3: Analiza normalnosti Pearsonovih ostataka



Izvor: Izrada autorice

Slika 3.4: Analiza Pearsonovih ostataka

Analiza rezultata za gradove

U Tablici 3.29 mogu se vidjeti sažeti rezultati za sva tri modela gledana u ovom poglavlju. Prikazane su varijable koje su u danim modelima ispale statistički značajne te kako one utječu - pozitivno ili negativno na varijablu odziva.

Ime varijable	Zasebni model	Puni model	Restringirani model
Por	✓+	✓+	✓+
St_nez	✓-	✓-	✓-
Obr	✓+	✓+	✓+
Br_stan	✓+	+	x
MI_stan	✓-	✓-	x
St_stan	+	-	x
More	✓+	✓-	✓-
Polj	+	-	x

✓ statistički značajan utjecaj
 + pozitivan utjecaj - negativan utjecaj
 x ne nalazi se u tom modelu

Izvor: Izrada autorice

Tablica 3.29: Rezultati za gradove

Analizom devijance pomoću naredbe *anova* dobilo se da je restringirani model (3.5) bolji od punog modela. Metodom iz Poglavlja 2 dobivena je sljedeća analiza hipoteza:

Hipoteza 1: Prosječni porez na dohodak po stanovniku značajno utječe na dohodak.

Prosječni porez na dohodak po stanovniku značajan je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je ostala u restringiranom modelu. Povećanje prosječnog poreza na dohodak po stanovniku za jednu jedinicu uzrokuje povećanje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za $9.2083 \cdot 10^{-5}$ jedinica.

Hipoteza 2: Stopa nezaposlenosti ima značajan utjecaj na dohodak.

Stopa nezaposlenosti značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0215 jedinica.

Hipoteza 3: Postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj na dohodak.

Varijabla postotka visoko obrazovanih stanovništva značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje povećanje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0233 jedinica.

Hipoteza 4: Broj stanovnika značajno i pozitivno utječe na rast dohotka.

Varijabla broja stanovnika nije bila značajna u punom modelu za opis varijable odziva i nije značajno smanjila rezidualnu devijancu modela te zato nije zadržana u restringiranom modelu.

Hipoteza 5: Postotak stanovnika mlađih od 14 godina negativno utječe na dohodak.

Varijabla postotka stanovnika mlađih od 14 godina bila je značajna u punom modelu za opis varijable odziva, ali nije značajno smanjila rezidualnu devijancu modela te zato nije zadržana u restringiranom modelu.

Hipoteza 6: Postotak stanovnika starijih od 65 godina negativno utječe na dohodak.

Varijabla postotka stanovnika starijih od 65 godina nije bila značajna u punom modelu za opis varijable odziva. Iako je značajno smanjila rezidualnu devijancu, nije značajno smanjivala ni *AIC* vrijednost pa nije zadržana u restringiranom modelu.

Hipoteza 7: Pristup moru ima pozitivan utjecaj na dohodak.

Varijabla pristupa moru značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu modela te je ostala u restringiranom modelu. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0785 jedinica.

Hipoteza 8: Broj kućanstava s pristupom poljoprivrednom zemljištu negativno i značajno utječe na dohodak.

Varijabla broja kućanstava s pristupom poljoprivrednom zemljištu nije bila značajna u punom modelu za opis varijable odziva i nije značajno smanjila rezidualnu devijancu modela te zato nije zadržana u restringiranom modelu.

Županije

Analiza nezavisnih varijabli

Utjecaj pojedinih varijabli na varijablu odziva možemo vidjeti u Tablici 3.30 gdje su prikazani procijenjeni parametri β_0 i β_1 modela oblika (3.1), njihove standardne greške te odgovarajuće p-vrijednosti. Sve vrijednosti dobivene su naredbom *glm* u R-u.

Ime varijable		Procijenjen koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Por	β_0	10.2810	0.0386	266.23	$< 2e - 16$ ***
	β_1	$9.0440 \cdot 10^{-5}$	$2.3810 \cdot 10^{-5}$	3.80	0.0012 **
St.nez	β_0	10.6152	0.0818	129.77	$< 2e - 16$ ***
	β_1	-0.0354	0.0103	-3.44	0.0027 **
Obr	β_0	9.9224	0.0632	157.09	$< 2e - 16$ ***
	β_1	0.0389	0.0053	7.38	$5.44e - 07$ ***
Br.stan	β_0	10.1951	0.0507	200.96	$< 2e - 16$ ***
	β_1	$8.0455 \cdot 10^{-7}$	$1.9510 \cdot 10^{-7}$	4.13	0.0006 ***
Ml.stan	β_0	11.3994	0.4552	25.04	$5.16e - 16$ ***
	β_1	-0.0680	0.0296	-2.29	0.0334 *
St.stan	β_0	10.4544	0.3856	27.11	$< 2e - 16$ ***
	β_1	-0.0052	0.0210	-0.25	0.8070
More	β_0	10.3151	0.0501	206.03	$< 2e - 16$ ***
	β_1	0.1325	0.0867	1.53	0.1430
Polj	β_0	10.2035	0.0478	213.58	$< 2e - 16$ ***
	β_1	$2.1529 \cdot 10^{-6}$	$5.0350 \cdot 10^{-7}$	4.28	0.0004 ***
Gus	β_0	10.2996	0.0379	271.81	$< 2e - 16$ ***
	β_1	$4.7241 \cdot 10^{-4}$	$1.3540 \cdot 10^{-4}$	3.49	0.0025 **
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.30: GLM na pojedinačnim kovarijatama za županije

Iz Tablice 3.30 mogu se iščitati vrijednosti β_1 koje označuju za koliko se promijeni očekivanje logaritmirane vrijednosti prosječnog dohotka po stanovniku pri povećanju pripadne nezavisne varijable za jednu jedinicu. Prosječan porez po stanovniku, postotak vi-

soko obrazovanog stanovništva, broja stanovnika i broj kućanstava s pristupom poljoprivrednom zemljištu značajno i pozitivno utječu na očekivani logaritmirani prosječni dohodak po stanovniku, μ , te njihovo povećanje za jednu jedinicu uzrokuje povećanje μ za $9.0440 \cdot 10^{-5}$, 0.0389, 0.1325 i $2.1529 \cdot 10^{-6}$ jedinica, respektivno. Značajno i negativno na varijablu odziva utječu stopa nezaposlenosti i postotak stanovništva mlađi od 14 godina čije povećanje za jednu jedinicu uzrokuje smanjenje μ za 0.0354 i 0.068 jedinica, respektivno. Varijabla dostupna samo za županije, gustoća naseljenosti u županijama, pozitivno i značajno utječe na varijablu odziva te njeno povećanje za jednu jedinicu uzrokuje povećanje μ za $4.7241 \cdot 10^{-4}$.

Po p-vrijednostima slijedi da su sve kovarijate, osim postotka stanovništva starijeg od 65 godina i varijable pristupa moru, značajne u opisu varijable odziva na razini značajnosti od 5%. Tu tvrdnju potvrđuju i p-vrijednosti u Tablici 3.31.

Za modele (3.1) za županije potvrđene su hipoteze koje pretpostavljaju da će prosječni porez, stopa nezaposlenosti, broj stanovnika, broj kućanstava s pristupom poljoprivrednom zemljištu i gustoća stanovnika značajno utjecati na prosječni dohodak, a završeno obrazovanje, broj stanovnika, varijabla pristupa moru i gustoća stanovnika imati pozitivan utjecaj na varijablu odziva te broj stanovnika mlađih od 14 i starijih od 65 godina biti negativnog utjecaja. Jedina hipoteza koja nije potvrđena je negativan utjecaj broja kućanstava s pristupom poljoprivrednom zemljištu na varijablu odziva.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
<i>NULL</i>			20	0.7486	
Por	1	0.3231	19	0.4255	0.0001 * **
St_nez	1	0.2873	19	0.4613	0.0006 * **
Obr	1	0.5550	19	0.1936	$1.59e - 13$ * **
Br_stan	1	0.3537	19	0.3950	$3.71e - 05$ * **
MI_stan	1	0.1623	19	0.5863	0.0218 *
St_stan	1	0.0024	19	0.7462	0.8039
More	1	0.0819	19	0.6667	0.1266
Polj	1	0.3671	19	0.3815	$1.90e - 05$ * **
Gus	1	0.2924	19	0.4562	0.0005 * **
Signif. codes: 0 * ** * 0.001 * * 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.31: Analiza varijance za sve kovarijate zasebno za županije

Pri provedbi testa na devijancama opisanog u 2.5 za modele oblika (3.1), dobivene su vrijednosti prikazane u Tablici 3.23. Visoke p-vrijednosti govore da je model koji se sastoji samo od slobodnog člana bolji od modela koji sadrži i danu kovarijatu. Nijedna od kovarijata predstavljena u ovom radu nije sama dovoljna kako bi značajno bolje opisala varijablu odziva nego slobodni član.

Ime varijable	Rezultat χ^2 testa
Por	0.5697
St_nez	0.5919
Obr	0.5919
Br_stan	0.5520
MI_stan	0.6871
St_stan	0.9608
More	0.7747
Polj	0.5446
Gus	0.5887

Izvor: Izrada autorice

Tablica 3.32: χ^2 test za razliku nul i rezidualne devijance te jednim stupnjem slobode

Analiza punog modela

U Tablici 3.33 prikazani su procijenjeni koeficijenti β_j , $j = 0, \dots, 9$ punog generaliziranog linearnog modela.

Analizom p-vrijednosti vidi se da varijable Obr, Br_stan, MI_stan, St_stan te Polj značajno utječu na varijablu odziva. Od njih postotak stanovništva sa završenim visokim obrazovanjem i broj stanovnika utječu pozitivno na μ , dok ostale negativno. Povećanje varijabli visoko obrazovanog stanovništva i broja stanovnika za jednu jedinicu uzrokuje povećanje μ za 0.0301 i $6.9609 \cdot 10^{-6}$ jedinica, respektivno. S druge strane, povećanje varijabli postotka stanovništva mlađeg od 14 te onog starijeg od 65 godina i varijable broja kućanstava s pristupom poljoprivrednom zemljištu za jednu jedinicu uzrokuje smanjenje μ za 0.0941, 0.0285 i $-2.0238 \cdot 10^{-5}$ jedinica, respektivno.

Gledajući početne hipoteze, potvrđene su završile pretpostavke da će postotak stanovništva sa završenim visokim obrazovanjem, broj stanovnika i broj kućanstava s pristupom poljoprivrednom zemljištu utjecati značajno na varijablu odziva, da će broj stanovnika i varijabla pristupa moru utjecati pozitivno na nju, a stanovništvo mlađe od 14 i

starije od 65 te broj kućanstava s pristupom poljoprivrednom zemljištu negativno. S druge strane, porez po stanovniku, stopa nezaposlenosti ni gustoća stanovnika u županiji nisu značajno utjecali na dohodak po stanovniku, a gustoća stanovništva negativno utječe na μ .

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	11.9628	0.5128	23.33	$1.02e - 10$ * **
Por	$2.1851 \cdot 10^{-4}$	$1.0620 \cdot 10^{-4}$	2.06	0.06418 .
St_nez	-0.0072	0.0059	-1.21	0.2529
Obr	0.0301	0.0109	2.77	0.0182 *
Br_stan	$6.9609 \cdot 10^{-6}$	$1.6530 \cdot 10^{-6}$	4.21	0.0015 * *
MI_stan	-0.0941	0.0215	-4.38	0.0011 * *
St_stan	-0.0285	0.0128	-2.23	0.0472 *
More	0.0268	0.0523	0.51	0.6183
Polj	$-2.0238 \cdot 10^{-5}$	$5.0201 \cdot 10^{-6}$	-4.03	0.0020 * *
Gus	$-6.6511 \cdot 10^{-4}$	$5.5270 \cdot 10^{-4}$	-1.20	0.2541
Signif. codes: 0 * * * 0.001 * * 0.01 * 0.05 . 0.1 1				
Nul devijanca: 0.748630 sa 20 stupnjeva slobode				
Rezidualna devijanca: 0.028851 sa 11 stupnjeva slobode				
AIC: -56.798				

Izvor: Izrada autorice

Tablica 3.33: GLM na podacima za županije

Također, proveden je test za značajnost modela naspram modela koji sadrži samo slobodan član. P-vrijednost χ^2 statistike od 0.7198 sa 8 stupnjeva slobode daje p-vrijednost od 0.9995 pa slijedi zaključak da puni model nije značajno bolji. To može slijediti iz jake koreliranosti kovarijata pa je zbog toga bitno pogledati restringirane modele.

Naredba *anova* daje da su kovarijate Por, St_nez, Obr, MI_stan te Polj značajne za smanjivanje devijance kad se dodaju u model. Ostale kovarijate imaju p-vrijednost veću od 0.05.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	$Pr(> Chi)$
<i>NULL</i>			20	0.7486	
Por	1	0.3231	19	0.4255	$< 2.2e - 16$ * **
St_nez	1	0.1594	18	0.2661	$6.44e - 15$ * **
Obr	1	0.1639	17	0.1022	$2.68e - 15$ * **
Br_stan	1	0.0049	16	0.0973	0.1710
ML_stan	1	0.0210	15	0.0764	0.0047 * *
St_stan	1	0.0010	14	0.0754	0.5316
More	1	0.0005	13	0.0749	0.6669
Polj	1	0.0422	12	0.0327	$6.02e - 05$ * **
Gus	1	0.0038	11	0.0289	0.2288
Signif. codes: 0 * ** * 0.001 * * 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.34: Analiza varijance za puni model za županije

Analiza restringiranog modela

Zbog jake korelacije među kovarijata te činjenice da nul model bolje opisuje podatke od punog modela, potrebno je pogledati mogućnosti za restringirani model. Uz pomoć gornjih tablica te naredbe *drop1*, koja govori koja od varijabli kad izbačena najmanje utječe na *AIC* vrijednost, iz modela su maknute varijable Por, Br_stan, St_stan, More i Gus. Varijabla prosječnog poreza po stanovniku ne utječe značajno na varijablu odziva, varijabla broja stanovnika ne utječe značajno na smanjenje rezidualne devijance modela, a obje su visoko korelirane s varijablom Obr koja je i značajna u opisu varijable odziva i po *anovi*. Zato je varijabla postotka visoko obrazovanog stanovništva ostala u modelu dok su navedene dvije izbačene. Varijabla postotka stanovništva starijeg od 65 godina ne utječe značajno na smanjenje rezidualne devijance niti njeno isključenje značajno povećava *AIC* vrijednost. Varijable More i Gus nisu značajne u opisu varijable odziva niti su značajne po *anovi*.

Restringirani model 1 je oblika:

$$\mu = \beta_0 + \beta_1 St_nez + \beta_2 Obr + \beta_3 ML_stan + \beta_4 Polj. \quad (3.6)$$

U Tablici 3.35 možemo vidjeti primijenjenu metodu na restringirani model.

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.6217	0.2538	41.844	$< 2e - 16$ ***
St_nez	-0.0221	0.0052	-4.26	0.0006 ***
Obr	0.0223	0.0083	2.68	0.0164 *
MI_stan	-0.0260	0.0138	-1.88	0.0779 .
Polj	$6.2876 \cdot 10^{-7}$	$5.1150 \cdot 10^{-7}$	1.23	0.2367
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1				
Nul devijanca: 0.74863 sa 20 stupnjeva slobode				
Rezidualna devijanca: 0.081933 sa 16 stupnjeva slobode				
AIC: -44.879				

Izvor: Izrada autorice

Tablica 3.35: GLM na restringiranom modelu 1 za županije

U tablici se također mogu vidjeti procijenjeni koeficijenti koji označuju koliko će se promijeniti μ ako se kovarijata poveća za jednu jedinicu. Pozitivno i značajno na očekivanu vrijednost prosječnog dohotka po stanovniku utječe postotak visoko obrazovanog stanovništva. Njeno povećanje za jednu jedinicu uzrokuje povećanje μ za 0.0223 jedinica. Stopa nezaposlenosti utječe negativno i značajno te njeno povećanje za jednu jedinicu smanji μ za 0.0221 jedinice. Iz nul devijance i rezidualne devijance χ^2 testom dobije se vrijednost 0.9554 koja govori da restringirani model nije značajno bolji od modela koji sadrži samo slobodan član.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	$Pr(> Chi)$
<i>NULL</i>			20	0.7486	
St_nez	1	0.2873	19	0.4613	$6.84e - 14$ ***
Obr	1	0.3590	18	0.1023	$< 2.2e - 16$ ***
MI_stan	1	0.0126	17	0.0897	0.1166
Polj	1	0.0077	16	0.0819	0.2190
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.36: Analiza varijance za restringirani model 1 za županije

U Tablici 3.36 vidi se provedena *anova* naredbu iz koje se vidi da su kovarijate St_nez

i Obr značajne na razini značajnosti od 5%.

Usporedbu adekvatnosti opisa podataka punog i restringiranog modela 1 ispituje se pomoću naredbe *anova*. Nul hipoteza je da puni model (Model 0) bolje opisuje podatke od restringiranog (Model 1).

<i>Model0 : log(Doh) ~</i>					
<i>Por + St_nez + Obr + Br_stan + ML_stan + St_stan + as.factor(More) + Polj</i>					
<i>Model1 : log(Doh) ~ St_nez + Obr + ML_stan + Polj</i>					
	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
0	11	0.0289			
1	16	0.0819	-5	-0.0531	0.0011 **
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.37: Analiza devijance

Odbacuje se nul hipoteza zbog p-vrijednosti manje od 0.05 te se može zaključiti da je restringirani model 1 bolji u opisu podataka od punog modela. No i dalje model nije značajno bolji od nul modela te uključuje varijable koje ne utječu značajno na varijablu odziva.

Zbog neznačajnog utjecaja na varijablu odziva te na smanjenje rezidualne devijance u rezidualnom modelu 1 ima smisla gledati novi smanjeni model bez varijabli ML_stan i Polj. S obzirom na to da je podataka najmanje za županije te postoje visoke korelacije među kovarijatima, ima smisla da restringirani model ima malo kovarijata za opis podataka.

Restringirani model 2 je oblika:

$$\mu = \beta_0 + \beta_1 St_nez + \beta_2 Obr. \quad (3.7)$$

U Tablici 3.38 možemo vidjeti primijenjenu metodu na restringirani model 2.

	Procijenjeni koeficijent	Standardna greška	t vrijednost	$Pr(> t)$
Slobodni član	10.1393	0.0718	141.28	$< 2e - 16$ ***
St_nez	-0.0212	0.0053	-4.01	0.0008 ***
Obr	0.0333	0.0042	7.95	$2.69e - 07$ ***
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1				
Nul devijanca: 0.74863 sa 20 stupnjeva slobode				
Rezidualna devijanca: 0.10228 sa 18 stupnjeva slobode				
AIC: -44.221				

Izvor: Izrada autorice

Tablica 3.38: GLM na restringiranom modelu 2 za županije

Po p-vrijednostima, koje su sve manje od 5%, vidi se da su sve kovarijate uključene u ovaj model značajne za opis zavisne varijable. U tablici se također mogu vidjeti procijenjeni koeficijenti koji označuju koliko će se promijeniti μ ako se kovarijata poveća za jednu jedinicu. Pozitivno na očekivanu vrijednost prosječnog dohotka po stanovniku utječe postotak visoko obrazovanog stanovništva. Njeno povećanje za jednu jedinicu uzrokuje povećanje μ za 0.0333 jedinica. Stopa nezaposlenosti utječe negativno i njeno povećanje za jednu jedinicu smanji μ za 0.0212 jedinice. Iz nul devijance i rezidualne devijance χ^2 testom se dobije vrijednost 0.7238 koja govori da restringirani model nije značajno bolji od modela koji sadrži samo slobodan član.

U Tablici 3.39 vidi se provedena *anova* naredbu iz koje slijedi da su sve kovarijate značajni na razini značajnosti od 5%.

Ime varijable	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	$Pr(> Chi)$
<i>NULL</i>			427	37.363	
St_nez	1	16.2193	426	21.143	$< 2.2e - 16$ ***
Obr	1	6.3830	425	14.760	$< 2.2e - 16$ ***
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.39: Analiza varijance za restringirani model 2 za županije

Usporedbu adekvatnosti opisa podataka punog i restringiranog modela 2 ispituje se

pomoću naredbe *anova*. Nul hipotezu je da puni model (Model 0) bolje opisuje podatke od restringiranog (Model 2).

<i>Model0 : log(Doh) ~</i>					
<i>Por + St_nez + Obr + Br_stan + Ml_stan + St_stan + as.factor(More) + Polj</i>					
<i>Model2 : log(Doh) ~ St_nez + Obr</i>					
	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
0	11	0.0289			
2	18	0.1023	-7	-0.0734	0.0002 ***
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

Izvor: Izrada autorice

Tablica 3.40: Analiza devijance

Odbacuje se nul hipoteza zbog p-vrijednosti manje od 0.05 te se može zaključiti da je restringirani model 2 bolji u opisu podataka. Dakle restringirani model koji sadrži samo varijable *St_nez* i *Obr* ima manje varijabli od restringiranog modela 1, sve varijable u modelu su značajne te je značajno bolji u opisu podataka od punog modela. Ostaje još ispitati nul hipotezu da je restringirani model 2 (Model 2) bolji od restringiranog modela 1 (Model 1).

<i>Model2 : log(Doh) ~ St_nez + Obr</i>					
<i>Model1 : log(Doh) ~ St_nez + Obr + Ml_stan + Polj</i>					
	Stupnjevi slobode (Df)	Devijanca	Rezidualni Df	Rezidualna devijanca	<i>Pr(> Chi)</i>
2	18	0.1023			
1	16	0.0819	2	0.0203	0.1372
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1					

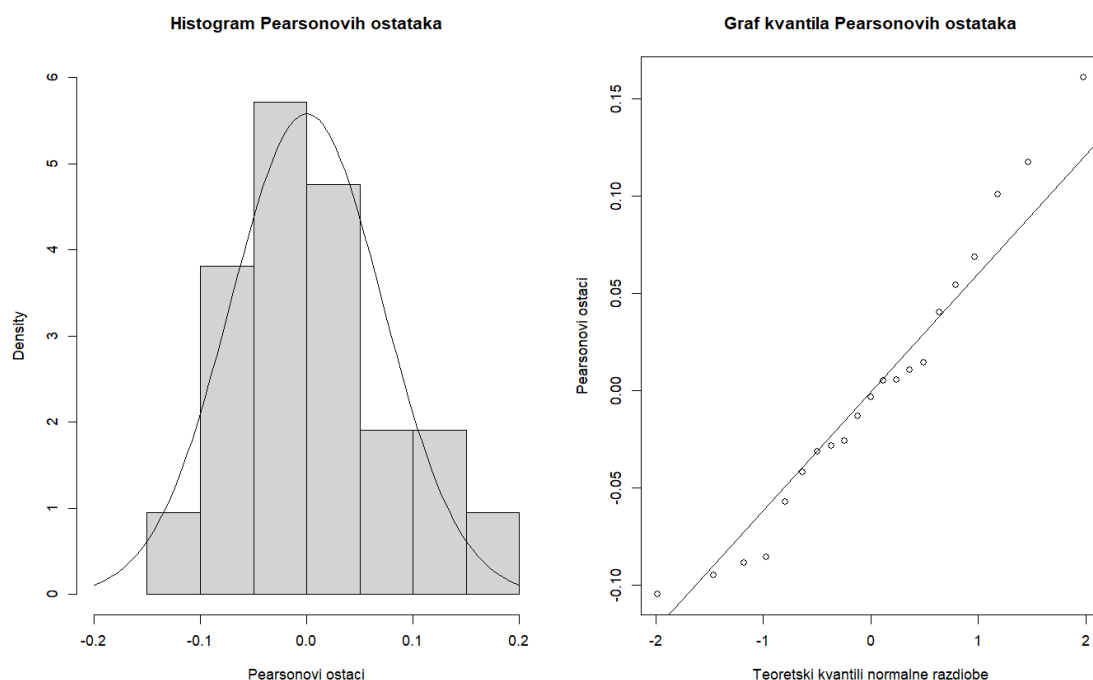
Izvor: Izrada autorice

Tablica 3.41: Analiza devijance

Na danoj razini značajnosti ne može se odbaciti hipoteza da je restringirani model 2 oblika (3.7) bolji u opisu podataka od restringiranog modela 1 (3.6). Ako se uspoređi AIC danih modela, vidi se da je vrijednost veća za restringirane modele nego za puni model.

Unatoč značajnom povećanju u AIC vrijednosti, zbog velikog smanjenja broja kovarijata te manje korelacije među kovarijatama i dalje je razumno koristiti restringirani model 2.

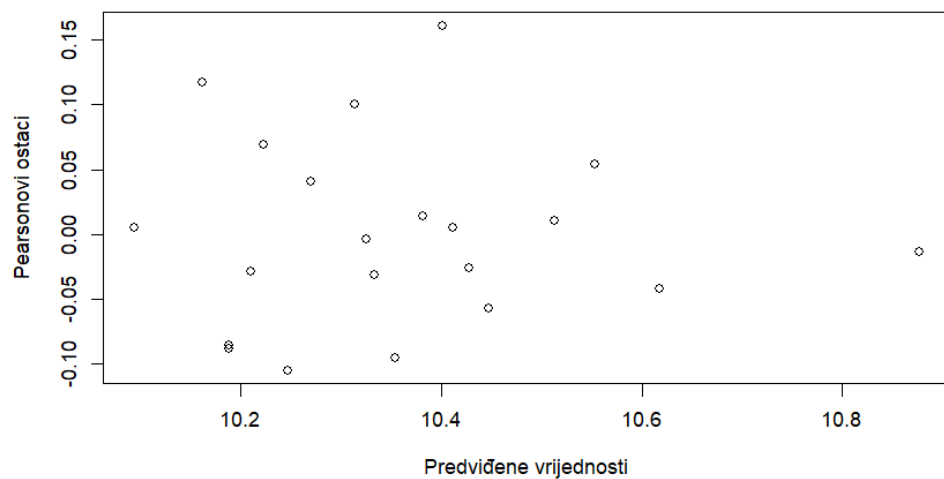
Sljedeće što treba analizirati su ostaci restringiranog modela kako bi se dodatno provjerilo koliko dobro model opisuje logaritmiranu vrijednost prosječnog dohotka po stanovniku. Na Slici 3.5 vide se histogram i graf kvantila Pearsonovih ostataka. Podataka ima malo pa se na slici ne može lako potvrditi dolaze li podaci iz normalne razdiobe, ali Lili-efors test normalnosti daje p-vrijednost od 0.3971 pa se ne može odbaciti nul hipoteza da podaci dolaze iz normalne razdiobe.



Izvor: Izrada autorice

Slika 3.5: Analiza normalnosti Pearsonovih ostataka

Na Slici 3.6 prikazane su predviđene vrijednosti modela naspram Pearsonovih ostataka. Na grafu nema uzorka ni trenda među podacima, dakle može se zaključiti da model adekvatno opisuje varijablu odziva.



Izvor: Izrada autorice

Slika 3.6: Analiza Pearsonovih ostataka

Analiza rezultata za županije

Ime varijable	Zasebni model	Puni model	Restringirani model
Por	✓+	+	<i>x</i>
St_nez	✓-	✓-	✓-
Obr	✓+	✓+	✓+
Br_stan	✓+	✓+	<i>x</i>
MI_stan	✓-	✓-	<i>x</i>
St_stan	-	✓-	<i>x</i>
More	+	+	<i>x</i>
Polj	✓+	✓-	<i>x</i>
Gus	✓+	-	<i>x</i>

✓statistički značajan utjecaj
 + pozitivan utjecaj - negativan utjecaj
x ne nalazi se u tom modelu

Izvor: Izrada autorice

Tablica 3.42: Rezultati za županije

U Tablici 3.42 mogu se vidjeti sažeti rezultati za sva tri modela gledana u ovom poglavlju. Prikazane su varijable koje su u danim modelima ispale statistički značajne te kako one utječu - pozitivno ili negativno na varijablu odziva.

Analizom devijance pomoću naredbe *anova* dobilo se da je restringirani model 2 (3.7) bolji od punog modela i restringiranog modela 1. Provedbom metode opisane u Poglavlju 2 dobivena je sljedeća analiza hipoteza:

Hipoteza 1: Prosječni porez na dohodak po stanovniku značajno utječe na dohodak.

Varijabla prosječnog poreza na dohodak po stanovniku nije značajna za opis varijable odziva te, iako je značajno smanjila rezidualnu devijancu u modelu, nije ostala u restringiranim modelima. Također je visoko korelirana s varijablom visoko obrazovanog stanovnika pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Hipoteza 2: Stopa nezaposlenosti ima značajan utjecaj na dohodak.

Varijabla stope nezaposlenosti značajno je smanjila rezidualnu devijancu u punom modelu i nije visoko korelirana ni s jednom od kovarijata pa je, iako nije značajno utjecala na varijablu odziva u punom modelu, ostala u restringiranim modelu 2. Povećanje te varijable za jednu jedinicu uzrokuje smanjenje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0212 jedinice.

Hipoteza 3: Postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj na dohodak.

Varijabla postotka stanovništva sa završenim visokim obrazovanjem značajna je za opis varijable odziva i značajno je smanjila rezidualnu devijancu u modelu te je ostala u restringiranom modelu 2. Povećanje te varijable za jednu jedinicu uzrokuje povećanje očekivanja logaritmirane vrijednosti prosječnog dohotka po stanovniku za 0.0333 jedinice.

Hipoteza 4: Broj stanovnika značajno i pozitivno utječe na rast dohotka.

Varijabla broja stanovnika nije značajno smanjila rezidualnu devijancu i visoko je korelirana s varijablom visoko obrazovanog stanovnika pa, iako je značajno utjecala na varijablu odziva u punom modelu, nije ostala u restringiranim modelima.

Hipoteza 5: Postotak stanovnika mlađih od 14 godina negativno utječe na dohodak.

Varijabla postotka stanovnika mlađih od 14 godina značajno je utjecala na varijablu odziva i na rezidualnu varijancu punog modela, ali isto nije vrijedilo i u restringiranom modelu 1 te nije uključena u restringirani model 2 koji je odabran kao najbolji u opisu podataka.

Hipoteza 6: Postotak stanovnika starijih od 65 godina negativno utječe na dohodak.

Varijabla postotka stanovništva starijih od 65 godina nije značajno smanjila rezidualnu devijancu u punom modelu pa, iako je značajna za opis varijable odziva, nije ostala u restringiranim modelima.

Hipoteza 7: Pristup moru ima pozitivan utjecaj na dohodak.

Varijabla pristupa moru nije značajna za opis varijable odziva i nije značajno smanjila rezidualnu devijancu u modelu pa nije ostala u restringiranim modelima. Također je visoko korelirana s varijablom visoko obrazovanog stanovnika pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Hipoteza 8: Broj kućanstava s pristupom poljoprivrednom zemljištu negativno i značajno utječe na dohodak.

Varijabla broja kućanstava s pristupom poljoprivrednom zemljištu značajno je utjecala na varijablu odziva i na rezidualnu varijancu punog modela, ali isto nije vrijedilo i u restringiranom modelu 1 te nije uključena u restringirani model 2 koji je odabran kao najbolji u opisu podataka.

Hipoteza 9: Gustoća naseljenosti značajno i pozitivno utječe na dohodak.

Varijabla gustoće naseljenosti nije značajna za opis varijable odziva i nije značajno smanjila rezidualnu devijancu u modelu pa nije ostala u restringiranim modelima. Također je visoko korelirana s varijablom visoko obrazovanog stanovnika pa bi njeno uključivanje u model zajedno s navedenom varijablom moglo utjecati na točnost modela u opisu podataka.

Analiza ukupnih rezultata za općine, gradove i županije

U Tablici 3.43 mogu se vidjeti sažeti rezultati za sve tri skupine podataka za sva tri modela gledana u ovom poglavlju. Prikazane su varijable koje su u danim modelima ispale statistički značajne te kako one utječu - pozitivno ili negativno na varijablu odziva.

Ime varijable	Općine			Gradovi			Županije		
	Z	P	R	Z	P	R	Z	P	R
Por	✓+	+	<i>x</i>	✓+	✓+	✓+	✓+	+	<i>x</i>
St_nez	✓-	✓-	✓-	✓-	✓-	✓-	✓-	✓-	✓-
Obr	✓+	✓+	✓+	✓+	✓+	✓+	✓+	✓+	✓+
Br_stan	+	✓-	<i>x</i>	✓+	+	<i>x</i>	✓+	✓+	<i>x</i>
Ml_stan	✓-	-	<i>x</i>	✓-	✓-	<i>x</i>	✓-	✓-	<i>x</i>
St_stan	✓-	✓-	✓-	+	-	<i>x</i>	-	✓-	<i>x</i>
More	✓+	-	<i>x</i>	✓+	✓-	✓-	+	+	<i>x</i>
Polj	+	✓+	✓-	+	-	<i>x</i>	✓+	✓-	<i>x</i>
Gus	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	✓+	-	<i>x</i>

✓ statistički značajan utjecaj
 + pozitivan utjecaj - negativan utjecaj
x ne nalazi se u modelu
 Z zasebni model, P puni model, R restringirani model
 Izvor: Izrada autorice

Tablica 3.43: Ukupni rezultati

Hipoteza 1: Prosječni porez na dohodak po stanovniku značajno utječe na dohodak.

U modelima koji sadrže samo slobodan član i varijablu prosječnog poreza na dohodak po stanovniku oblika (3.1), varijabla utječe pozitivno i značajno na varijablu odziva za općine, gradove i županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe neznačajno i pozitivno za općine i županije te značajno i pozitivno na varijablu odziva za gradove.

U izabranom optimalnom modelu opisa podataka, varijabla prosječnog poreza na dohodak po stanovniku značajno i pozitivno utječe na prosječni dohodak po stanovniku za gradove. Međutim, nije zadržana u optimalnom modelu za općine i županije.

Povećanje dohotka i povećanje poreza na dohodak u pozitivnoj su korelaciji te je međusobni odnos najviše izražen u podacima za gradove. Osnovica poreza na dohodak u Republici Hrvatskoj ukupni je iznos dohotka od nesamostalnog rada, dohotka od samostalne djelatnosti, dohotka od imovine i imovinskih prava, dohotka od kapitala i drugog dohotka, koji se ostvari u tuzemstvu i u inozemstvu (načelo svjetskog dohotka), a umanjeno za osobni odbitak [14]. Budući da je značajni dio dohotka (iznos dohotka do visine odbitka) izuzet od oporezivanja, tek kod viših iznosa dohodaka dolazi do značajnog oporezivanja porezom na dohodak, a ti viši dohodci najviše su izraženi u gradovima gdje je najveći prosječni dohodak po stanovniku.

Zbog značajnog utjecaja dohotka na porez na dohodak, što je suprotan smjer utjecaja od onog koji se ispituje u ovom radu, potrebno bi bilo provesti dodatna istraživanja na podacima iz više vremenskih razdoblja u kojima su korištene različite stope poreza. Tada bi bila veća vjerojatnost da su rezultati u skladu sa stvarnošću i da nisu pod utjecajem nekih drugih faktora koji nisu ispitani u ovom radu..

Hipoteza 2: Stopa nezaposlenosti ima značajan utjecaj na dohodak.

U modelima koji sadrže samo slobodan član i varijablu stope nezaposlenosti oblika (3.1), varijabla utječe negativno i značajno na varijablu odziva za općine, gradove i županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva značajno i negativno za općine i gradove te neznajčajno i negativno za županije.

U izabranom optimalnom modelu opisa podataka, varijabla stope nezaposlenosti značajno i negativno je utjecala na prosječni dohodak po stanovniku u općinama, gradovima i županijama RH.

Stopa nezaposlenosti konzistentno je bila od značajnog negativnog utjecaja na prosječni dohodak po stanovniku. Nezaposlene osobe ponekad ostvaruju neke dohotke (dohotak od imovine i imovinskih prava, dohotak od kapitala i drugi dohodak), ali oni su manji od dohotka zaposlenih osoba pa to sveukupno rezultira manjim prosjekom dohotka po stanovniku.

Hipoteza 3: Postotak stanovništva sa završenim visokim obrazovanjem ima pozitivan utjecaj na dohodak.

U modelima koji sadrže samo slobodan član i varijablu postotka visoko obrazova-

nih stanovništva oblika (3.1), varijabla utječe pozitivno i značajno na varijablu odziva za općine, gradove i županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe značajno i pozitivno na varijablu odziva za općine, gradove i županije.

U izabranom optimalnom modelu opisa podataka, varijabla postotka visoko obrazovanog stanovništva značajno je i pozitivno utjecala na prosječni dohodak po stanovniku u općinama, gradovima i županijama RH.

Završeno visoko obrazovanje, uz stopu nezaposlenosti, najkonzistentnije je značajno utjecalo na povećanje prosječnog dohotka po stanovniku u svim jedinicama lokalne i područne samouprave RH. Više obrazovano stanovništvo ostvaruje više dohotke te time povećava i prosjek dohotka po stanovniku.

Hipoteza 4: Broj stanovnika značajno i pozitivno utječe na rast dohotka.

U modelima koji sadrže samo slobodan član i varijablu broja stanovnika oblika (3.1), varijabla utječe na varijablu odziva pozitivno i neznajčajno za općine te pozitivno i značajno za gradove i županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva značajno i negativno za općine, neznajčajno i pozitivno za gradove te značajno i pozitivno za županije.

Varijabla broja stanovnika nije u izabranom optimalnom modelu opisa podataka ni za jednu od skupina podataka.

Broj stanovnika ima varijabilni smjer utjecaja na varijablu odziva i nestalnu statističku značajnost u različitim modelima i u različitim skupovima podataka. Što je više stanovnika to je veća mogućnost da su među njima pojedinci s većim dohotkom od prosječnog koji bi time povećali prosjek dohotka po stanovniku u danoj teritorijalnoj jedinici. Isto tako veći broj stanovnika uzrokuje da se ukupan dohodak u danoj jedinici dijeli s većim brojem kako bi se dobio prosječan dohodak po stanovniku. Jedino u županijama broj stanovnika kroz modele više povećava ukupan dohodak nego što veći nazivnik smanjuje razlomak prosječnog dohotka po stanovniku i to na značajnoj razini.

Hipoteza 5: Postotak stanovnika mlađih od 14 godina negativno utječe na dohodak.

U modelima koji sadrže samo slobodan član i varijablu postotka stanovnika mlađih od 14 godina oblika (3.1), varijabla utječe na varijablu odziva negativno i značajno za općine, gradove i županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva negativno i neznačajno za općine te negativno i značajno za gradove i županije.

Varijabla postotka stanovnika mlađih od 14 godina nije u izabranom optimalnom modelu opisa podataka ni za jednu od skupina podataka.

Broj stanovništva mlađe od 14 godina uvijek utječe negativno na prosjek dohotka po stanovniku, ali najčešće nije od velikog značaja. U RH raditi mogu osobe s navršenih 15 godina, uz suglasnost roditelja ili zakonskog skrbnika, a po Pravilniku o poslovima na kojima se ne smije zaposliti maloljetnik [19], zabranjen je rad maloljetnim učenicima na poslovima koji su utvrđeni kao poslovi s posebnim uvjetima rada i poslovi koji su utvrđeni kao osobito teški i štetni za zdravlje. Također, zabranjen je rad maloljetnim učenicima na poslovima koji su objektivno iznad njihovih fizičkih ili psiholoških sposobnosti. Zabranjen je rad na poslovima u kockarnicama, salonima za igre na sreću, disko klubovima te noćnim barovima i noćnim klubovima. Dakle, osobe do 14 godina mogu ostvarivati dohodak od imovine i imovinskih prava te dohodak od kapitala, što su značajno manji izvori dohotka od prosječne zaposlene osobe pa je razuman zaključak negativnog utjecaja postotka stanovnika mlađih od 14 na prosječan dohodak po stanovniku.

Hipoteza 6: Postotak stanovnika starijih od 65 godina negativno utječe na dohodak.

U modelima koji sadrže samo slobodan član i varijablu postotka stanovnika starijih od 65 godina oblika (3.1), varijabla utječe na varijablu odziva negativno i značajno za općine, pozitivno i neznačajno za gradove te negativno i neznačajno za županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva značajno i negativno za općine i županije te neznačajno i negativno za gradove.

Varijabla postotka stanovnika starijih od 65 godina nije u izabranom optimalnom modelu opisa podataka za gradove i županije, dok je u modelu za općine bila značajnog i negativnog utjecaja.

Dana varijabla utječe negativno na prosječni dohodak, kao i pretpostavljeno, ali je utje-

caj te varijable neznačajan osim u modelu za općine. Iznos prosječne mirovine u RH [22] značajno je manji od prosječnog dohotka zaposlene osobe. Izvori drugog dohotka kao ostvarivani dohodak od imovine i imovinskih prava i dohotka od kapitala nisu dovoljni da osobe starije od 65 godina budu značajna varijabla u ostvarenju dohotka.

Hipoteza 7: Pristup moru ima pozitivan utjecaj na dohodak.

U modelima koji sadrže samo slobodan član i varijablu pristupa moru oblika (3.1), varijabla utječe na varijablu odziva pozitivno i značajno za općine i gradove te pozitivno i neznačajno za županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva negativno i neznačajno za općine, značajno i negativno za gradove te pozitivno i neznačajno za županije.

Varijabla pristupa moru nije u izabranom optimalnom modelu opisa podataka za općine i županije, dok je u modelu za gradove bila značajnog i negativnog utjecaja.

Iako je turizam značajan dohodak za RH, pristup moru po ovom radu najčešće nema velik značaj na prosječni dohodak po stanovniku te se utjecaj dane varijable često mijenja s pozitivnog na negativni ovisno o skupini podataka. Negativan i značajan utjecaj za gradove moguć je zbog kopnenih gradova s visokim prosječnim dohotkom kao što su Zagreb, Osijek, Velika Gorica i drugi. Velikoj razlici između dobivenih rezultata mogla bi pridonositi i siva ekonomija u turizmu. Siva ekonomija podrazumijeva obavljanje ekonomske aktivnosti izvan odgovarajuće pravne regulative ili se protivi propisima ([12]). Zbog neprijavlivanja dijela dohotka jako je teško dobiti točne podatke za područja u kojima je siva ekonomija zastupljena, a jedno od tih je područja turizam.

Hipoteza 8: Broj kućanstava s pristupom poljoprivrednom zemljištu negativno i značajno utječe na dohodak.

U modelima koji sadrže samo slobodan član i varijablu broja kućanstava s pristupom poljoprivrednom zemljištu oblika (3.1), varijabla utječe na varijablu odziva pozitivno i neznačajno za općine i gradove te pozitivno i značajno za županije.

U punom modelu oblika (3.2) (tj. (3.3) za županije), dana varijabla utječe na varijablu odziva pozitivno i značajno za općine, negativno i neznačajno za gradove te negativno i značajno za županije.

Varijabla broja kućanstava s pristupom poljoprivrednom zemljištu nije u izabranom optimalnom modelu opisa podataka za gradove i županije, dok je u modelu za općine bila značajnog i negativnog utjecaja.

Dana varijabla različito utječe na većinu predstavljenih modela, nekad imajući pozitivan, a nekad negativan utjecaj na prosječan dohodak. Također varira i značajnost tog utjecaja. Poljoprivreda je jedna od značajnijih djelatnosti u RH te potrebno je imati više podataka vezanih uz to područje ili provesti dodatne analize na podacima za značajniji rezultat.

Hipoteza 9: Gustoća naseljenosti značajno i pozitivno utječe na dohodak.

U modelu koji sadrži samo slobodan član i varijablu gustoće naseljenosti oblika (3.1), varijabla utječe pozitivno i značajno na varijablu odziva za županije.

U punom modelu oblika (3.3) za županije dana varijabla utječe neznačajno i negativno na varijablu odziva.

Varijabla gustoće naseljenosti nije u izabranom optimalnom modelu opisa podataka za županije.

Gustoća naseljenosti najveća je u gradovima te se nastojalo povezati je s većim dohotkom, ali u ovom radu nije dobiven značajan rezultat koji bi potvrdio njenu značajnost u opisu prosječnog dohotka po stanovniku. Točnijim rezultatima pridonijela bi lakša dostupnost gustoće naseljenosti za općine i gradove RH.

Utjecaj gornje analize na početne hipoteze

Na kraju se mogu prikazati svi prethodno dobiveni podaci u Tablici 3.44 gdje su prikazane sve hipoteze te jesu li potvrđene na temelju podataka i provedene analize iz ovog rada.

Hipoteza 1 koja pretpostavlja značajnost prosječnog poreza na dohodak po stanovniku djelomično je potvrđena. Utjecaj je konzistentno značajan u modelima za gradove, ali nije bio značajan u punom modelu za općine i županije. Hipoteza 2 koja pretpostavlja značajnost stope nezaposlenosti potvrđena je na svim osim jednom modelu za županije dok je pozitivnost utjecaja visokog obrazovanja hipoteze 3 potvrđena na svim modelima u svim skupinama podataka. S druge strane, hipoteza 4 koja pretpostavlja da broj stanovnika utječe značajno i pozitivno nije potvrđena ni za jedan skup podataka. U različitim modelima je bila različitog smjera utjecaja i statističkog značaja te je potrebna detaljnija analiza

kako bi se hipoteza potvrdila. Negativnost utjecaja postotka stanovništva mlađeg od 14 te onog starijeg od 65 iz hipoteza 5 i 6 potvrđena je u većini obrađenih modela u kojima su te varijable bile pristune. Ipak, u te dvije varijable postoje manja odstupanja od hipoteze te bi bilo potrebno obaviti još analiza npr. na novijim podacima. Hipoteze 7 i 8 sa svojim pretpostavkama pozitivnog utjecaja pristupa moru te negativnom i značajnom utjecaju pristupa poljoprivrednom zemljištu imale su najnestabilnije zaključke među modelima i skupovima podataka. Kako bi se hipoteze potvrdile ili opovrgnule potrebno je provesti još analiza, pogotovo na novijim podacima. Gustoća naseljenosti bila je dostupna samo za jednu skupinu podataka te je također potrebna daljnja analiza.

Hipoteza	Općine	Gradovi	Županije	Ukupno
Hipoteza 1	x	✓	x	x
Hipoteza 2	✓	✓	$x+$	✓
Hipoteza 3	✓	✓	✓	✓
Hipoteza 4	x	x	$x+$	x
Hipoteza 5	✓	✓	✓	✓
Hipoteza 6	✓	$x+$	✓	✓
Hipoteza 7	x	x	$x+$	x
Hipoteza 8	x	x	x	x
Hipoteza 9	NA	NA	x	x
✓ Hipoteza je potvrđena				
$x+$ Hipoteza nije potvrđena, ali podaci ukazuju da bi mogla biti ispravna				
x Hipoteza nije potvrđena				

Izvor: Izrada autorice

Tablica 3.44: Hipoteze

Poglavlje 4

Zaključak

Ovaj rad bavi se odrednicama prosječnog dohotka po stanovniku u jedinicama lokalne i područne samouprave Republike Hrvatske. Prvo se predstavila literatura na temu odrednica dohotka i nejednakosti dohotka u raznim državama svijeta iz kojih se izabralo devet odrednica i postavilo devet hipoteza. Ekonomski podaci, tj. dohodak, porez i stopa nezaposlenosti, dobiveni su za 2019. godinu, dok se za demografske varijable koristio popis stanovništva 2011.

Zbog normalnosti logaritmiranih vrijednosti prosječnog dohotka po stanovniku, koristio se generalizirani linearni model zajedno s normalnom distribucijom. Generalizirani linearni model također je omogućio korištenje kategorijalne varijable pristupa moru. Analiza se provela nad tri modela za svaku od tri skupine podataka: općine, gradove i županije. Prvi od tih modela provjerava utjecaj svake varijable zasebno na dohodak dok drugi, puni model, sadržava sve varijable. Zadnji je model restringirana verzija prijašnjeg te su se oni usporedili u točnosti i optimalnosti opisa dohotka.

Dobiveni rezultati u različitoj mjeri podupiru početne hipoteze. Hipoteza da prosječni porez po stanovniku utječe značajno na prosječni dohodak po stanovniku potvrđena je za gradove, dok je za općine i županije imala različitu statističku značajnost ovisno o korištenom modelu. Stopa nezaposlenosti sa svojim negativnim i značajnim utjecajem te postotak sa završenim visokim obrazovanjem s pozitivnim i značajnim utjecajem na prosječni dohodak po stanovniku u skladu su s početnim hipotezama. Hipoteza da broj stanovnika značajno i pozitivno utječe na prosječni dohodak po stanovniku nije potvrđena jer je u promatranim modelima i teritorijalnim jedinicama utjecaj bio različitog smjera i različite statističke značajnosti. Potvrđene su hipoteze da postotak stanovnika mlađih od 14 godina te onih starijih od 65 godina negativno utječu na prosječni dohodak po stanovniku. Varijabla pristupa moru imala je varijabilan utjecaj na dohodak u modelima i

u različitim skupinama podataka te nije potvrđena hipoteza da pristup moru ima pozitivan utjecaj. Slična je situacija i za varijablu broja kućanstva s pristupom poljoprivrednom zemljištu, gdje je uz varijabilan smjer utjecaja također i varijabilan značaj tog utjecaja te nije moguće potvrditi hipotezu da dana varijabla utječe negativno i značajno na prosječni dohodak po stanovniku. Iz istih razloga se nije mogla potvrditi ni hipoteza da gustoća naseljenosti ima značajan i pozitivan utjecaj na dohodak.

Iako je ovaj rad uzeo u obzir devet varijabli, više različitih modela i proveo analizu nad tri skupa podataka, i dalje treba uzeti rezultate s oprezom i potrebna je daljnja analiza za točnije rezultate. Rezultate varijable prosječnog poreza na dohodak treba uzeti s dozom opreza zbog jake korelacije sa samim dohotkom i utjecaja dohotka na porez na dohodak što je obrnut smjer utjecaja naspram onog koji se ispituje u ovom radu. Trebala bi se izvesti dodatna analiza s podacima iz više različitih godina u kojima su bile korištene različite stope poreza. Varijable kao što su broj stanovnika, varijabla pristupa moru i pristupa poljoprivrednom zemljištu bi trebalo dodatno ispitati ili uvesti druge varijable slične njima kako bi se objasnio njihov varirajući utjecaj na razne modele i skupove podataka. Varijabla pristupa moru mogla bi se zamijeniti postotkom dohotka dobivenog od turizma, a broj kućanstava s pristupom poljoprivrednom zemljištu postotkom dohotka dobivenog od poljoprivrede. Također, gustoća naseljenosti bila je dostupna samo za županije te bi ta varijabla za općine i gradove pomogla boljem testiranju početne hipoteze. Mogla bi se provesti i analiza na podacima s novog popisa stanovništva iz 2022. nakon što budu dostupni te bi se mogli usporediti s ovdje dobivenim rezultatima.

Bibliografija

- [1] B. Basrak, H. Planinić, *Financijski praktikum, Generalizirani linearni modeli*, Svi-banj 2022, https://www.pmf.unizg.hr/images/50025954/GLM2022_Notes.pdf, posjećena 2022-09-27.
- [2] B. Su, A. Heshmati, *Analysis of the determinants of income and income gap between urban and rural China*, *China Economic Policy Review* **2** (2013), br. 01, 1350002.
- [3] D. C. Monchuk, J. A. Miranowski, D. J. Hayes, B. A. Babcock, *An analysis of regional economic growth in the US Midwest*, *Applied Economic Perspectives and Policy* **29** (2007), br. 1, 17–39.
- [4] D. S. Mapa, A. M. Balisacan, K. J. S. Briones, *Robust determinants of income growth in the Philippines*, *Philippine Journal of Development* **33** (2006), br. 1-2, 1–32.
- [5] Ministarstvo financija, *PR-RAS i RAS-funkc za razdoblje 2014. - 2020., 2021.*, <https://mfin.gov.hr/istaknute-teme/lokalna-samouprava/financijski-izvjestaji-jlp-r-s/pr-ras-i-ras-funkc-za-razdoblje-2014-2020/3173>, posjećena 2022-10-03.
- [6] G. Auten, R. Carroll, *The effect of income taxes on household income*, *Review of economics and statistics* **81** (1999), br. 4, 681–693.
- [7] H. N. Mocan, *Structural unemployment, cyclical unemployment, and income inequality*, *Review of Economics and Statistics* **81** (1999), br. 1, 122–134.
- [8] J. A. Nelder, R. W. M. Wedderburn, *Generalized linear models*, *Journal of the Royal Statistical Society: Series A (General)* **135** (1972), br. 3, 370–384.
- [9] M. Blejer, I. Guerrero, *The impact of macroeconomic policies on income distribution: an empirical study of the Philippines*, *The Review of Economics and Statistics* (1990), 414–423.

- [10] M. Huzak, *Statistika, Opisna statistika 2*, 2021, <https://web.math.pmf.unizg.hr/nastava/stat/files/StatOpisnastat2.pdf>, posjećena 2022-12-17.
- [11] M. O. Odedokun, J. I. Round, *Determinants of income inequality and its effects on economic growth: evidence from African countries*, br. 2001/103, WIDER Discussion Paper, 2001.
- [12] M. Topić, *SIVA EKONOMIJA U HRVATSKOJ*, 2006, <https://www.matica.hr/hr/342/siva-ekonomija-u-hrvatskoj-20998/>, posjećena 2022-12-16.
- [13] Ministarstvo Financija, Porezna Uprava, *Dohodak*, https://www.porezna-uprava.hr/pozivni_centar/Stranice/dohodak.aspx, posjećena 2022-12-17.
- [14] ———, *Porez na dohodak*, https://www.porezna-uprava.hr/HR_porezni_sustav/Stranice/porez_na_dohodak.aspx, posjećena 2022-11-15.
- [15] N. D. Bogović, S. Drezgić, S. Čegar i ostali (2017), *EVALUACIJA POSTOJEĆEG I PRIJEDLOG NOVOG MODELA ZA IZRAČUN INDEKSA TE IZRAČUN NOVOG INDEKSA RAZVIJENOSTI JEDINICA LOKALNE I PODRUČNE SAMOUPRAVE U REPUBLICI HRVATSKOJ [Studija]*, Centar za lokalni ekonomski razvoj Ekonomskog fakulteta Sveučilišta u Zagrebu i Rijeci.
- [16] P. K. Dunn, G. K. Smyth i ostali, *Generalized linear models with examples in R*, sv. 53, Springer, 2018.
- [17] R. G. Holcombe, D. J. Lacombe, *The effect of state income taxation on per capita income growth*, *Public Finance Review* **32** (2004), br. 3, 292–312.
- [18] R. H. Myers, D. C. Montgomery, *A tutorial on generalized linear models*, *Journal of Quality Technology* **29** (1997), br. 3, 274–291.
- [19] Ministarstvo rada i mirovinskoga sustava, *Pravilnik o poslovima na kojima se ne smije zaposliti maloljetnik*, Kolovoz 2015, https://narodne-novine.nn.hr/clanci/sluzbeni/2015_08_89_1742.html, posjećena 2022-11-15.
- [20] S. Démurger, J. Sachs, W. T. Woo, S. Bao, G. Chang, A. Mellinger, *Geography, economic policy, and regional development in China*, *Asian Economic Papers* **1** (2002), br. 1, 146–197.
- [21] Zach Statology, *How to Interpret glm Output in R*, Studeni 2021, <https://www.statology.org/interpret-glm-output-in-r/>, posjećena 2022-10-27.

- [22] Hrvatski zavod za mirovinsko osiguranje, *Aktualna statistika za rujan 2022.- isplata u listopadu 2022.*, <https://www.mirovinsko.hr/hr/aktualna-statistika-za-rujan-2022-isplata-u-listopadu-2022/148>, posjećena 2022-11-15.
- [23] Državni zavod za statistiku, *Popis stanovništva, kućanstava i stanova 2011.*, 2011, <https://web.dzs.hr/Hrv/censuses/census2011/results/censustabsxls.htm>, posjećena 2022-09-27.

Sažetak

Ovaj rad analizira utjecaj odabranih varijabli na prosječni dohodak po stanovniku u jedinicama lokalne i područne samouprave Republike Hrvatske. Analizira se devet varijabli: prosječni porez po stanovniku, stopa nezaposlenosti, postotak stanovništva sa završenim visokim obrazovanjem, broj stanovnika, postotak stanovništva mlađeg od 14 godina, postotak stanovništva starijeg od 65 godina, pristup moru, broj kućanstava s pristupom poljoprivrednom zemljištu i gustoća naseljenosti te se provjerava njihov utjecaj na prosječni dohodak po stanovniku. Analiza se provodi zasebno za 428 općina, 128 gradova i 21 županiju. Ekonomski podaci (dohodak, porez i stopa nezaposlenosti) dobiveni su za 2019. godinu, dok su se za demografske varijable koristio popis stanovništva iz 2011.

Analiza podataka temelji se na generaliziranom linearnom modelu. Na tri su se modela putem programskoj jezika R dobile procjene utjecaja varijabli na dohodak po stanovniku na svakoj skupini podataka. Prvo se obradio model u kojem se nalaze svaka od izabranih varijabli zasebno, potom model koji je sadržavao sve varijable. Posljednji je model restringirana verzija punog modela dobiven analizom značajnosti varijabli u modelu i visine korelacije između tih varijabli. Puni i restringirani modeli usporedili su se pomoću statistike omjera vjerodostojnosti i u svakom skupu podataka zaključilo se da postoji model s manje varijabli koji jednako dobro ili bolje opisuje podatke od punog modela.

Dobiveni rezultati ukazuju da prosječni porez po stanovniku utječe pozitivno na prosječni dohodak po stanovniku, ali s različitom statističkom značajnošću u različitim teritorijalnim jedinicama. Značajan utjecaj imaju stopa nezaposlenosti, koja utječe negativno, i postotak stanovništva sa završenim visokim obrazovanjem, koji ima pozitivan utjecaj. Varijable broja stanovnika, postotka stanovništva mlađe od 14 su u većini modela bile statistički neznačajne. Postotak stanovništva starijeg od 65 ima konzistentno negativan utjecaj, ali zbog različite razine značajnosti nije dobiven jasan rezultat. Pristup moru i pristup poljoprivrednom zemljištu imali su različit utjecaj i statističku značajnost ovisno o odabranoj teritorijalnoj jedinici i modelu te nije moguće donijeti zaključak na razini države bez daljnje analize. Gustoća naseljenosti, koja je bila dostupna samo za županije, ima različit smjer utjecaja i različitu statističku značajnost te također nije jasno definiran njezin utjecaj.

Summary

This paper analyzes the impact of selected variables on the average income per capita (p.c.) in local and regional self-government units of the Republic of Croatia. Nine variables are analyzed: average tax p.c., unemployment rate, percentage of the population with completed tertiary education, number of inhabitants, percentage of the population younger than 14, percentage of the population older than 65, access to the sea, number of households with access to agricultural land and population density. The analysis is carried out separately for 428 municipalities, 128 cities and 21 counties. Economic data (income, tax and unemployment rate) were obtained for the year 2019, while the 2011 census was used for demographic variables.

The data analysis is based on a generalized linear model. Estimates of the impact of variables on income p.c. on each group of data were obtained using the R programming language. First, the model containing each of the selected variables was processed separately, and then the model that contained all variables was analysed. The last model is a restricted version of the full model obtained by analysing the significance of the variables in the model and the level of correlation between those variables. The full and restricted models were compared using the likelihood ratio statistic, and in each data set it was concluded that there was a model with fewer variables that described the data as well or even better than the full model.

The obtained results indicate that the average tax p.c. has a positive effect on the average income p.c., but with different statistical significance in different territorial units. The unemployment rate, which has a negative impact, and the percentage of the population with completed tertiary education, which has a positive impact, have a significant influence. The number of inhabitants, the percentage of the population under the age of 14 were statistically insignificant in the majority of models. The percentage of the population older than 65 has a consistently negative impact, but due to different levels of significance, no clear result was obtained. The variables signifying access to the sea and access to agricultural land had different influence and statistical significance depending on the selected territorial unit and model, so it was not possible to draw a conclusion at state level without further

analysis. Population density, which was available only for counties, has different direction of influence and statistical significance, and its impact is also not clearly defined.

Životopis

Tina Krištofić rođena je 18. srpnja 1998. u Zagrebu. Prvih sedam godina školovanja provodi u osnovnoj školi u Svetoj Nedelji te zadnju godinu osnovne škole završava u područnoj školi Vladimira Deščaka u Rakitju. Nakon toga upisuje opću gimnaziju Lucijana Vranjanina u Zagrebu koju završava 2017. Preddiplomski sveučilišni studij Matematika na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu završava 2020. i dobiva titulu sveučilišne prvostupnice matematike, nakon čega upisuje sveučilišni diplomski studij Financijska i poslovna matematika na istom fakultetu.