

# Modeliranje i prognoziranje prihoda u hotelu Nautica

---

Grčar-Lovrić, Marko

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Tourism and Hospitality Management / Sveučilište u Rijeci, Fakultet za menadžment u turizmu i ugostiteljstvu**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:191:440218>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2025-02-07**



Repository / Repozitorij:

[Repository of Faculty of Tourism and Hospitality Management - Repository of students works of the Faculty of Tourism and Hospitality Management](#)



**SVEUČILIŠTE U RIJECI**  
**Fakultet za menadžment u turizmu i ugostiteljstvu u Opatiji**  
**Sveučilišni diplomski studij**

**MARKO GRČAR LOVRIĆ**

**Modeliranje i prognoziranje prihoda u hotelu Nautica**

**Income modelling and forecasting in Hotel Nautica**

Diplomski rad

Opatija, 2023.

**SVEUČILIŠTE U RIJECI**  
**Fakultet za menadžment u turizmu i ugostiteljstvu**  
**Sveučilišni diplomski studij**  
**Menadžment u hotelijerstvu**

**Modeliranje i prognoziranje prihoda u hotelu Nautica**  
**Income modelling and forecasting in Hotel Nautica**

Diplomski rad

Kolegij: **Ekonometrija**

Student: **Marko GRČAR LOVRIĆ**

Mentor: **prof. dr. sc. Tea BALDIGARA**

Matični broj: **3614/21**

Komentor: **doc. dr. sc. Jelena MUŠANOVIĆ**

Opatija, 2023.

Sadržaj:

1. Regresijska analiza .....	3
1.1. Korelacijska analiza .....	3
1.2. Jednostavni regresijski modeli .....	5
1.3. Višestruki regresijski model .....	8
1.4. Pretpostavke klasičnoga linearnog regresijskog modela .....	9
1.5. Procjena i interpretacija regresijskih parametara .....	16
1.6. Regresijski modeli s kvalitativnim regresorskim varijablama .....	19
1.7. Nelinearni regresijski modeli .....	20
2. Analiza valjanosti regresijskog modela.....	24
2.1. Analiza varijance .....	24
2.2. Pokazatelji pouzdanosti regresijskog modela.....	26
2.3. Testiranje hipoteza u regresijskom modelu .....	28
2.3.1. Testiranje značajnosti regresijskih parametara .....	29
2.3.2. Testiranje značajnosti regresijskog modela.....	30
3. Prognoziranje – pojmovno određenje.....	32
3.1. Definicija i faze prognoziranja .....	32
3.2. Metode prognoziranja u turizmu .....	36
3.3. Mjere prognostičke efikasnosti .....	40
4. Modeliranje i prognoziranje ukupnih prihoda u hotelu <i>Nautica</i> .....	44
4.1. Deskriptivna analiza varijabli korištenih u analizi .....	46
4.2. Ekonometrijsko modeliranje ostvarenih ukupnih prihoda .....	52
4.3. Prognoziranje ukupnih prihoda .....	58



# IZJAVA O AUTORSTVU RADA I O JAVNOJ OBJAVI RADA

**Marko Grčar Lovrić**

**ds3614/21**

---

(ime i prezime studenta)

(matični broj studenta)

---

**Modeliranje i prognoziranje prihoda u hotelu *Nautica***

---

(naslov rada)

Izjavljujem da sam ovaj rad samostalno izradila/o, te da su svi dijelovi rada, nalazi ili ideje koje su u radu citirane ili se temelje na drugim izvorima, bilo da su u pitanju knjige, znanstveni ili stručni članci, Internet stranice, zakoni i sl. u radu jasno označeni kao takvi, te navedeni u popisu literature

Izjavljujem da kao student–autor završnog rada, dozvoljavam Fakultetu za menadžment u turizmu i ugostiteljstvu Sveučilišta u Rijeci da ga trajno javno objavi i besplatno učini dostupnim javnosti u cjelovitom tekstu u mrežnom digitalnom repozitoriju Fakulteta za menadžment u turizmu i ugostiteljstvu Sveučilišta u Rijeci.

U svrhu podržavanja otvorenog pristupa završnim radovima trajno objavljenim u javno dostupnom digitalnom repozitoriju Fakulteta za menadžment u turizmu i ugostiteljstvu Sveučilišta u Rijeci, ovom izjavom dajem neisključivo imovinsko pravo iskorištavanja bez sadržajnog, vremenskog i prostornog mog završnog rada kao autorskog djela pod uvjetima *Creative Commons* licencije CC BY Imenovanje, prema opisu dostupnom na <http://creativecommons.org/licenses/>.

U Opatiji, 28. 6. 2023.

---

Potpis studenta

## Sažetak

Cilj diplomskog rada jest modelirati i prognozirati kretanje ukupnih mjesečnih prihoda u hotelu *Nautica* s obzirom na ostvareni broj noćenja i prosječnu cijenu po noćenju.

Istraživanje obuhvaća vremensko razdoblje od siječnja 2017. godine do prosinca 2022. godine. U analizi su korišteni dvostruko logaritamski modeli višestruke regresijske analize. Pored kvantitativnih regresijskih varijabli, korištene su i sezonske *dummy* varijable. Rezultati analize pokazali su da kretanje ukupnih prihoda pokazuje znakove sezonalnosti te da su svi regresijski koeficijenti značajni su na razini od 1%. Broj noćenja i prosječna cijena po noćenju u velikoj mjeri objašnjavaju kretanje ukupnih mjesečnih prihoda u hotelu *Nautica*. Uz sve zadovoljene pretpostavke klasične linearne regresije, procijenjeni model koji se koristi za prognozu sklon je prognostičkim pogreškama koje je nemoguće predvidjeti.

Rezultati istraživanja od koristi su menadžerima te mogu poslužiti kao temelj za donošenje poslovnih odluka, planiranje poslovnih procesa, nabavu potrošne robe, stalnu i tekuću imovinu, za zadovoljavanje potrebe za ljudskim resursima, efikasnost rada, i druge poslovne aktivnosti te očekivani pokazatelji uspješnosti poslovanja.

Ključne riječi: *modeliranje; prognoziranje; broj noćenja, cijena noćenja, prihodi, hotel*

## Uvod

U suvremeno se doba poslovne promjene dešavaju veoma rapidno. Brza prilagodba promjenama i uvjetima na tržištu neophodna je za konkurentan nastup poduzeća. Kako bi poduzeće moglo brzo odgovoriti na poslovne promjene, ono mora kontinuirano analizirati i modelirati podatke iz poslovanja u potrazi za određenim objašnjenjima.

Ekonometrijsko analiziranje i modeliranje podataka omogućava pretvorbu poslovnih podataka u složene informacije kojima se otkrivaju razne osobnosti podataka odnosno poslovanja, a mogu se i anticipirati poslovne promjene. Sposobnost razumijevanja i predviđanja kretanja podataka veoma je moćan alat u svim djelatnostima i industrijama jer omogućava pravodobnu prilagodbu novonastalim uvjetima na tržištu.

Turizam, kao jedan od najdinamičnijih sektora posebno je izazovno polje za ekonometrijske analize. Modeliranje i prognoziranje ključnih determinanti turističkoga sektora može se odnositi kako na turističku potražnju tako i na turističku ponudu. Rezultati ekonometrijskoga modeliranja i prognoziranja predstavljaju temelje gotovo svih poslovnih odluka.

Predmet istraživanja diplomskoga rada jest modeliranje i prognoziranje mjesečnih prihoda Hotela *Nautica* kao funkcije ostvarenih noćenja i prosječne cijene po noćenju u razdoblju od siječnja 2017. godine do prosinca 2022. godine. U analizi su korišteni dvostruko logaritamski modeli višestruke regresijske analize. Pored kvantitativnih regresijskih varijabli, korištene su i sezonske *dummy* varijable.

Temeljni cilj diplomskoga rada jest dobivanje uvida u kretanje ukupnih mjesečnih prihoda u Hotelu *Nautica* te prognoziranje njihova kretanja za 2023. godinu.

Rad se sastoji od četiri poglavlja. U prvome se poglavlju teorijski opisuje regresijska analiza i regresijski modeli korišteni u diplomskome radu. Drugo poglavlje bavi se pojmovima testiranja valjanosti regresijskih modela, a u trećem se poglavlju pojmovno određuje prognoziranje. U posljednjem, četvrtom poglavlju, modeliraju se i prognoziraju ukupni mjesečni prihodi hotela *Nautica*.

# 1. Regresijska analiza

Prilikom analiziranja podataka za više varijabli, dobro je poznavati stupanj povezanosti između varijabli što se postiže korelacijskom analizom. Ako se nadalje želi opisati navedena povezanost, to se može učiniti uz pomoć regresijskog modela.

## 1.1. Korelacijska analiza

Često korištena metoda numeričke analize dviju varijabli kojom se razmatra njihova povezanost i smjer kretanja naziva se korelacijska analiza.

Korelacijska analiza nastoji izmjeriti stupanj povezanosti između dvije varijable. Korelacija je analiza kovarijacije između dviju (ili čak više) varijabli, a za dvije varijable se smatra da su u korelaciji ako promjena u bilo kojoj od njih rezultira odgovarajućom promjenom u drugoj. Kod analiziranja korelacije nije cilj koristiti jednu varijablu za predviđanje druge, već pokazati snagu linearnog odnosa koji postoji između dviju numeričkih varijabli. Mjerenje jačine linearne povezanosti između dviju numeričkih varijabli izražava se koeficijentom korelacije. Koeficijent korelacije označava se s grčkim slovom ' $\rho$ ' (rho), a koeficijent korelacije uzorka slovom ' $r$ ' te se njihov raspon kreće između -1 do +1. Takav koncept korelacijskog koeficijenta nazvan je po Karlu Pearsonu, (1857. – 1936.), a naziva se i linearni koeficijent korelacije produkt-moment ili koeficijent korelacije Karla Pearsona. Ovaj se koeficijent općenito koristi kada su varijable kvantitativne prirode, tj. varijable intervala, omjera ili ljestvice. Koeficijent korelacije izražava se sljedećim izrazom:

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum_{i=1}^n (X_i - \bar{X})^2) (\sum_{i=1}^n (Y_i - \bar{Y})^2)}} \quad (1)$$

gdje su:

$\bar{X}, \bar{Y}$  – aritmetičke sredine varijabli

Ako je apsolutna vrijednost koeficijenta korelacije blizu apsolutnoj jedinici tada je riječ o usko povezanim varijablama. Niska apsolutna vrijednost koeficijenta korelacije znači da je veza

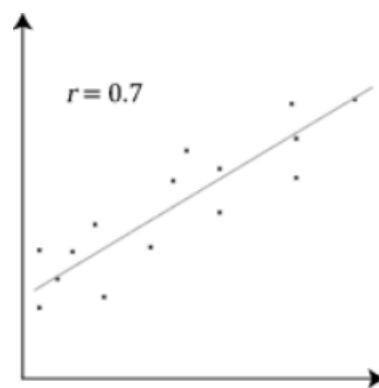


između varijabli slaba, no ne nužno. Povezanost između varijabli može biti krivolinijska, pa je upotreba koeficijenta linearne korelacije ponekad neprimjerena.<sup>1</sup>

Postoje dvije osnovne skupine korelacije:

1. pozitivna, negativna i nulta korelacija i
2. linearna i nelinearna korelacija.

Pozitivna korelacija prisutna je kada vrijednosti dviju varijabli odstupaju u istom smjeru, tj. ako povećanje ili smanjenje vrijednosti jedne varijable rezultira odgovarajućim povećanjem ili smanjenjem u vrijednostima druge varijable, tada se kaže da je korelacija pozitivna. Dvije su varijable u pozitivnoj korelaciji ako imaju tendenciju promjene (povećanja ili smanjenja) u istom smjeru. Na slici 1. prikazan je visok stupanj pozitivne korelacije gdje sve su točke u dijagramu raspršenosti u blizini linije s pozitivnim odnosno negativnim nagibom. Ako sva opažanja leže na ravnoj liniji koja je nagnuta prema gore, to ukazuje na savršenu pozitivnu korelaciju.

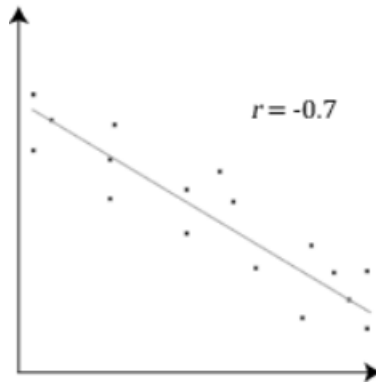


Slika 1.: Pozitivna korelacija (Izvor: [https://bs.wikipedia.org/wiki/Koeficijent\\_korelacije#/media/Datoteka:Pearson\\_Correlation\\_Coefficient\\_and\\_associated\\_scatterplots.png](https://bs.wikipedia.org/wiki/Koeficijent_korelacije#/media/Datoteka:Pearson_Correlation_Coefficient_and_associated_scatterplots.png))

Negativna korelacija prisutna je kada vrijednosti dviju varijabli odstupaju u suprotnom smjeru, odnosno ukoliko povećanje ili smanjenje vrijednosti jedne varijable rezultira odgovarajućim smanjenjem ili povećanjem vrijednosti druge varijable. Na slici 2. prikazan je dijagram raspršenosti gdje su sve točke u blizini linije s negativnim nagibom. Kada bi sva promatranja ležala na pravcu s negativnim nagibom bilo bi riječ o savršenoj negativnoj korelaciji.

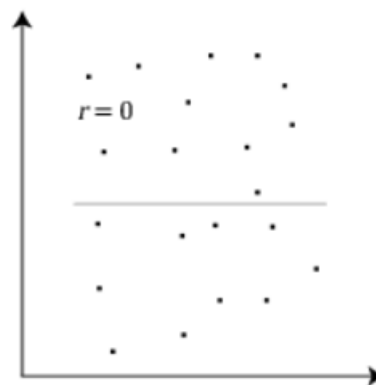
---

<sup>1</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 16.



Slika 2.: Negativna korelacija (Izvor: [https://bs.wikipedia.org/wiki/Koeficijent\\_korelacije#/media/Datoteka:Pearson\\_Correlation\\_Coefficient\\_and\\_associated\\_scatterplots.png](https://bs.wikipedia.org/wiki/Koeficijent_korelacije#/media/Datoteka:Pearson_Correlation_Coefficient_and_associated_scatterplots.png))

Kada između varijabli ne postoji veza govori se o nultoj ili nepostojećoj korelaciji. Za takve se dvije varijable kaže da su ne korelirane te imaju tendenciju mijenjanja bez ikakve međusobne veze. Na slici 3. prikazane su varijable s nultom korelacijom.



Slika 3.: Nekorelirane varijable (Izvor: [https://bs.wikipedia.org/wiki/Koeficijent\\_korelacije#/media/Datoteka:Pearson\\_Correlation\\_Coefficient\\_and\\_associated\\_scatterplots.png](https://bs.wikipedia.org/wiki/Koeficijent_korelacije#/media/Datoteka:Pearson_Correlation_Coefficient_and_associated_scatterplots.png))

Za korelaciju između dviju varijabli kaže se da je linearna ako promjena jedne jedinice u jednoj varijabli rezultira odgovarajućom (konstantnom) promjenom druge varijable u cijelom rasponu vrijednosti.

## 1.2. Jednostavni regresijski modeli

Kada je potrebno, ne samo ispitati jakost veze između varijabli, već i objasniti povezanost između varijabli koristi se regresijska analiza. Regresijski model je model u kojemu je prikazan odnos između jedne promatrane varijable (zavisna varijabla) te jedne ili više drugih varijabli (regresori, nezavisne varijable). Takvim se modelom nastoji numerički prikazati utjecaj nezavisne ili nezavisnih varijabli na zavisnu varijablu odnosno kako će promjene u kretanju nezavisne varijable utjecati na zavisnu varijablu.

Općenito, regresijska analiza se bavi opisivanjem i ocjenjivanjem odnosa između dane varijable i jedne ili više drugih varijabli. Drugim riječima, regresijska analiza je nastojanje da se objasni kretanje vrijednosti u jednoj varijabli uz pomoć vrijednosti drugih varijabli.<sup>2</sup>

Razlikuju se četiri glavna cilja regresijske analize:

1. kvantificiranje utjecaja jedne varijable na drugu varijablu,
2. prognoziranje budućih vrijednosti varijable u pitanju,
3. identifikacija varijable ili varijabli koje bi mogle imati utjecaja na zavisnu varijablu i
4. ispitivanje učinkovitosti određene varijable.

Najjednostavniji oblik stohastičkog odnosa između dvije varijable  $X$  i  $Y$  zove se *jednostavni linearni regresijski model*. Taj se model formalno izražava u obliku:

$$Y_i = \beta_0 + \beta_1 X_1 + e_i \quad (2)$$

gdje je:

$Y_i$  – zavisna varijabla

$X$  – nezavisna varijabla

$\beta_0, \beta_1$  – nepoznati koeficijenti ili parametri

$\beta_0$  – konstantni član, predstavlja odsječak na osi ordinate

$\beta_1$  – koeficijent nagiba ili regresijski koeficijent, označava vrijednost promjene  $Y$  za jediničnu promjenu  $X$ .

$e_i$  – slučajno odstupanje ili rezidual

$i$  –  $i$ -to opažanje

Model pokazuje prosječnu vrijednost  $Y_i$  za dani  $X_1$  i nepoznate parametre  $\beta_0$  populacije i dok indeks  $i$  pokazuje da je riječ o podacima vremenskoga presjeka. Model podrazumijeva da za svaku varijablu  $X$  postoji cijela distribucija vjerojatnosti za vrijednost varijable  $Y$  (Baldigara, 2023.). Neizvjesnost kod varijable  $Y$  se javlja zbog slučajnog odstupanja .

Model (2) sastavljen je od dva karakteristična dijela (Baldigara, 2023):

---

<sup>2</sup> Brooks, C. (2019): *Introductory Econometrics for Finance. 4th edn.*, Cambridge University Press. str. 38.

1. Deterministički dio modela – sustavni dio modela , koji izražava postuliranu teorijsku vezu danu ekonomskom teorijom pri kojoj je  $Y_i$  zavisna od  $X_i$ , ako su drugi čimbenici konstantni (*ceteris paribus* klauzula). Taj dio predstavlja sistematske varijacije  $Y$  u zavisnosti od promjene u  $X$ .
2. Stohastički dio modela ( $e_i$ ) – nesustavni dio modela koji predstavlja slučajne varijacije kojima se uzima u obzir djelovanje promjena ostalih varijabli koje su izostavljene iz modele.

### *Zavisna varijabla*

Varijabla na lijevoj strani jednadžbe naziva se zavisna varijabla te se uobičajeno označava sa „ $Y$ “. Drugi nazivi za zavisnu varijablu su regresand varijabla, endogena varijabla, output varijabla, i dr. Kretanje te varijable nastoji se objasniti uz pomoć nezavisnih varijabli.

Za zavisne se varijable kaže da su one koje su determinirane sustavom, a model se konstruira sa svrhom da ih objasni stoga je broj jednadžbi jednak broju zavisnih varijabli.<sup>3</sup>

### *Nezavisne varijable*

Nezavisna varijabla (regrosor, egzogena varijabla, input varijabla, objasnidbena varijabla) je varijabla za koju se očekuje da će objasniti kretanje zavisne varijable uz pretpostavku da su ostali utjecajni faktori ne promijenjeni. Nezavisnu varijablu u modelu označavamo s  $X$ .

Nezavisne se varijable odabiru s obzirom na poznavanje područja primjene, a isto vrijedi i za klasificiranje varijabli na nezavisne i zavisne.

### *Nepoznati koeficijenti ili parametri*

Regresijski koeficijent  $\beta_1$ , je vrijednost koja određuje nagib pravca regresije. Vrijednost regresijskog koeficijenta  $\beta_1$  kvantificira promjenu zavisne varijable u slučaju jedinične promjene nezavisne varijable, uz pretpostavku da su ostale varijable nepromijenjene.

Konstantni član  $\beta_0$ , predstavlja odsječak na osi  $y$  odnosno točku na ordinati kroz koju prolazi pravac regresije. Vrijednost  $\beta_0$  pokazuje vrijednost zavisne varijable u slučaju da nezavisna varijabla poprimi vrijednost 0.

---

<sup>3</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 8.

### Slučajno odstupanje

Slučajno odstupanje, mjera je odstupanja stvarnih od regresijskih vrijednosti.

Postoji nekoliko razloga nastajanja slučajnoga odstupanja:

- Regresijski model gotovo nikada ne sadrži sve čimbenike koji utječu na zavisnu varijablu te se zbog toga očekuju određena odstupanja. Osim odabranih regresora, na zavisnu varijablu  $Y$ , mogu utjecati i drugi čimbenici koji su izostavljeni iz modela. Razlozi tomu mogu ili nedovoljno poznavanje navedenih čimbenika ili nemogućnost prikupljanja podatka.
- U nekim je slučajevima poželjno izbjeći uključivanje prevelikoga broja nezavisnih varijabli u model te se stoga nastoji sažeti sveukupni broj odnosa u samo jednoj ili nekoliko najvažnijih varijabli. Kao rezultat sažimanja dobiju se dobre procijenjene vrijednosti zavisne varijable  $Y_i$ , s odstupanjima koja su obuhvaćena u slučajnim odstupanjima.
- Mogućnost pojave slučajnih odstupanja može biti rezultat pogrešne specifikacije modela. Na primjer, može se dogoditi da trenutna vrijednost nezavisne varijable  $X_1$  ne utječe na vrijednosti zavisne varijable  $Y_i$ , ali na vrijednosti nezavisne varijable u prethodnim razdobljima  $X_{t-1}$  imaju utjecaja na kretanje vrijednosti zavisne varijable.
- Smanjenja točnost u mjerenju nezavisnih varijabli također može također biti uzrok pojavljivanja slučajnih odstupanja.<sup>4</sup>

### 1.3. Višestruki regresijski model

Ako se u regresijskoj jednadžbi koriste dvije ili više nezavisne varijable riječ je o modelu višestruke regresije. Model višestruke regresije dan je izrazom:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \dots + \beta_j X_{ij} + e_i \quad (3)$$

gdje je:

$Y_i$ - zavisna varijabla

$X_{i1}, X_{i2}, X_{ij}$  – nezavisne varijable

$e_i$  – slučajno odstupanje

---

<sup>4</sup> Asteriou, D. and Hall, S. (2021): *Applied Econometrics. 4th edn.* Bloomsbury Publishing., str. 65.

$i$  –  $i$ -to opažanje

$\beta_0$  – konstantni član, odnosno odsječak na ordinati, a predstavlja prosječnu vrijednost  $Y_i$  kada su  $X_{i1}, X_{i2}, \dots, X_{ij}$  jednaki nuli

$\beta_1, \beta_2, \dots, \beta_j$  – parcijalni regresijski koeficijenti

Kao i kod jednostavne regresije, jednadžba se sastoji od sustavne ili determinističke komponente ( $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$ ) i nesustavne ili stohastičke komponente ( $e_i$ ) određenu čimbenicima koji su drukčiji od korištenih nezavisnih varijabli.

#### 1.4. Pretpostavke klasičnoga linearnog regresijskog modela

Standard prema kojemu se utvrđuju rezultati regresijske analize dan je postavkama klasičnog linearnog regresijskog modela (CLRM – *Classical Linear Regression Model*). Model je 1821. godine definirao C. F. Gauss. Ako pretpostavke nisu ispunjene, rezultati su netočni i obmanjujući. Pretpostavke Gaussovog klasičnog linearnog regresijskog modela su slijedeće:

1. *Korektna specifikacija i linearnost regresijskog modela*
2. *Sredina jednaka nuli*
3. *Odsutnost autokorelacije*
4. *Homoskedastičnost*
5. *Odsutnost multikolinearnosti*
6. *Nestohastičnost varijable  $X$*
7. *Normalnost slučajnog odstupanja*

*Korektna specifikacija i linearnost regresijskog modela*

Prva pretpostavka zahtijeva mogućnost predstavljanja zavisne varijable kao linearne funkcije određenog skupa nezavisnih varijabli kojem se pribrojavaju i slučajna odstupanja. Linearnost modela u parametrima omogućava dobivanje linearnih ocjena (Baldigara, 2023.). Pretpostavka linearnosti nije tako restriktivna kao što se na prvi pogled može činiti, jer zavisna varijabla i regresori mogu biti transformacije varijabli.

*Sredina jednaka nuli*

Očekivana vrijednost člana prognostičke pogreške je nula. To se može izraziti kao  $E(e_i) = 0$ . S obzirom da slučajna odstupanja nemaju utjecaja na zavisnu varijablu jer ona predstavljaju

čimbenike koji su izostavljeni iz modela, njegovi se pozitivni i negativni utjecaji međusobno negiraju te je njihov ukupan utjecaj naposljetku jednak nuli.<sup>5</sup>

#### *Nestohastičnost regresorskih varijabli*

Regresori  $X_1$  su nestohastični te fiksiran u ponovljenim uzorcima. Ova pretpostavka zahtijeva da vrijednosti regresora nisu određene nekim slučajnim mehanizmom. Nadalje, pretpostavlja mogućnost ponavljanja uzoraka s istim vrijednostima nezavisnih varijabli. Navedeno implicira da je  $cov(X_1, ) = 0$  za sve  $s, i = 1, 2, \dots, n$  odnosno da  $X_1$  i nisu korelirani.

#### *Homoskedastičnost*

Homoskedastičnost zahtijeva konstantnu varijancu odstupanja, tako da je  $var(e_i) = \sigma^2 =$  konstanta za sve  $i$ . To znači da odstupanja imaju svojstvo jednake raspršenosti odnosno homoskedastičnosti. Ako pretpostavka homoskedastičnosti nije ispunjena te odstupanja nisu konstantna za sve varijable tada je riječ o heteroskedastičnosti. To znači da varijanca pogreške zavisio o opažanju o kojemu je riječ. Reprerentativnost modela ovisi o tome kako su blizu distribuirane empirijske vrijednosti varijable  $Y$  oko njihovih sredina, i to je suština regresije (Baldigara, 2023.). Kada je riječ o modelima koji se ocjenjuju na osnovi podataka vremenskog prosjeka, gdje su velike razlike između najvećih i najmanjih vrijednosti opažanja tada se najčešće koristi heteroskedastičnost.

Za bilo koju vrijednost nezavisne varijable  $X$ , varijanca reziduala oko regresijske linije u populaciji pretpostavlja se da je konstantna. Kada je zadovoljena pretpostavka o konstantnoj varijanci reziduala bez obzira na vrijednost  $X$ , ovaj se uvjet naziva homoskedastičnost. S druge strane, ako se varijanca mijenja s obzirom na kretanje varijable  $X$  tada se to naziva heteroskedastičnost.

Kada je prisutna heteroskedastičnost, procijenjeni regresijski koeficijenti ostaju nepristrane, ali će standardne pogreške, a time i testovi značajnosti i intervali pouzdanosti biti netočni. U praksi će testovi značajnosti i intervali pouzdanosti biti vrlo blizu točnih vrijednosti osim u slučaju kada je stupanj nekonstantne varijance velik.<sup>6</sup>

---

<sup>5</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 51.

<sup>6</sup> Cohen, J. et al. (2013): *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd edn. Taylor and Francis., str. 33.

Testova heteroskedastičnosti ima također mnogo, a uobičajeno se koriste White test, Park test, Goldfeld-Quandt test, Breusch-Pagan test, i dr. U empirijskome dijelu diplomskoga rada koristiti će se White test.

Test veličina White testa temelji se na procjeni pomoćne regresijske jednadžbe. Neka je dan višestruki regresijski model oblika:

$$Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u_i \quad (4)$$

Kako bi dobili pomoćne jednadžbe potrebno je metodom najmanjih kvadrata izračunati rezidualne polaznog modela  $\hat{\varepsilon}_i^2$ . Zatim se procjenjuju pomoćne regresijske jednadžbe:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i}^2 X_{3i}^2 + \varepsilon_i \quad (5)$$

Kod White testa nulatom se hipotezom pretpostavlja da su varijance nepromjenjive odnosno homoskedastične:  $H_0: \alpha_1 = \alpha_2 = \alpha_3 \dots \alpha_n$ . Izraz za White-ov dan je:

$$W = nR^2 \quad (6)$$

Gdje je  $n$  veličina uzorka, a  $R^2$  je koeficijent determinacije pomoćne regresijske jednadžbe. White test također pripada  $\chi^2$  distribuciji s  $r$  stupnjeva slobode, gdje je  $r$  broj regresora pomoćne regresijske jednadžbe. Ukoliko je vrijednost izračunatoga White testa, uz razinu signifikantnosti  $\alpha$ , veća od tablične vrijednosti  $\chi^2$  distribucije tada se odbacuje nulta hipoteza homoskedastičnosti.

#### *Odsustvo autokorelacije*

Serijska neovisnost podrazumijeva da su svi članovi slučajnih odstupanja neovisno raspoređeni odnosno da nisu međusobno povezani, tako da je  $cov(e_i, e_s) = E(e_i - Ee_i)(e_s - Ee_s) = E(e_i e_s) = 0$  za sve  $i \neq s$ .

Autokorelacija označava korelaciju sa „samim sobom“ (*auto-*) odnosno korelaciju varijable s njom samom.

Autokorelacija mjeri stupanj korelacije između vrijednosti dvaju promatranja iste varijable. Ona pokazuje u kojoj je mjeri vrijednost u prethodnom promatranju povezana s vrijednošću



promatranja u pitanju. Autokorelacija se naziva i serijska korelacija te pomaže otkriti ponavljajuće periodične uzorke među podacima koji mogu pomoći pri daljnjoj analizi.<sup>7</sup>

Autokorelacija se uobičajeno javlja kod ocjenjivanja modela na osnovi podataka vremenskih nizova te predstavlja očekivanu vrijednost između dviju različitih komponenti varijable  $u$  koja je različita od 0. Rezultat autokorelacije može biti broj između -1 i 1, gdje vrijednosti od -1 do 0 označavaju negativnu autokorelaciju, a vrijednosti od 0 do 1 označavaju pozitivnu autokorelaciju. Kod negativne autokorelacije pozitivna odstupanja slijede negativna, zatim ponovo pozitivna, dok kod pozitivne autokorelacije odstupanja imaju najčešće isti predznak.

U praksi se najčešće susreće autokorelacija prvog reda koja se može izraziti autoregresijskom funkcijom AR(1)

$$u_t = \rho u_{t-1} + v_t \tag{7}$$

gdje je:

$u_t$  – odstupanje u razdoblju  $t$

$u_{t-1}$  – odstupanje u prethodnom razdoblju

$\rho$  – jednostavni korelacijski koeficijenti između  $u_t$  i  $u_{t-1}$ ,  $|\rho| < 1$

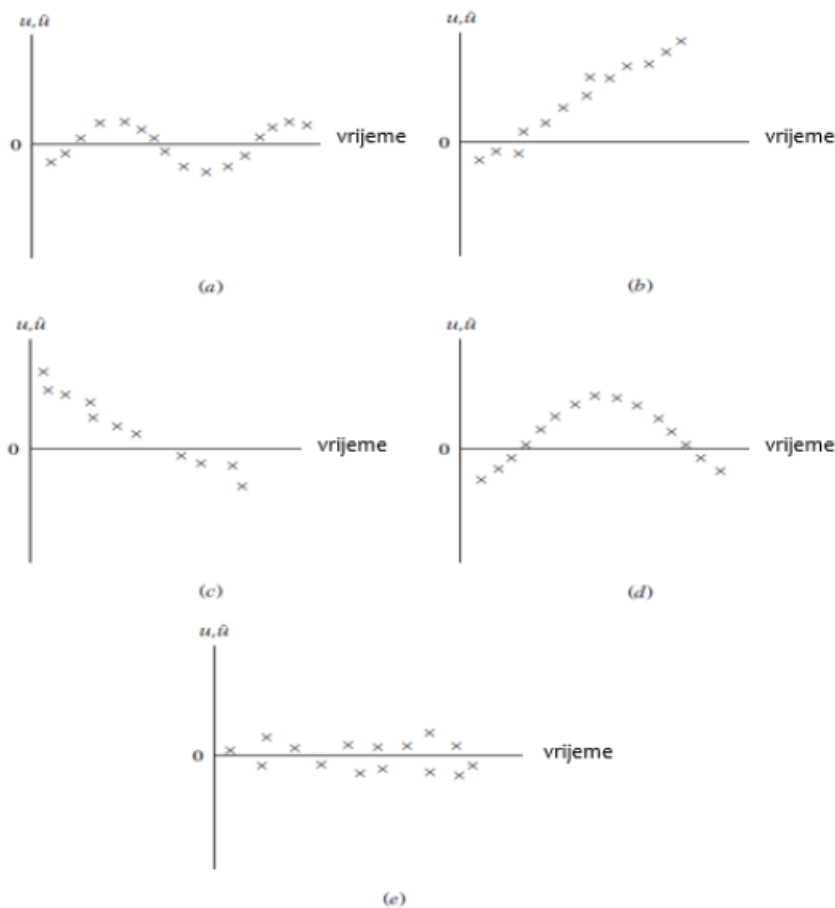
$v_t$  – normalno distribuirana nezavisna odstupanja koja su u skladu s klasičnim pretpostavkama, tj.  $v_t \sim N(0, \sigma^2)$ <sup>8</sup>

Ukoliko nedostaju neke vrijednosti opažanja među analiziranim podacima te ako se model procjeni metodom najmanjih kvadrata, autokorelaciju nije moguće ispitati matematičkom formulom. No, ona se može ispitati i grafički. Autokorelirana odstupanja uglavnom će pratiti određenu zajedničku liniju na prikazu odnosno kretati će se prema određenom uzorku.

---

<sup>7</sup> <https://corporatefinanceinstitute.com/resources/data-science/autocorrelation/>, Pristupio: 30.05.2023., 22:12

<sup>8</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 59.



Slika 4. (Izvor: Miles, J. and Shevlin, M. (2000) *Applying Regression and Correlation. 1st edn.* SAGE Publications. str. 19.)

Slika 4. a, b, c i d prikazuje prisutnost autokorelacije, dok je na slici e prikazana situacija nepostojanja autokorelacije među rezidualima.

Autokorelacija se najčešće javlja iz dva razloga:

1. Prisutnost uzorka među samim podacima (Prava korelacija)
2. Specifikacijska pogreška (neprava korelacija).

Kada je među podacima prisutan uzorak, na osnovi kojeg se model ocjenjuje, to nazivamo pravom autokorelacijom. Nepravom autokorelacijom nazivamo specifikacijske pogreške poput izostavljanja značajne varijable ili odabir pogrešne funkcijske veze.

#### *Normalnost reziduala*

Normalnost reziduala pretpostavlja se da su odstupanja  $e_1, e_2, \dots, e_n$  neovisna i jednako normalno raspoređeni, sa srednjom nulom i zajedničkom varijancom  $\sigma^2$ .

Normalnost se reziduala može ispitati na više načina, no u nastavku će biti opisan Jarque-Bera (JB) test. Test koristi koeficijent asimetrije i koeficijent zaobljenosti reziduala procijenjenih metodom najmanjih kvadrata kako bi se utvrdilo odstupaju li procijenjene veličine značajno od vrijednosti tih mjera za normalnu distribuciju. Kako bi distribucija bila normalna koeficijent asimetrije treba biti 0, a koeficijent zaobljenosti 3. Jarque-Bera test veličina može se izraziti sljedećom formulom:

$$JB = \left[ \frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24} \right] \quad (8)$$

Pod pretpostavkom normalnosti, test veličina pripada  $\chi^2$  distribuciji s 2 stupnja slobode. Nulta hipoteza glasi: „greška je normalno distribuirana“. Nulta se hipoteza odbacuje kao lažna ako je  $JB > \chi_{\alpha}^2(2)$  odnosno ako je empirijska razina signifikantnosti  $p$  manja od teorijske razine signifikantnosti.<sup>9</sup> Uobičajena teorijska razina signifikantnosti koja se koristi je  $\alpha(0,05)$  što znači da je razina pouzdanosti 95%.

#### *Odsustvo multikolinearnosti*

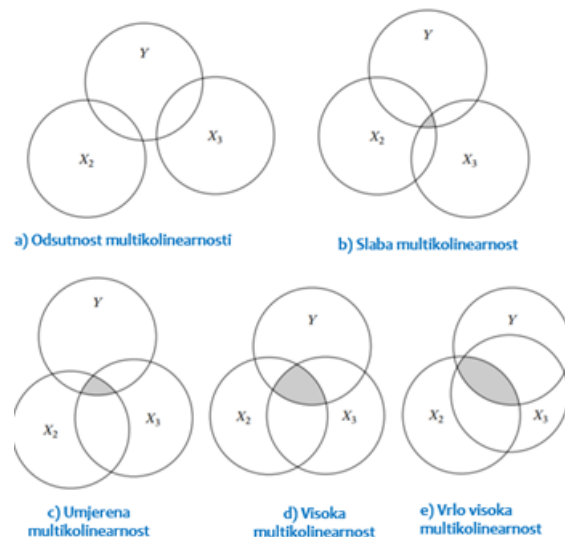
Nezavisna varijabla ne smije biti korelirana s nekom drugom nezavisnom varijablom niti s bilo kojom drugom linearnom kombinacijom nezavisnih varijabli. Kada postoji korelacija između nezavisnih varijabli riječ je o problemu multikolinearnosti.

#### *Kolinearnost*

Regresijski model procijenjen metodom najmanjih kvadrata, mora zadovoljiti i pretpostavku o nepostojanju multikolinearnosti između nezavisnih varijabli. Među nezavisnim varijablama ne smije postojati savršena kolinearnost. Pojam „kolinearnost“ upućuje da se radi o zajedničkoj povezanosti varijabli (*co-*) oko određenog pravca.

---

<sup>9</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 68.



Slika 5. Multikolinearnost, Izvor: Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 54.

Na slici je prikazan problem multikolinearnosti između dvije nezavisne varijable. Pretpostavimo da se kretanje varijable Y opisuje varijablom  $X_1$  koja objašnjava određeni dio promjena u varijabli Y. Ukoliko u tom slučaju postoji i druga nezavisna varijabla  $X_2$  koja je u korelaciji s  $X_1$  tada, najjednostavnije rečeno,  $X_2$  objašnjava gotovo isti dio promjena u varijabli Y kao i  $X_1$ .<sup>10</sup>

Problem multikolinearnosti analizira se uz pomoć inflacijskog faktora varijance (*Variance Inflation Factor-VIF*). VIF se računa sljedećom formulom:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, k. \quad (9)$$

gdje je  $R_j^2$  – koeficijent determinacije u modelu višestruke linearne regresije u kojemu je  $j$ -ta regresorska varijabla zavisna.<sup>11</sup> Idealan rezultat bio bi kada bi vrijednost  $R_j^2$  bila nula, tada bi VIF iznosio 1 što bi značilo da nema kolinearnosti među nezavisnim varijablama. No to je rijetko slučaj, kako vrijednost  $R_j^2$  raste, tako raste i vrijednost VIF-a. Ozbiljan problem multikolinearnosti se javlja kada je  $R_j^2=0,8$  odnosno  $VIF=5$ .

<sup>10</sup> Hoffmann, J. (2021) *Linear Regression Models. 1st edn.* CRC Press., str. 18.

<sup>11</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 57.

## RAMSEY RESET

Jedna od početnih pretpostavki regresijskih modela je linearnost odnosno da je veza između  $x$  i  $y$  linearna u parametrima koja se može prikazati ravnom linijom. Doduše, ta pretpostavka nije uvijek podržana. Da li bi model trebao biti linearan može se ispitati uz pomoć Ramsey RESET testa (Ramsey 1969.). To je test pogrešne specifikacije regresijske jednadžbe (RESET-Regression Equation Specification Error Test) dizajniran je za otkrivanje pogrešno specificiranog modela zbog izostavljanja važnih eksplanatornih varijabli ili netočnog izbora funkcionalnog oblika.

Ramsey-ev RESET se provodi na način da se odrede pomoćne regresijske jednadžbe. Pomoćne jednadžbe sadrže nove zavisne i nezavisne varijable čije su vrijednosti podignute na višu potenciju.

$$\hat{Y}_t^2 = (\hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \dots + \hat{\beta}_n X_{2n})^2 \quad (10)$$

Potencirane vrijednosti imaju tendenciju otkrivanja nelinearnih odnosa između nezavisne i zavisne varijable.

Rezultati testa se ocjenjuju uz pomoć  $\chi^2$  distribucije za  $(k-1)$  stupnjeva slobode. Ukoliko je empirijska vrijednost testa veća od kritične vrijednosti  $\chi^2$  odbacuje se nulta hipoteza koja kaže da je model točno specificiran. Kao i kod ostalih testova, rezultati se mogu ocijeniti i uz pomoć empirijske  $p$ -vrijednosti testa. Ako je izračunata  $p$ -vrijednost testa manja od vrijednosti  $\alpha$  na zadanoj razini signifikantnosti, tada se također odbacuje nulta hipoteza što je nepoželjan rezultat za daljnju analizu.

Treba napomenuti, kako Ramsey-ev RESET pokazuje samo da li je model točno specificiran ili nije. Ne pokazuje koji je bolji oblik specifikacije modela.

### 1.5. Procjena i interpretacija regresijskih parametara

Određivanje pravca regresije započinje određivanjem vrijednosti regresijskih koeficijenata. Dobivene vrijednosti, osim što će odrediti pravac regresije, objasniti će kretanje zavisne varijable pod uvjetom da su regresijski koeficijenti značajni te ako je model reprezentativan.

Postoji više metoda procjene parametara. Neke od njih su:

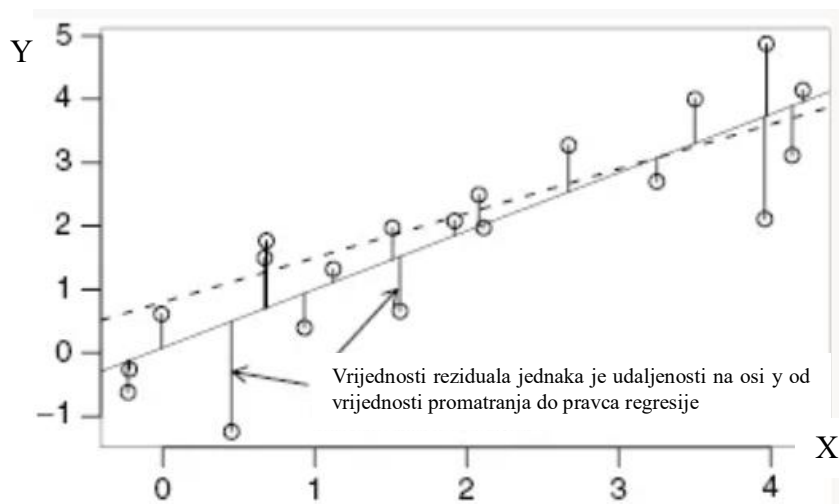
- Metoda najmanjih kvadrata (*Method of Ordinary Least Squares Estimators*, OLS)
- Najbolje linearno nepristrano ocjenjivanje (*Best Linear Unbiased Estimator*, BLUE)

- Metoda maksimalne vjerodostojnosti (*Method of Maximum Likelihood Estimator, MLE*)

U nastavku biti će opisana metoda najmanjih kvadrata s obzirom da je to najčešće korištena metoda u regresijskoj analizi, te će ista biti korištena i u empirijskom dijelu diplomskoga rada.

Za procjenu parametra kod jednostavne i višestruke linearne regresije najčešće se koristi metoda najmanjih kvadrata. Cilj procjene parametra jest odrediti jednadžbu pravca  $\hat{Y}$  koji će se najbolje prilagoditi prikupljenim podacima.

Regresijski koeficijenti su vrijednosti koje minimiziraju funkciju na način da tvore pravac regresije koji će imati minimalna prosječna odstupanja od zadanih vrijednosti.<sup>12</sup>



Slika 6., Izvor: Weisberg, S. (2013) *Applied Linear Regression. 4th edn, str:8.*)

Ovu je metodu postavio Carl Friedrich Gauss, početkom 19. st.. Osnovni cilj metode jest određivanje jednadžbe pravca  $\hat{Y}$  koja će se najbolje prilagoditi stvarnim podacima. Ako je zadan regresijski model s dvije varijable:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{11}$$

Regresijsku funkciju populacije nije moguće odrediti izravno, stoga se ocjenjuje temeljem regresijske funkcije uzorka:

<sup>12</sup> Weisberg, S. (2013): *Applied Linear Regression. 4th edn.*, Wiley. str. 8.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (12)$$

Koja se može zapisati i kao:

$$e_i = \text{stvarni } Y_i - \text{ocijenjeni } Y_i = Y_i - \hat{Y}_i \quad (13)$$

Zbog statističkih se razloga ne minimizira  $\sum e_i$  već  $\sum e_i^2$ , te se dobije:

$$\min \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad (14)$$

Za zadane vrijednosti X i Y suma kvadrata reziduala je funkcija parametra  $\hat{\beta}_0$  i  $\hat{\beta}_1$ . Različite vrijednosti regresijskih koeficijenata rezultirati će različitim vrijednostima slučajnog odstupanja. Shodno tome, dobiju se i različite vrijednosti sume kvadrata slučajnih odstupanja.

Regresijski koeficijenti predstavljaju vrijednost s kojima suma kvadrata slučajnih odstupanja ima najmanju moguću vrijednost.

Parametri regresije koji minimiziraju sumu kvadrata reziduala računaju se tzv. sustavom normalnih jednadžbi koje su prikazane u nastavku:

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \quad (15)$$

$$\sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \quad (16)$$

Rješavanjem ovoga sustava dolazi se do formule za izračunavanje regresijskih koeficijenata koje se najjednostavnije mogu zapisati na sljedeći način:

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} \quad (17)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (18)$$

Regresijski pravac dobiven metodom najmanjih kvadrata prolazi kroz srednje vrijednosti varijable X i Y te se može zapisati kao:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \quad (19)$$

gdje je:

$\bar{X}, \bar{Y}$  – jednostavne aritmetičke sredine od X i Y

$x_i - (X_i - \bar{X})$

$y_i - (Y_i - \bar{Y})$

*Interpretacija procijenjenih parametra funkcije*

Konstantni član  $\beta_0$  pokazuje prosječnu vrijednost zavisne varijable, ako je vrijednost nezavisne varijable 0. Dok s druge strane, vrijednost  $\beta_1$  kvantificira promjenu zavisne varijable za jediničnu promjenu vrijednosti nezavisne varijable.

Dobivene vrijednosti regresijskih parametara koriste se za opisivanje kretanja zavisne varijable, uz pretpostavku njihove značajnosti.

## 1.6. Regresijski modeli s kvalitativnim regresorskim varijablama

Kvalitativne se varijable, bilo da su određene kao nezavisne ili zavisne, uključuju u model pomoću *dummy* (indikator, binarnih) varijabli. *Dummy* varijable se koriste za prikazivanje varijabli koje se ne mogu brojčano iskazati. U tome slučaju raspon vrijednosti varijable je veoma ograničen, preciznije, varijabla može poprimiti samo dvije vrijednosti. Upravo tako funkcioniraju lažne varijable, svoju svrhu ostvaruju sa vrijednostima nula i jedan.<sup>13</sup>

Uobičajeno je da se s 0 označava odsustvo nekog svojstva, a s 1 prisustvo svojstva, osobine ili druge kategorije koje se ne mogu brojčano prikazati. Koeficijenti uz *dummy* varijable predstavljaju diferencijalne koeficijente, te ukazuju koliko se vrijednost kategorije koja poprima vrijednost 1 *dummy* varijable razlikuje od konstantnoga člana bazne kategorije.

Postoje tri vrste *dummy* varijabli:

1. *dummy* varijable konstantnog člana,
2. *dummy* varijable za promjene u nagibu i

---

<sup>13</sup> Zax, J. (2011) *Introductory Econometrics. 1st edn.* Stanford University Press., str. 145.



### 3. sezonske *dummy* varijable.

U diplomskome radu detaljnije se razmatraju sezonske *dummy* varijable.

Često podaci u svom kretanju kroz prošla razdoblja pokazuju određene oscilacije. Uzroci oscilacija mogu biti različiti, kao što su trajanje turističke sezone, trajanje skijaške sezone, povećana prodaja u vrijeme blagdana i sl. Kada postoje periodične oscilacije unutar jedne godine među podacima to označava prisutnost sezonalnosti. Kako bi podaci bili adekvatniji za daljnju analizu, poželjno je otkloniti komponentu sezonalnosti koja je prisutna među podacima. To se postiže uvođenjem, kao regresora, sezonskih *dummy* varijabli. Broj *dummy* varijabli manji je za jedan od broja perioda tj. ako se podaci mjere na mjesečnoj razini tada je broj *dummy* varijabli 11. Regresijska jednadžba za mjesečne podatke glasi:<sup>14</sup>

$$Y = \beta_1 X_1 + \gamma_2 D_2 + \gamma_3 D_3 + \gamma_4 D_4 + \dots + \gamma_{12} D_{12} + u_i \quad (20)$$

$$D_i = \begin{cases} 1 & \text{za } i - \text{ti mjesec} \\ 0 & \text{za ostale mjesece} \end{cases}, i = 1, 2, 3, \dots, 12. \quad (21)$$

Parametri *dummy* varijable procjenjuju se metodom najmanjih kvadrata, a njihova se značajnost i procijenjeni model testiraju na isti način kao i kod linearne regresije s kvantitativnim regresorima. Prilikom uvođenja *dummy* varijabli uz model koji posjeduje konstantni član, diferencijalni koeficijenti uz *dummy* varijable označavaju kretanje zavisne varijable za dani period u odnosu na bazni period.

Kod uvođenja *dummy* varijabli u regresijski model bez konstantnog člana vrijednosti diferencijalnih parametara za svaku *dummy* varijablu pokazuju prosječnu vrijednost zavisne varijable u određenom periodu.

## 1.7. Nelinearni regresijski modeli

Linearna regresijska analiza razmatra modele u kojima zavisna varijabla predstavlja linearnu funkciju nezavisnih varijabli. U empirijskim je istraživanjima, međutim veza između nezavisnih varijabli i zavisne varijable često nelinearna. U tom slučaju jedinična promjena u

---

<sup>14</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 148.

nezavisnoj varijabli ne rezultira konstantnom apsolutnom promjenom zavisne varijable, već je učinak različit s obzirom na različite vrijednosti nezavisne varijable.

Međutim, većinu nelinearnih modela moguće je linearizirati korištenjem transformacije varijabli te je u konačnici moguće koristiti linearne modele i uobičajene metode procjene parametara (pomoću metode najmanjih kvadrata).

Jedan od načina na koji se podaci mogu transformirati, kako bi bili pogodni za daljnju analizu i modeliranje, jest logaritmiranje.

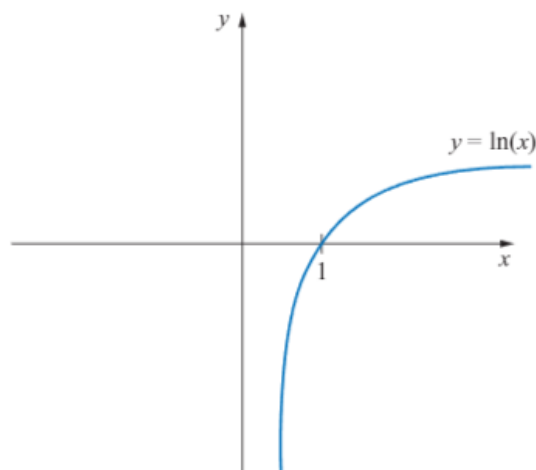
Opći izraz logaritma zapisuje se na sljedeći način:

$$\text{Ako je } a^b = c, \text{ tada je } \log_a c = b \quad (22)$$

Za potrebe analize i modeliranja podataka adekvatniji su prirodni logaritmi odnosno logaritmi na bazi vrijednosti  $e$ . Takvi se još logaritmi nazivaju i *Napierovi logaritmi* koji su dobili ime po Johnu Napier-u, škotskom matematičaru i fizičaru. Korištenje prirodnog logaritma je inverzna operacija eksponenciranja, stoga se često eksponencijalna funkcija naziva antilogaritam. Prirodni logaritmi za vrijednost manju od 1 su negativnog predznaka, a logaritmiranje s nulom i negativnim vrijednostima nije moguće. Prirodni algoritmi te njihovo pretvaranje dano je sljedećim izrazom:

$$\ln(e^x) = e^{\ln(x)} = x \quad (23)$$

Logaritamske funkcije posjeduju određeno svojstvo koje govori da se vrijednost varijable  $y$  mijenja sporije u odnosu na varijablu  $x$ . Krivulja logaritamske funkcije prikazana je na slici 7.



Slika 7.: Logaritamska funkcija ( Izvor: <https://ereader.perlego.com/1/book/2957643/50>)

Postoje više razloga logaritmiranja vrijednosti varijabli prilikom analize i modeliranja:<sup>15</sup>

1. Postavljanje nove ljestvice vrijednosti (npr. vrijednosti s određenom mjernom jedinicom logaritmiranjem se transformiraju u brojeve bez mjerne jedinice) koja će podatke „približiti“ središnjem pravcu distribucije vrijednosti opažanja. Na taj se način savladava problem heteroskedastičnosti.

2. Logaritmiranjem se kod pozitivno iskrivljenih slučajnih odstupanja postiže normalna distribucija.

3. Logaritmi se mogu koristiti za postizanje linearnosti između varijabli.

Nelinearni regresijski modeli imaju mnogo sličnosti s modelom linearne regresije. Nelinearni modeli mogu također biti jednostavni i višestruki. Ako se uspoređuje odnos između zavisne varijable i jedne nezavisne varijable tada je riječ o jednostavnoj nelinearnoj regresiji, a ukoliko jednadžba podrazumijeva dvije ili više nezavisne varijable tada se radi o višestrukoj nelinearnoj regresiji.

Općenito se razlikuju tri oblika nelinearnih regresijskih modela:<sup>16</sup>

1. Dvostruki logaritamski model (log-log model)
2. Logaritamsko linearni model (*log-lin* model)
3. Linearno logaritamski model (*lin-log* model).

U dvostruko logaritamskom modelu logaritmirane su zavisna i nezavisna varijabla. Kod polu-logaritamskih modela logitmirana se samo jedna varijabla. Kada se koristi log-lin model logaritmirana je samo zavisna varijabla, a kada je riječ o lin-log modelu logaritmirana je samo nezavisna varijabla. U tablici 1. su prikazani osnovni oblici polaznih nelinearnih regresijskim modela i njihovi linearizirani oblici.

---

<sup>15</sup> Brooks, C. (2019): *Introductory Econometrics for Finance. 4th edn.*, Cambridge University Press., str. 119.

<sup>16</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 95.

Nelinearni regresijski model	Linearizirani regresijski model	Interpretacija parametara
Model potencijske regresije $Y = \beta_0 \cdot X_i^{\beta_1}$	Dvostruko logaritamski model $\ln Y_i = \ln \beta_0 + \beta_1 \ln X_i$	$\% \Delta \hat{Y} \approx \hat{\beta}_1 \% \Delta X$
Model eksponencijalne regresije $Y_i = \beta_0 \cdot \beta_1^{X_i}$	Logaritamsko linearni model $\ln Y_i = \ln \beta_0 + \ln \beta_1 X_i$	$\% \Delta \hat{Y} \approx (\hat{\beta}_1 \cdot 100) \Delta X$
Model jednostavne logaritamske regresije $e^{Y_i} = e^{\beta_0} \cdot X_i^{\beta_1}$	Linearno logaritamski model $Y_i = \beta_0 + \beta_1 \ln X_i$	$\Delta \hat{Y} \approx (\hat{\beta}_1 / 100) \% \Delta X$

Tablica 1: Nelinearni i linearizirani regresijski modeli (Izvor: Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 96.)

U diplomskome radu detaljnije će biti opisan dvostruko logaritamski (log-log) model. Kako bi se dvostruko logaritamski model mogao koristiti potrebno je da su sve vrijednosti Y i X pozitivne te veće od 1 kako bi rezultat logaritma bio također pozitivan.<sup>17</sup>

Kod dvostruko logaritamskoga modela regresijski se koeficijenti također mogu procijeniti metodom najmanjih kvadrata, a njihovo se testiranje provodi kao i kod linearnoga regresijskoga modela. Osim toga, poželjno je da se i kod log-log modela ostvare pretpostavke koje su dane za linearni model kao što su normalnost reziduala, linearnost, homoskedastičnost, odsustvo multikolinearnosti, odsustvo korelacije, serijska neovisnost, prisutno odstupanje i nestohastičnost.

Razlika dvostruko logaritamskog i linearnog modela javlja se kod interpretacije vrijednosti regresijskih koeficijenata. Kod linearne regresije, regresijski koeficijent uz nezavisne varijable tumače se kao vrijednost promjene zavisne varijable za jediničnu promjenu nezavisne varijable. S druge strane, kod dvostruko logaritamskog modela regresijski se koeficijent uz varijablu X tumači kao postotna promjena u varijabli Y ako se varijabla X poveća za 1%.

<sup>17</sup> Hill, C., Griffiths, W. and Lim, G. (2018): *Principles of Econometrics. 5th edn.*, Wiley., str. 176.

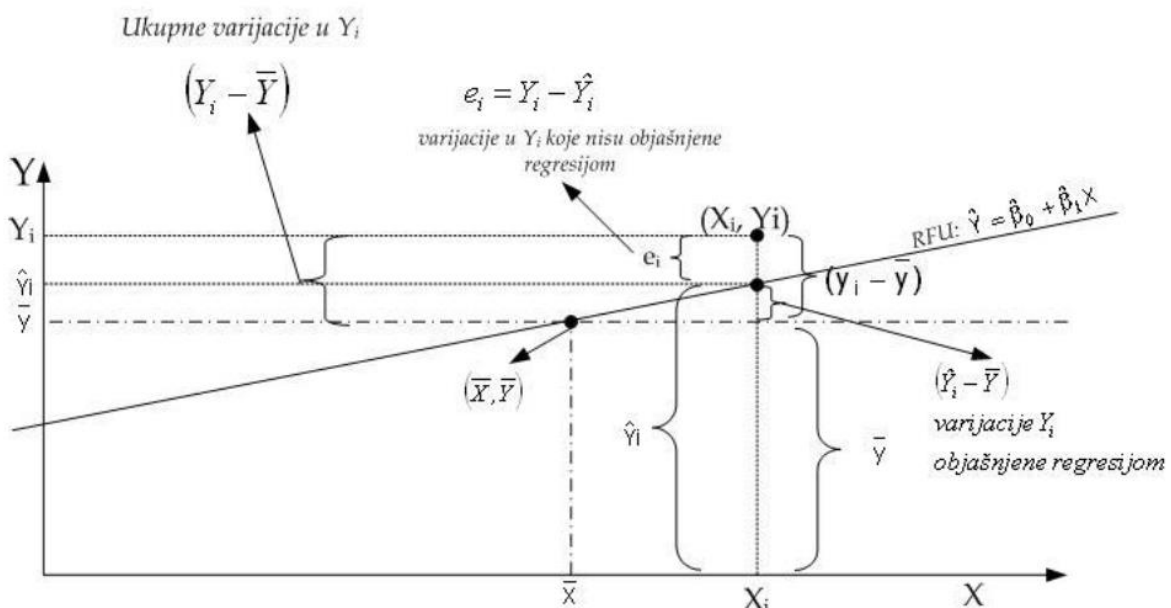
## 2. Analiza valjanosti regresijskog modela

Nakon ocjenjivanja ekonometrijskog modela pristupa se vrednovanju dobivenih ocjena parametara sa stajališta ekonomskih, statističkih i ekonometrijskih kriterija.

### 2.1. Analiza varijance

U ekonometriji se za mjerenje prilagođenosti procijenjenih vrijednosti stvarnim vrijednostima koriste odstupanja empirijskih podataka od njihove aritmetičke sredine. Osnova za definiranje statističkih mjera prilagođenosti su upravo ta odstupanja.

Uspješnost modela ovisi o rezidualnom odstupanju  $e_i = Y_i - \hat{Y}_i$  odnosno o razlici stvarnih vrijednosti zavisne varijable i regresijskih vrijednosti varijable.



Slika 8: Objasnjena i neobjasnjena komponenta varijable  $Y_i$  (Izvor: Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 30.)

Na slici vidljivo je da se odstupanja empirijskih vrijednosti zavisne varijable  $Y_i$  od aritmetičke sredine  $\bar{Y}$  može podijeliti na:

1. Odstupanje koje je protumačeno modelom i
2. Odstupanje koje nije protumačeno modelom.

Odstupanjima koja su protumačena modelom smatraju se odstupanja regresijske vrijednosti zavisne varijable od aritmetičke sredine, a ne protumačena odstupanja predstavljaju odstupanje empirijskih vrijednosti zavisne varijable od regresijskih vrijednosti zavisne varijable.

Odstupanje empirijskih podataka od njihove sredine odnosno *ukupna odstupanja* dana su izrazom:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \quad (24)$$

Ako se za svako opažanje kvadriraju i zbroje obje strane jednadžbe dobije se sažeta mjera za sva opažanja koja se naziva varijacija:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n e_i^2 + 2 \sum_{i=1}^n (Y_i - \bar{Y}) * e_i \quad (25)$$

Ukoliko je zadnji pribrojnik na desnoj strani jednadžbe 0 dobije se jednadžba analize varijance:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n e_i^2 \quad (26)$$

Jednadžbom analize varijance se računa ukupna suma kvadrata (TSS) čiji je rezultat jednak zbroju objašnjenje sume kvadrata (ESS) i neobjašnjene sume kvadrata (RSS).

TSS (*Total Sum of Squares*) ili ukupna varijacija je mjera ukupne varijacije zavisne varijable u odnosu na središnju vrijednost.

ESS (*Estimated Sumo of Squares*) ili objašnjena varijacija predstavlja dio ukupnih varijacija zavisne varijable koji je objašnjen varijacijama regresorske varijable.

RSS (*Residual Sumo of Squares*) ili rezidualna suma kvadrata predstavlja dio ukupnih varijacija zavisne varijable koji je objašnjen varijacijama parametara.

Varijacija je suma kvadrata odstupanja varijable od njene središnje vrijednosti, a varijanca je suma podijeljena s odgovarajućim stupnjem slobode. Proces računanja varijanci naziva se analiza varijance, a uobičajeno se prikazuje tablično (ANOVA tablica – *Analysis of Variance*)

izvor varijacije	Suma kvadrata	stupnjevi slobode	sredina kvadrata	F-vrijednost
objašnjena regresijom $\sum \hat{Y}_i^2$	$ESS = \sum (\hat{Y}_i - \bar{Y})^2$ ili $ESS = \beta_1^2 \sum x_i^2$	k	$\frac{ESS}{k} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{k}$ ili $\frac{ESS}{k} = \frac{\beta_1^2 \sum x_i^2}{k}$	$F = \frac{ESS/k}{RSS/(n-k-1)}$
neobjašnjena regresijom (rezidualna odstupanja) $\sum e_i^2$	$RSS = \sum e_i^2$	n-k-1	$\frac{RSS}{n-k-1} = \frac{\sum e_i^2}{n-k-1} = se^2$	
Ukupna $\sum Y_i^2$	$TSS = \sum (Y_i - \bar{Y})^2$	n-1		

Slika 9.: ANOVA tablica (izvor: Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 30.)

Na slici 9. je prikazan proces računanja F vrijednosti uz pomoć ranije navedenih suma kvadrata. Svaka se suma kvadrata dijeli s pripadajućim stupnjem slobode zatim se dobivenim vrijednostima računa F vrijednost.

ANOVA tablica je temelj za računanje pokazatelja adekvatnosti regresijskog modela kao što su procjena varijance regresije, procjena standardne pogreške regresije, procjena koeficijenata varijacije regresije, koeficijent determinacije i korigirani koeficijent determinacije.<sup>18</sup>

## 2.2. Pokazatelji pouzdanosti regresijskog modela

Koeficijent determinacije,  $R^2$ , opća je mjera reprezentativnosti procijenjenog regresijskog modela te ukazuje na udio ili postotak ukupne varijacije u zavisnoj varijabli Y.<sup>19</sup>

Koeficijent determinacije kvantificira postotak varijacija zavisne varijable koji se može objasniti postavljenim modelom odnosno odabranim regresorskim varijablama.

Koeficijent determinacije iznosi:

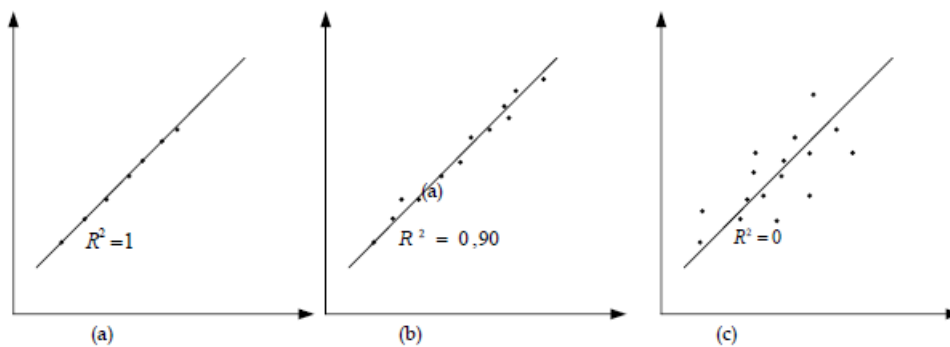
<sup>18</sup> Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 115.

<sup>19</sup> Gujarati, D. (2017): *Econometrics by Example. 2nd edn.*, Bloomsbury Publishing. str. 114.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2} \quad (27)$$

Koeficijent determinacije je najčešće korištena mjera prilagođenosti ocijenjene funkcije empirijskim podacima, koja mjeri proporciju modelom protumačenog dijela zbroja kvadrata u ukupnom broju kvadrata (Baldigara, 2023). Koeficijent determinacije može poprimiti vrijednost između nule i jedinice. Što je vrijednost koeficijenta determinacije bliža jedinici model je reprezentativniji.

Kada koeficijent determinacije iznosi 1 (slika 10. (a)) riječ je o savršenoj prilagodbi odnosno svaka varijacija nezavisne varijable objašnjena je procijenjenim modelom. S druge strane, ako je  $R^2=0$  (slika, (c)), tada to predstavlja nepostojanje veze između zavisne i nezavisne varijable.



Slika 10.: Koeficijent determinacije kao mjera uspješnosti prilagodbe linearne funkcije empirijskim podacima (Izvor: Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija, str. 34.)

Jedan nedostatak koeficijenta determinacije jest da je to funkcija koja raste s povećanjem broja regresora. To jest, ako se dodaju dodatne nezavisne varijable u model vrijednost koeficijenta determinacije raste. Stoga ponekad istraživači nastoje maksimizirati koeficijent determinacije dodajući dodatne objasnidbene varijable. Kako bi se izbjeglo takvo maksimiziranje  $R^2$  dobro je koristiti mjeru koja uzima u obzir uključen broj regresora u model. Takva mjera naziva se prilagođeni koeficijent determinacije te se označava s  $\overline{R^2}$  te se računa uz pomoć sljedeće formule:

$$\overline{R^2} = 1 - (1 - R^2) \frac{n - 1}{n - k} \quad (28)$$

Sposobnost modela da objasni kretanje zavisne varijable ovisiti će o vrijednosti slučajnih odstupanja tog modela. Što je manji udio rezidualnih odstupanja u ukupnim varijacijama, model



se bolje prilagođava empirijskim podacima. Mjera takve prilagođenosti dobije se iz procijenjene varijance regresije  $se^2(\hat{\sigma}^2)$  te se računa na sljedeći način:

$$se^2 = \frac{RSS}{n - k - 1} = \frac{\sum e_i^2}{n - k - 1} \quad (29)$$

Iz prethodne se jednadžbe računa mjera prilagođenosti modela stvarnim podacima koja se naziva standardnom pogreškom regresije:

$$se = \sqrt{se^2} = \sqrt{\frac{RSS}{n - k - 1}} = \sqrt{\frac{\sum e_i^2}{n - k - 1}} \quad (30)$$

Standardna pogreška regresija se interpretira kao prosječno odstupanje empirijskih vrijednosti od regresijskih, a može se računati i kao relativni pokazatelj odnosno koeficijent varijacije:

$$V = \frac{se}{\bar{Y}} \cdot 100 \quad (31)$$

Koeficijent varijacije od 20% ili više iskustvena je vrijednost prihvatljivosti ocijenjenog regresijskog modela (Baldigara, 2023).

### 2.3. Testiranje hipoteza u regresijskom modelu

Proces donošenja odluka o prihvaćanju ili ne prihvaćanju neke tvrdnje na temelju podatka iz slučajnog uzorka naziva se testiranje hipoteze. Svako testiranje započinje postavljanje nulte hipoteze ( $H_0$ ) te alternativne hipoteze ( $H_1$ ) koja proturječi nultoj hipotezi.

Ukoliko je  $H_0$  istinita tada je ispravna odluka prihvatiti  $H_0$ , a ako je  $H_0$  lažna tada je ispravno odbaciti nultu hipotezu. Iz navedenog se mogu prikazati i dvije vrste pogreška koje se javljaju prilikom odlučivanja:

1. Pogreška tipa I – odbacivanje istinite  $H_0$
2. Pogreška tipa II – prihvaćanje lažne  $H_0$ .

Pogreška tipa I prikazuje se vrijednošću  $\alpha$  koja se naziva teorijskom razinom signifikantnosti. Vjerojatnost da se prihvati lažna nulta hipoteza prikazuje se vrijednošću  $\beta$ , a s izrazom  $(\beta-1)$  definirana je vjerojatnost odbacivanja lažne  $H_0$  koja se naziva snagom statističkog testa. U nastavku diplomskoga rada detaljnije se razmatraju:

- Test značajnosti regresijskih parametara (pojedinačni t-test)
- Test značajnosti regresijskog modela (skupni F-test).

### 2.3.1. Testiranje značajnosti regresijskih parametara

Valjanost procijenjenog regresijskog modela ocjenjuje se testiranjem značajnosti objasnidbenih varijabli modela. Kod procijenjenog regresijskog modela metodom običnih kvadrata potrebno je testirati značajnost parametara uz nezavisnu varijablu. Drugim riječima, potrebno je li značajna prisutnost varijable  $X$  u modelu. Značajnost parametara se utvrđuje usporedbom vrijednosti testa veličina s odgovarajućim vrijednostima sampling-distribucije.

„Statistički značajan parametar znači zapravo da su podaci konzistentni s pretpostavkom o uzročnosti veze između zavisne i nezavisne varijable.“ (Baldigara, 2023)

Poznajući distribuciju naših procijenjenih koeficijenata, u mogućnosti smo provesti testiranje hipoteza *t-testom* kako bismo procijenili njihovu statističku značajnost. Općenito, treba slijediti sljedeće korake:<sup>20</sup>

1. Postavljanje nulte i alternativne hipoteze.

$$H_0: \beta = 0; H_1 \neq 0 \tag{32}$$

2. Izračun t-statistike izrazom:

$$t = (\hat{\beta} - \beta) / se\hat{\beta} \tag{33}$$

3. Iz t-tablice je potrebno pronaći kritičnu vrijednost za  $n-k-1$  stupnja slobode gdje je  $n$  broj promatranja, a  $k$  broj nezavisnih varijabli.
4. Ako je  $|t_{emp}| > |t_c|$  odbaciti nultu hipotezu.

Pravilo t-testa se može primijeniti koristeći i empirijsku  $p$  vrijednost kako bi se testirala značajnost regresijskih koeficijenata. Umjesto da se traži kritična vrijednost u statističkim

---

<sup>20</sup> Asteriou, D. and Hall, S. (2021): *Applied Econometrics. 4th edn.* Bloomsbury Publishing, str. 34.

tablicama kao kod t-testa, usporediti se može p-vrijednost testa s odabranim koeficijentom pouzdanosti. Primjerice, ako odaberemo da se analiza provodi uz pouzdanost od standardnih 95%, tada bi razina signifikantnosti iznosila  $\alpha(0,05)$ . U tome slučaju, ako je p-vrijednost manja od 0,05 tada se također odbacuje nulta hipoteza kao neistinita odnosno prihvaća se alternativna hipoteza koja nam govori da su regresijski koeficijenti značajni.<sup>21</sup>

*p-vrijednost* predstavlja vjerojatnost odbacivanja istinite nulte hipoteze, a računa se uz pomoć empirijskih podataka.

$$p - \text{vrijednost} = P(t_c(n - k - 1) \geq |t_{emp}|) \quad (34)$$

Što je niža *p-vrijednost* to je manja vjerojatnost odbacivanja istinite nulte hipoteze i obrnuto. Prilikom testiranja *p-vrijednosti*, postupak donošenja odluke provodi se na sljedeći način:

- $p - \text{vrijednost} > \alpha$  – nulta se hipoteza ( $H_0$ ) prihvaća
- $p - \text{vrijednost} < \alpha$  – nulta se hipoteza ( $H_0$ ) odbacuje.

### 2.3.2. Testiranje značajnosti regresijskog modela

Ocjena značajnosti regresijskog modela provodi se uz pomoć *F-testa*. Nulta hipoteza u navedenom testu tvrdi da regresorska varijabla X nije značajna te je suvišno uključivati ju u model. S druge strane, alternativna hipoteza tvrdi da je nezavisna varijabla X značajna odnosno objašnjava određeni dio varijacija u nezavisnoj varijabli Y.

Nulta se hipoteza može izraziti na sljedeći način:

$$H_0: R^2 = 0 \quad (35)$$

Takva se hipoteza testira tehnikom analize varijance:

$$F = \frac{ESS/k}{RSS/(n - k - 1)} = \frac{R^2/k}{(1 - R^2)(n - k - 1)} \sim F(k; n - k - 1) \quad (36)$$

Odluka se donosi na način da se usporedi rezultat empirijske F-vrijednosti izračunatu uz pomoć izraza s teorijskom vrijednosti F-distribucije za određenu razinu signifikantnosti i određen broj stupnjeva slobode. Nulta se hipoteza prihvaća ako je empirijski F-omjer manji od

<sup>21</sup> Hayashi, F. (2011): *Econometrics*, Princeton University Press., str. 177.

teorijske vrijednosti F-distribucije. Odluka o nultoj hipotezi se može donijeti i na temelju  $p$ -vrijednosti. Ako je  $p$ -vrijednost veća od razine signifikantnosti  $\alpha$  tada se također prihvaća  $H_0$ .

Kod testiranja značajnosti regresijskog modela, prihvaćanjem nulte hipoteze ( $H_0$ ) prihvaća se pretpostavka da nezavisne varijable nisu značajne u modelu te ih nema smisla uključivati u model. Ako je odbačena  $H_0$ , to znači da barem jedna objasnidbena varijabla pridonosi objašnjavanju kretanja zavisne varijable.

### 3. Prognoziranje – pojmovno određenje

Općenito govoreći, prognoza je input u aktivnost planiranja čiji je output splet ciljeva poslovnom sustavu (Branco & Pavlović, 1995.). U području poslovne ekonomije prognozirati znači donositi sudove o budućim stanjima promatrane pojave za koju su prikupljeni povijesni podaci.

Prognoziranje se smatra predviđanje nekih budućih događaja koji će se desiti uz određenu vjerojatnost na način da se analizira prošlost tih događaja. Prognoziranje se nastoji stvoriti nepristrana procjena koja se odnosi na buduće promjene neke varijable na temelju nekakvog prethodnog znanja i iskustva.

#### 3.1. Definicija i faze prognoziranja

Proces prognoziranja se može definirati kao procjena kretanja pojave kroz buduće vremenske intervale koja je generirana uz pomoć tekućih podataka o potražnji u prošlosti. Drugim riječima, prognoza je predviđanje varijabli bazirano na poznatim vrijednostima iz prošlosti i sadašnjosti tih varijabli ili druge povezane varijable (Babu, Venkataramanaiah, & Tech, 2013).

Prognoza je kvantitativna procjena o vjerojatnosti budućih događaja. Ta je informacija izražena s jednom odnosno više jednadžbi ili uz pomoć modela vremenske serije. Ekstrapolacijom modela izvan razdoblja za koje se procjenjuje moguće je prognozirati buduće događaje, a najbolja je ona prognoza koja ima najmanju pogrešku.<sup>22</sup>

Prognoziranje je jedna od najvažnijih poslovnih funkcija jer ona predstavlja temelj mnogih poslovnih odluka.<sup>23</sup> Glavna je svrha prognoziranja dobivanje znanja o nesigurnim pojavama koje su bitne za donošenje takvih poslovnih odluka.

Postoje različite metode prognoziranja koje se koriste u ekonomiji, a to su:

- pogađanje, pravilo palca ili neformalni modeli
- procjene eksperta
- ekstrapolacija
- vodeći indikatori

---

<sup>22</sup> Babu, M., Venkataramanaiah, M. and Jahnavi, M. (2013): *Forecasting with Econometric Methods. 1st edn.* Scholars, str. 245.

<sup>23</sup> Sanders, N. (2016): *Forecasting Fundamentals*, Business Expert Press, str. 26.

- upitnici
- modeli vremenskih serija i
- ekonomski sustav i dr.

U turističkoj se djelatnosti prognozom koristi mnogo menadžera kako bi reducirali rizik koji za sobom nosi donošenje odluka koje se odnose na buduće događaje. Na primjer, marketing menadžeri mogu koristiti prognozu kako bi:

- postavili marketinške ciljeve, strateški ili godišnji plan marketinga
- istražili potencijalna tržišta s ciljem povećanja obima prodaje
- simulirali utjecaj budućih događaja na potražnju ili poslovne aktivnosti konkurencije.

Operativni i taktički menadžeri u turizmu koriste prognozu kod:

- određivanja operacionih potreba poput kadrova, inventara ili kapaciteta
- proučavanje isplativosti projekta poput financijska stabilnost izgradnje novog objekta ili proširenja restorana i sl.

Prognoziranje se mogu i javne službe uključene u turizam s ciljem:

- predviđanja ekonomskih, socio-kulturnih i ekoloških posljedica turističkih kretanja
- procjene potencijalnog utjecaja zakona i pravilnika na kvalitetu kontrole okoliša
- osiguranja dovoljnog kapaciteta infrastrukture što podrazumijeva prometnice, mostove, aerodrome, piste, ostale oblike javnog prijevoza, energiju i vodu.<sup>24</sup>

Danas je teško zamisliti da turističko poduzeće, ili pak bilo koje poduzeće, ne koristi prognoziranje kako bi se nosilo s rizikom koje donosi budućnost. Berstein (1996) tvrdi kako je uspješan rukovoditelj prije svega prognozer. Kupnja, proizvodnja, marketing, cijene i organiziranje slijedi potom.“

#### *Faze prognoziranja*

Proces prognoziranja se u praksi prema Frechtlingu (2012) općenito sastoji od sljedećih faza:

1. faza dizajniranja
2. faza specifikacije
3. faza implementacije
4. faza evaluacije.

<sup>24</sup> Frechtling, D. (2012) *Forecasting Tourism Demand*. 1st edn. Taylor and Francis, str 10.

Rezultat faze dizajniranja je odabir adekvatnog prognostičkog modela. U ovoj se fazi analizira problem, dostupni resursi i veze koje će pomoći pri preliminarnom određivanju metode. Faza se definiranja može podijeliti u nekoliko koraka:

1. definiranje problema
2. određivanje potreba korisnika
3. određivati varijable koje će se prognozirati
4. određivanje dostupnih resursa
5. postavljanje hipoteze
6. pregled dostupnosti podataka
7. popisivanje dostupnih metoda prognoziranja
8. primjenjivanje preliminarnih kriterija odabira metode
9. preliminarni odabir metode prognoziranja.

U fazi specifikacije određuju se odnosi koje će sadržavati modeli te se odabiru adekvatni modeli. Glavni je cilj u ovoj fazi odabrati najbolji mogući prognostički model koji se bazira na uzorcima i vezama među podacima iz prošlosti. Faza se specifikacije sadrži sljedeće korake:

1. određivanje odnosa
2. prikupljanje, priprema i provjera podataka (inputa)
3. odabir i programiranje preliminarnog modela
4. provjera razumnosti modela
5. određivanje preciznosti modela u prošlosti
6. testiranje alternativnih modela
7. uspoređivanje preciznosti različitih modela te njihove mogućnosti za predviđanjem trendova i odabir najpovoljnijeg modela
8. dokumentiranje rezultata.

Faza implementacije se sastoji od primjene odabranog modela kako bi se generirala prognoza i pripreme izvješća prognoze koje se prezentira korisnicima. Ova faza prolazi kroz sljedeće korake:

1. dobivanje potrebne prognoze
2. izvođenje subjektivnih korekcija ukoliko je potrebno
3. dokumentiranje modela i rezultata
4. prezentiranje prognoze.

Faza evaluacije podrazumijeva nadzor nad prognozom kroz vrijeme sa svrhom određivanja potreba za promjenom modela te implementacijom određenih korekcija kako bi se postigle najtočnije vrijednosti prognoziranog niza. Ova faza podrazumijeva:

1. praćenje preciznosti prognoze
2. određivanje uzroka devijacija
3. revizija modela, ukoliko je potrebno
4. određivanje da li dolazi do promjene među parametrima
5. generiranje nove prognoze iz postojećeg modela ili razvijanje novog modela.<sup>25</sup>

Prognoziranje je veoma zahtjevna i složena aktivnost, te mnoge prognoze mogu naposljetku biti pogrešne.

U poslovnom svijetu, postoji potreba za stalnim donošenjem odluka. Važne odluke potiču stvaranje predodžbe o budućnosti i predviđanje. Veoma je važno da odluka nije pogrešna jer to može rezultirati neuspješnim poslovanjem i gubicima stoga je potrebno angažirati niz raznih vještina kako bi donesena odluka bila uspješna (Morrell, 2018).

Ekonomске prognoze se često oslanjaju na statističke i ekonometrijske metode, ali to ne mora uvijek biti slučaj. Neke prognoze ne uključuju uopće matematičke tehnike, npr. Upitnici i anketna istraživanja mogu također proizvesti kvalitetnu prognozu bez upotrebe ekonometrijskih metoda prognoziranja (Evans, 2003.).

Prognoziranje nije rezervirano samo za budžet poduzeća, već se prognoze mogu koristiti i za prognoziranje različitih varijabli kao što su npr. procjena isplativosti projekta, procjena količine prodaje, procjena prodaje kod konkurencije i sl. Samim time, prognoziranje je ključni dio menadžmenta te se upotrebljava gotovo uvijek na jedan ili drugi način (Morrell, 2018).

Svaka prognoza ima veliki broj zahtjeva i ograničenja. Ti su zahtjevi često neispunjivi ili veoma teško ispunjivi, a to su:

- vremenska serija mora biti stacionarna (slična središnja vrijednost i standardna devijacija u preliminarnom testiranju modela) i ponavljajuća
- prognoza zahtjeva vremensku seriju sa velikom brojem povijesnih opažanja (ne zahtjeva samo mnogo podatkovnih točaka, već zahtjeva i podatkovne točke koje uzrokuju mnoge događaje)
- struktura vremenske serije mora polaziti iz unutarnjeg ponašanja sustava

---

<sup>25</sup> Frechtling, D. (2012): *Forecasting Tourism Demand 1st edn.*, Taylor and Francis., str. 86.



- čak i najmanje smetnje u vremenskoj seriji mogu sakriti nelinearnost, stoga serija podataka mora biti relativno bez smetnji
- dugoročna prognoza je manje precizna od kratkoročne prognoze
- potreba za testiranjem nulte hipoteze (testiranje značajnosti rezultata prognoze).<sup>26</sup>

U turizmu se prognoziranje koristi kako bi se prognozirale različite pojave, kao što su očekivani prihodi, očekivani troškovi, kretanje tekuće imovine, potražnju, strukturu gostiju i sl. Kao i u ostalim djelatnostima, nastoji se prognozirati što je više pojava moguće kako bi odluke na svim menadžerskim razinama bile što efikasnije usmjerene na buduće događaje. Turističkom se potražnjom smatra mjera koja pokazuje u kojoj je količini kupljen neki proizvod ili usluga, u slučaju ovoga rada radi se o usluzi iznajmljivanja ležaja odnosno noćenju. Prognoze ponude mogu se koristiti kako bi se predvidjela potražnja za željeno razdoblje te kako bi se minimizirao rizik vezan za odluke koje se odnose na budućnost. U djelatnosti turizma moguće je prognozirati potražnju kako bi se:

- postavili tržišni ciljevi, strateški ili godišnji planovi
- istražila potencijalna tržišta te ispitali kupovni potencijal istog
- simulirali budući događaji te testirale akcije koje se namjeravaju implementirati
- odredile operacijske potrebe za inventarom, osobljem i kapacitetom
- ispitala isplativost novih projekta poput troškova održavanja novog hotela, proširenje restorana i sl.<sup>27</sup>

### 3.2. Metode prognoziranja u turizmu

Neka je dana je vremenska serija sa  $t$  članova  $X_1, X_2, X_3, \dots, X_t$  i potrebno je prognozirati buduću vrijednost niza  $X_{t+n}$ . „ $n$ “ se naziva horizont prognoze (lead time) te označava određeno buduće razdoblje, a „ $t$ “ označava razdoblje u kojemu se vrši prognoza. Izraz  $X_{t+n}$  možemo tumačiti kao prognoziranu vrijednost u vremenu „ $t$ “ za „ $n$ “ razdoblja unaprijed. Prognoza se prema tome može izraziti kao  $X_t(n)$ .

Metode prognoziranja su procedure za računanje budućih vrijednosti koristeći se sadašnjim i prošlim vrijednostima promatrane varijable. Kao takve one mogu jednostavno biti skup algoritamskih pravilo i ne moraju ovisiti o temeljnom modelu. Alternativno, mogu proizaći iz

<sup>26</sup> Babu, M., Venkataramanaiah, M. and Jahnavi, M. (2013): *Forecasting with Econometric Methods. 1st edn.*, Scholars Press., str. 98.

<sup>27</sup> Frechtling, Douglas C. (2001.): *Forecasting Tourism Demand: Methods and Strategies*, Routledge by Taylor and Francis Group, New York, str. 111.

identificiranja određenog modela za dane podatke i pronalaženja optimalne prognoze uvjetovane tim modelom.<sup>28</sup>

Chatfield (2000) je generalno podijelio metode prognoziranja u tri skupine:

1. prosudbene prognostičke metode – baziraju se na subjektivnoj procjeni, znanju ili nekoj relevantnoj informaciji
2. univarijatne metode – ovise isključivo o sadašnjim i prošlim podacima vremenske serije koja se prognozira uvećanima funkcijom vremena kao što je linearni trend
3. multivarijatne metode – kretanje promatrane varijable ovisiti će, barem djelom, o kretanju neke druge varijable (*predictor* ili *explanatory variables*). Ovakve metode podrazumijevaju više jednadžbi ukoliko su varijable zajednički ovisne.

Prognozer može odabrati metodu prognoziranja s obzirom na njegovo znanje i informiranost o problemu. Pri tome, dok se prognoza odvija, moguće je korigirati metode prognoziranja koje se primjenjuju kako bi se zadovoljile trenutne potrebe.<sup>29</sup>

Metoda prognoziranja je način sustavne organizacije informacija iz prošlosti o nekoj pojavi sa ciljem boljeg shvaćanja kretanja pojave u budućnosti. Riječ „sustavno“ se odnosi na poduzimanje skupa postupaka prema određenom redoslijedu.

Cilj svake prognostičke metode je minimiziranje pogrešaka u samoj prognozi, stoga će uspješnost prognoze ovisiti o efikasnosti prognostičkog modela.<sup>30</sup>

Prognoze se mogu podijeliti s obzirom na nekoliko parametra. Jedan od načina na koji se mogu klasificirati prognoze jest prema njenim osnovnim značajkama:

1. kvantitativne metode prognoziranja
2. kvalitativne metode prognoziranja.

Kvantitativne metode prognoziranja temelje se upotrebi matematičkoga i statističkoga instrumentarija. Takve metode su objektivne i konstantne te su sposobne obraditi velike količine podataka te otkrivati složene uzročno posljedične veze među podacima. Uz pretpostavku da su dostupni adekvatni podaci, ova je metoda u načelu preciznija od prosudbene prognostičke metode.

---

<sup>28</sup> Chatfield, C. (2000): *Time-Series Forecasting. 1st edn.*, CRC Press., str. 16.

<sup>29</sup> Babu, Mohan A. i dr. (2013): *Forecasting with Econometrics Methods*, AV Akedemikerverlag GmbH & Co., Saarbrücken, ISBN: 978-3-639-51858-0, str 43.

<sup>30</sup> Babu, M., Venkataramanaiah, M. and Jahnavi, M. (2013): *Forecasting with Econometric Methods. 1st edn.* Scholars' Press., str.19.

Osnovne prednosti korištenja kvantitativnih metoda prognoziranja jesu:

- objektivnost
- konzistentnost
- mogućnost analize velikoga broja podataka
- mogućnost korištenja velikoga broja varijabli i razumijevanje složenih odnosa među njima.

Glavni nedostaci korištenja kvantitativnih metoda prognoziranja jesu:

- sporo reagiraju na promjene u okolini
- dobivena je prognoza onoliko kvalitetna koliko su kvalitetni i vjerodostojni podaci koji se obrađuju.
- zahtjeva tehnička znanja.

Prosudbene prognostičke metode, često nazvane kvalitativnim metodama prognoziranja temelje se na subjektivnom mišljenju i individualnim procjenama. Takve se metode primarno koriste kada nema relevantnih podataka iz prošlosti vezanih za određenu pojavu ili kada je pojavu teško izraziti broječno. Kvalitativne metode prognoziranja, kao Delphi metoda, metoda upitnika ili metoda intervjua, često se koriste u društvenim istraživanjima.

Osnovne prednosti korištenja kvalitativnih metoda prognoziranja su:

- lako se prilagođavaju promjenama na tržištu
- mogu uključivati informacije koje je teško kvantificirati
- mogu poslužiti za predviđanje jednokratne ili neobične aktivnosti.

Neki od nedostataka korištenja kvalitativnih metoda prognoziranja su (Sanders, 2017.):

- ograničen opseg pozornosti
- kratkoročnost
- poteškoće u shvaćanju jednostavnih veza između varijabli
- manjak konzistentnosti
- optimističnost
- priželjkivanje
- politička manipulacija.

Drugi način klasifikacije prognoziranja može se izvesti s obzirom na vremenski period kojeg obuhvaća prognoza. Taj vremenski period podrazumijeva buduće vrijeme kretanja varijable koje se nastoji prognozirati te se prema tome prognoze dijele na:

- kratkoročne prognoze – prognostičke metode gdje se prognozira pojava za narednih 3 do 6 mjeseci unaprijed. Takve su prognoze pretežito podloga za operativne odluke. Primjer problema gdje se koriste kratkoročne prognoze bilo kretanje i kontrola zaliha, planiranje proizvodnje i distribucije za kraći period.
- srednjoročne prognoze – metoda prognoziranja kod koje se prognozira pojava za narednih 6 mjeseci do dvije godine. Srednjoročne prognoze podloga su taktičkom planiranju i odlučivanju. Odluke koje zahtijevaju ovakve metode prognoziranja mogu biti odluke o promjeni zaposlenika ili odluke o upravljanju objektom ili proizvodnjom.
- dugoročne prognoze – podrazumijevaju prognoziranje dužeg vremensko razdoblje, obično više od dvije godine. Pomaže kod formuliranja strateških odluka i planova. Neke takve odluke bile bi akvizicije i stapanja, promjene proizvoda, istraživanje i razvoj i sl.<sup>31</sup>

Treba napomenuti kako je klasifikacija prognoziranja prema vremenu za koje se prognozira veoma relativna. Točan vremenski period koji će obuhvaćati kratkoročna ili srednjoročna prognoza ovisi će o varijabli koja se prognozira. U nekim slučajevima kratkoročno može biti nekoliko dana, tjedana ili mjeseci dok s druge strane kratkoročno se može odnositi i na period do godine dana.

#### *Kvantitativne metode prognoziranja*

Kvantitativne metode prognoziranja uobičajeno koriste povijesne podatke i prognostički model. Model izražava statistički odnos između prethodnih i trenutnih vrijednosti varijable. te se koristi za ekstrapolaciju prošlog i sadašnjeg ponašanja u budućnost.

Postoji nekoliko vrsta prognostičkih modela, no tri se modela najčešće pojavljuju u praksi:

1. regresijski model
2. model izgladivanja
3. model vremenske serije.

---

<sup>31</sup> Babu, Mohan A. i dr. (2013): *Forecasting with Econometrics Methods*, AV Akademikerverlag GmbH & Co., Saarbrücken, str. 56.

Regresijski modeli ili uzročni prognostički model pretpostavljaju da jedna varijabla utječe na kretanje druge varijable od interesa (prognoziranu varijablu). Uzročne metode nastoje matematički objasniti uzročno-posljedične veze između varijabli. To se postiže određivanjem uzročne varijable (objašnjavajuća varijabla) koja utječe na varijablu za koju se vrši prognoza te postavljanje matematičkog izraza.<sup>32</sup>

Modeli izgladivanja primjenjuju jednostavnu funkciju podataka iz prošlosti kako bi predvidjeli kretanje varijable u budućnosti.

Modeli vremenskih serija koriste statistička svojstva povijesnih podataka kako bi odredili formalni model i zatim procijenili nepoznate parametre.<sup>33</sup> Te se metode nazivaju i ekstrapolativne metode te polaze od pretpostavke da prošle vrijednosti pojave predstavljaju ključ prognoze budućih vrijednosti.

Srž kvantitativnih metoda prognoziranja leži u identifikaciji određenih uzoraka ponašanja u vremenskoj seriji odnosno u uočavanju karakteristika vremenske serije.

### 3.3. Mjere prognostičke efikasnosti

Jedan među najvažnijim kriterijima uspješnosti prognoziranja jest točnost. Mjere ocjenjivanja točnosti temelje se na razlikama između stvarnih i prognostičkih vrijednosti. Što je izražena veća razlika, prognoze su manje uspješne, i obratno. Dakle, pouzdanost prognoze ovisi o tome koliko su bliske prognoze varijable stvarnim vrijednostima iste varijable. Razlika između stvarne i prognozirane vrijednosti, poznata pod nazivom prognostička pogreška, definirana je izrazom:

$$e_f = Y_f - \hat{Y}_f \quad (37)$$

gdje je:

$e_f$  – prognostička pogreška

$Y_f$  – stvarna vrijednost promatrane pojave

$\hat{Y}_f$  – prognozirana vrijednost promatrane pojave

---

<sup>32</sup> Frechtling, D. (2012): *Forecasting Tourism Demand. 1st edn.*, Taylor and Francis., str 11.

<sup>33</sup> Montgomery, D., Jennings, C. and Kulahci, M. (2015): *Introduction to Time Series Analysis and Forecasting. 2nd edn.*, Wiley., str.6

Ukoliko je stvarna vrijednost veća od prognozirane vrijednosti u vremenu  $t$ , onda je prognostička pogreška pozitivna, i obrnuto ukoliko je stvarna vrijednost manja od prognozirane tada je prognostička pogreška negativna.

Svaki model dizajniran za prognoziranje pojava koje su posljedica ljudskih djelovanja sadržati će prognostičke pogreške. Pogreške su prisutne zbog najmanje tri čimbenika koji ponekad međusobno djeluju:

1. Izostavljanje utjecajnih varijabli – niti jedan prognostički model ne može uključiti sve varijable koje djeluju na pojavu. Čak i kada bi se mogli prikupiti podaci za sve varijable koje utječu na pojavu, mala je vjerojatnost da bi se ikakvim modelom mogle objasniti veze između aktualnih i prognoziranih vrijednosti. Primjerice, na broj dolazaka mogu utjecati nekakve izvanredne i neuobičajene varijable poput katastrofalnog vremena, kvara na sustavu javnog prijevoza ili radnički štrajkovi i sl.
2. Pogreška u mjerenju – neke varijable je iznimno teško točno izmjeriti kroz vrijeme kao što su primjerice brojevi turističkih dolazaka za određenu regiju ili turistička potrošnja u destinaciji. Također postoje varijable koje nemaju brojčanu vrijednost kao npr. atraktivnost destinacije, a kako bi se izvršila prognoza taj će se podatak kvantificirati nekakvom ljestvicom koja neće biti u potpunosti točan odraz atraktivnosti zbog prisutnosti ljudskog faktora. Opće je poznato u sferi turizma da su turistički utjecaji na gospodarstvo teško mjerljivi te čak postoje i metode poput turističke satelitske bilance kako bi se što preciznije odredio ukupni utjecaj turizma.
3. Ljudska neodlučnost – ljudska bića ne djeluju uvijek racionalno i predvidljivo. Primjerice, unatoč nepovoljnoj ekonomskoj situaciji, gosti se i dalje odlučuju na turistička putovanja ili iznenadna promjena uobičajene turističke destinacije nakon više godina uzastopnog posjećivanja. Osim toga ljudi se mogu i razboljeti, ozlijediti, rastati te na bilo koji drugi način promijeniti planove što je jako teško bilježiti i predvidjeti.<sup>34</sup>

U literaturi se navode brojni pokazatelji koji se mogu koristiti u evaluaciji prognostičkih vrijednosti, kao što su prosječna apsolutna pogreška i prosječna apsolutna postotna pogreška.

Prosječna apsolutna pogreška (*Mean Average Error* – MAE ili *Mean Absolute Deviation* - MAD), kao mjera varijabilnosti podataka u seriji, predstavlja apsolutnu vrijednost sume

---

<sup>34</sup> Frechtling, D. (2012): *Forecasting Tourism Demand. 1st edn.*, Taylor and Francis, str. 18.

pogreška za svako promatranje podijeljena s brojem promatranja. Prosječna apsolutna pogreška dana je izrazom:

$$MAE = \frac{\sum_{i=1}^n |e_t|}{n} \quad (28)$$

Prosječna apsolutna postotna pogreška (*Mean Absolute Percentage Error* – MAPE). Dobije se računanjem apsolutne pogreške za svaki vremenski period što se zatim dijeli sa odgovarajućom stvarnom vrijednošću te množi sa sto. Tada MAPE glasi:<sup>35</sup>

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|Y_t - \hat{Y}_t|}{Y_t} \cdot 100\% \quad (39)$$

Razine pouzdanosti ovisno o veličini prosječne postotne apsolutne pogreške u literaturi se interpretiraju na sljedeći način (Frechtling, 2012):

- <10% - visoko precizna prognoza
- 10% < x < 20% - dobra prognoza
- 20% < x < 50% - prihvatljiva prognoza
- >50% - netočna prognoza.

Takav standard može biti ponekad pogrešan jer ignorira karakteristiku promjene varijabli. Primjerice, ako je prosječna godišnja stopa rasta 2% za vremensku seriju, tada prognostički model s MAPE od 8% nije naročito koristan. U takvim se slučajevima uspoređuje MAPE i MAPE naivnog modela. Ukoliko vrijednost MAPE prvobitnog model pokazuje manju preciznost od vrijednosti MAPE naivnog modela tada se ne isplati razvijati prvobitni model. Kako bi se postavili standardi za evaluaciju MAPE prognostičkog modela, potrebno je izračunati MAPE za naivni model, usporediti MAPE i MAPE za naivni model te primijeniti sljedeće smjernice:

- Model gdje vrijednost MAPE iznosi manje od jedne polovine vrijednosti MAPE naivnog modela je visoko precizan prognostički model

---

<sup>35</sup> *Ibidem*, str. 20.

- Model gdje vrijednost MAPE iznosi 50% do 100% vrijednosti MAPE naivnog modela je razumno precizan model.
- Model gdje je vrijednost MAPE veća od vrijednosti MAPE naivnog modela smatra se lošim modelom.

U posljednjem se slučaju više isplati primijeniti jednostavniji naivni model, a prognostičke pogreške bile bi niže nego kada bi se zadržao prvobitni model.

MAPE je potrebno primijeniti na finalni, ne transformirani niz koji se planira prognozirati, u suprotnome MAPE se ne može računati.<sup>36</sup>

Još jedna mjera prognostičke pogreške korisna kod analize vremenske serije je (*Root mean square percentage error* – RMSPE). Dana je sljedećim izrazom:

$$RMSPE = \sqrt{\frac{\left(\frac{e_t}{Y_t}\right)^2}{n}} \cdot 100 \quad (40)$$

Gdje je:

n – broj perioda (promatranja)

$e_t$  – prognostička pogreška

$Y_t$  – stvarna vrijednost varijable koja se prognozira

t – vremenski period.

Ova mjera prognostičke pogreške također računa pogrešku u postocima i može biti uspoređena sa stvarnim promjenama koje su prisutne u podacima iz prošlosti. No, ovakva mjera više kažnjava veće greške u odnosu na male pogreške.

Također kao i kod MAPE, RMSPE pristup procjene pogrešaka trebao bi biti primijenjen samo na finalnom, ne transformiranom nizu podataka pojave koja se prognozira.

---

<sup>36</sup> *Ibidem*, str. 24.



## 4. Modeliranje i prognoziranje ukupnih prihoda u hotelu

### *Nautica*

Hotel *Nautica* je hotel s četiri zvjezdice koji se nalazi u gradu Novigradu u Istri. Smješten je na samoj obali, a u sklopu hotela postoji i marina. Operativan je već 17 godina te je od kraja 2021. godine u vlasništvu češkog investicijskog fonda EMMA Capital, a hotelom upravlja hrvatsko poduzeće SIGMA posloводство d.o.o. ima cjelogodišnje poslovanje, sa iznimkom razdoblja predviđenih za radove renoviranja. U strukturi gostiju prema zemlji dolaska, najveći broj gostiju dolazi iz Njemačke, Austrije, Italije i Slovenije, a čak 60% rezervacija smještaja vrši se posredstvom *online* turističke agencije Booking.com. S obzirom da je broj soba manji od 50, hotel *Nautica* se smatra malim hotelom što se odražava i na njegovu organizacijsku strukturu. Većina aktivnosti vezanih za prodaju smještaja obavljaju se u hotelu te ne postoje zasebni odjeli podrške poput odjela prodaje, upravljanja prihodima, rezervacijski centar i sl. Hotel zapošljava oko 60 radnika, a od toga je pet djelatnik zaposleno na recepciji hotela. Dosadašnje poslovanje hotela je uspješno što pokazuju godišnji prihodi koji iz godine u godinu rastu, a sudeći prema ocjenama sa Booking.com platforme kvaliteta usluge je na razini ostalih dobrih hotela na Jadranu te ona iznosi 8,8.

Cilj istraživanja provedenoga u diplomskome radu jest modelirati i prognozirati kretanje ukupnih mjesečnih prihoda u hotelu *Nautica*, a s obzirom na ostvareni broj noćenja i prosječnu cijenu po noćenju.

Korištena metodologija slijedila je temeljene faze metodologije ekonometrijskoga istraživanja te se sažeto može prikazati kao u tablici 2.:

<b>KORACI</b>	<b>KORIŠTENE METODE I MODELI</b>
<b>Postavljanje istraživačke hipoteze</b>	Broj noćenja i prosječna cijena noćenja statistički značajno pozitivno utječu na ukupne prihode
<b>Prikupljanje podataka</b>	Prikupljanje podataka o broju ostvarenih noćenja, prosječnoj cijeni po noćenju te ostvarenim ukupnim prihodima u razdoblju od siječnja 2017. godine do prosinca 2022. godine
<b>Statističko deskriptivna analiza korištenih varijabli</b>	<ul style="list-style-type: none"> <li>– Korelacijska analiza i grafički prikazi varijabli</li> <li>– Analiza sezonalnosti korištenih varijabli primjenom ANOVA regresijskih modela</li> </ul>
<b>Ekonometrijsko modeliranje ukupnih prihoda modelima regresijske analize</b>	<ul style="list-style-type: none"> <li>– Dvostruko logaritamski model višestruke regresijske analize:</li> <li>– ANCOVA regresijski model za utvrđivanje statistički značajne razlike kretanja prihoda prije i poslije pojave pandemije koronavirusa</li> </ul>
<b>Evaluacija valjanosti modela</b>	Ekonometrijski testovi i dijagnostika
<b>Evaluacija prognostičke efikasnosti modela</b>	Srednje apsolutna postotna prognostička pogreška
<b>Upotreba modela</b>	<ul style="list-style-type: none"> <li>– Prognoziranje ukupnih prihoda</li> <li>– Usporedba stvarnih i prognoziranih vrijednosti ostvarenih ukupnih prihoda</li> </ul>

Tablica 2.: Postupak analize i modeliranja podataka (Izvor: izrada autora)

## 4.1. Deskriptivna analiza varijabli korištenih u analizi

U nastavku se ispituju osnovne karakteristike prikupljenih podataka kao što su obim nedostajućih opažanja, korelacijska analiza i analiza sezonalnosti.

U analizi su korištene sljedeće varijable:

- mjesečni prihodi u eurima (UPE, u €)
- broj noćenja (N)
- prosječna cijena po noćenju (PCPN, €).

Podaci korišteni u analizi su mjesečni te pokrivaju razdoblje od siječnja 2017. godine do prosinca 2022. godine. Prije ekonometrijske analize, potrebno je napomenuti kako uzorak nije potpun. Slika 11. prikazuje strukturu i obim nedostajućih opažanja.

```
2018:02      3 missing values
2020:04      3 missing values
2020:05      3 missing values
2020:06      3 missing values
2020:07      3 missing values
2020:08      3 missing values
2020:09      3 missing values
2020:10      3 missing values
2020:11      3 missing values
2020:12      3 missing values
2021:01      3 missing values
2021:02      3 missing values

Number of observations (rows) with missing data values = 12 (16,67%)
Total number of missing data values = 36 (16,67% of total data values)

N: 12 missing values (16,67%); 60 valid values (83,33%)
UPE: 12 missing values (16,67%); 60 valid values (83,33%)
PCPN: 12 missing values (16,67%); 60 valid values (83,33%)
```

Slika 11.: Obim i struktura podataka, (Izvor: Gretl)

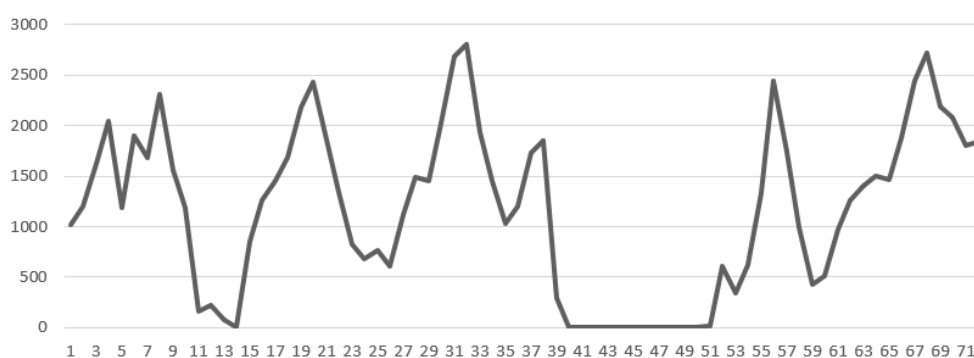
Vrijednosti koje nedostaju odnose se na mjesec veljaču 2018. godine, te na razdoblje od travnja 2020. godine do veljače 2021. godine. Od ukupno 216 vrijednosti, 36 vrijednosti nedostaje odnosno 16,67%. Razlozi nedostataka opažanja leže u činjenici da se u to vrijeme pojavila globalna pandemija uzrokovana pojavom koronavirusa koja je u velikoj mjeri ograničavala kretanje osoba izvan uobičajene sredine. Shodno tome, hotel je zatvorio svoja vrata za goste u razdoblju od gotovo godinu dana. Unatoč vrijednostima koje nedostaju, podaci nisu imputirani te je modeliranje i prognoziranje temeljeno na postojećim podacima.

Na slici 12. su prikazani osnovni statistički pokazatelji korištenih varijabli.

	Mean	Median	S.D.	Min	Max
N	1396	1449	701,1	15,00	2807
UPE	94109	84906	57938	720,0	2,436e+005
PCPN	66,15	64,00	14,85	42,00	100,0

Slika 12.: Deskriptivno – statistički pokazatelji korištenih varijabli (Izvor: Gretl)

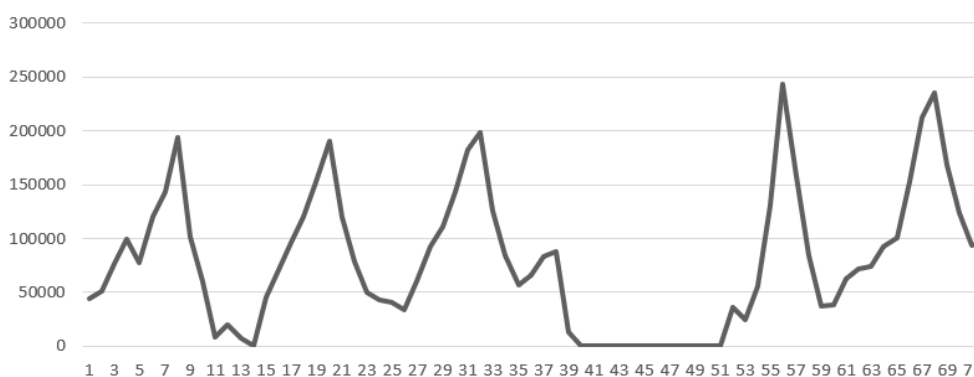
U periodu od siječnja 2017. godine do prosinca 2022. godine prosječan broj noćenja iznosi je 1396 noćenja sa standardnom devijacijom od 701 noćenja. Visoka vrijednost standardne devijacije može biti posljedica sezonalnosti turizma. Kretanje broja noćenja prikazano je na slici 13.



Slika 13.: Kretanje ostvarenih noćenja u razmatranom razdoblju. (Izvor: Gretl)

Najveći broj noćenja ostvaren je u kolovozu 2019. godine kada je u hotelu ostvareno 2 807 noćenja, a najmanje je noćenja ostvareno u ožujku 2021. godine kada je ostvareno svega 15 noćenja.

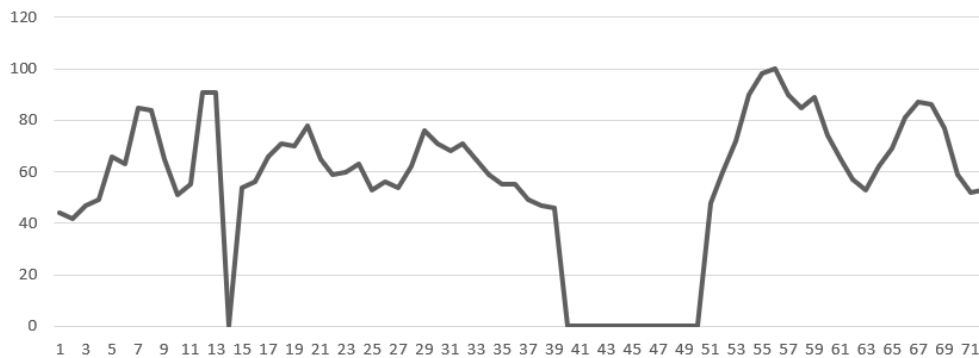
U istome periodu, mjesečni prihodi prosječno iznose 94 109 € sa standardnom devijacijom od visokih 84 906 €. Kretanje ostvarenih prihoda prikazano je na slici 14.



Slika 14.: Kretanje ostvarenih prihoda u razmatranom razdoblju. (Izvor: Gretl)

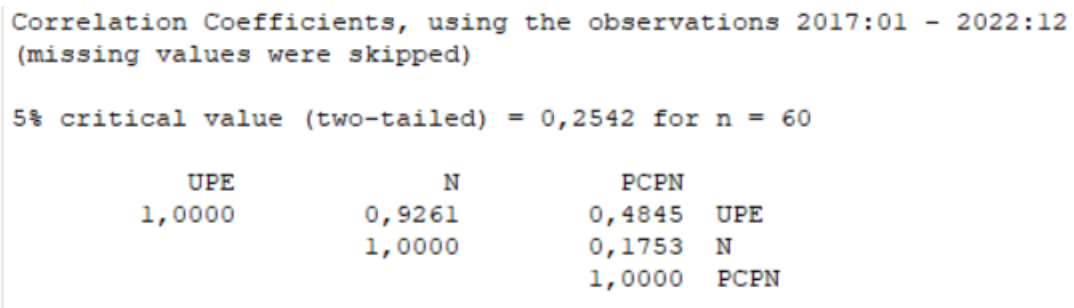
U kolovozu 2021. godine ostvareni su najveći prihodi u iznosu od 243 559 €, a najmanji prihodi ostvareni su također u ožujku 2021. godine kada je hotel zaradio svega 720 €.

Prosječna cijena po noćenju iznosila je 66,15 € uz standardnu devijaciju od 14,85 €. Kretanje prosječne cijene po noćenju prikazano je na slici 15. Najveća prosječna cijena po noćenju iznosila je 100 € dok je najniža prosječna cijena po noćenju iznosila 42 €.



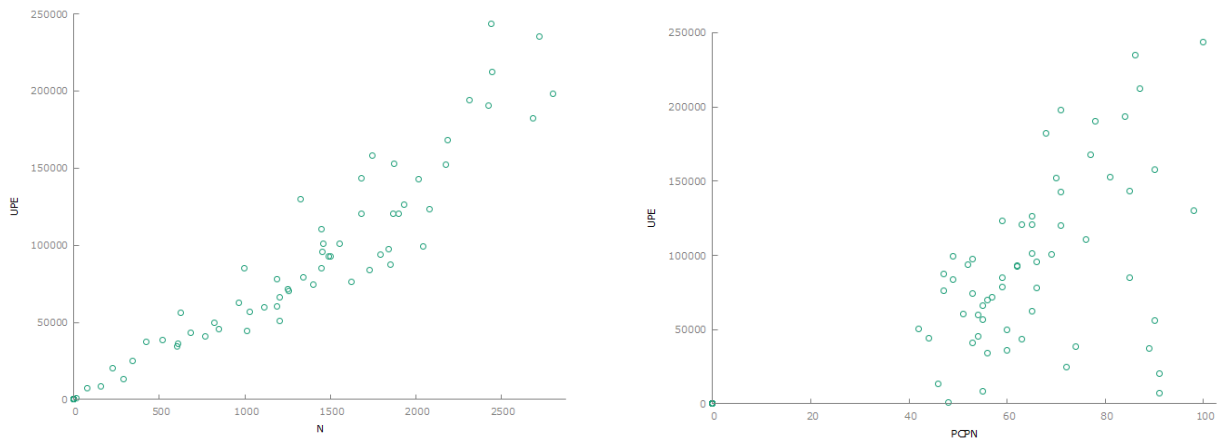
Slika 15.: Kretanje prosječne cijene po noćenju u razmatranom razdoblju. (Izvor: Gretl)

U cilju analize postojanja povezanosti među odabranim varijablama izvršena je korelacijska analiza prikazana na slici 16.



Slika 16.: Rezultati korelacijske analize (Izvor: Gretl)

Iz rezultata korelacijske matrice vidljivo je da između ukupnih prihoda, broja noćenja te prosječne cijene po noćenju postoji pozitivna korelacija. Korelacija između prihoda i noćenja znatno je jača nego li korelacija između prihoda i prosječne cijene po noćenju. Koeficijent korelacije između prihoda i noćenja iznosi  $\rho=0,9261$  što predstavlja gotovo potpunu linearnu povezanost. Stupanj povezanosti između noćenja i prihoda te između prosječne cijene po noćenju i prihoda prikazani su na slici 17.



Slika 17: Grafički prikaz povezanosti korištenih varijabli, (Izvor: Gretl)

U nastavku su postavljena tri modela višestruke linearne regresije kako bi se ispitalo postojanje sezonskih fluktuacija u korištenim varijablama. Sezonalnost je u korištenim varijablama posljedica turističke potražnje koja svoj vrhunac postiže u ljetnim mjesecima. U zimsko i jesensko doba, kada vrijeme nije sunčano niti povoljno za kupanje još uvijek ne postoji dovoljno razvijena ponuda koja bi privukla potražnju te smanjila sezonske oscilacije.

S obzirom na postojanje sezonalnosti u korištenim varijablama, procijenjena su tri ANOVA regresijska modela oblika:

$$Y = \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \gamma_4 D_4 + \dots + \gamma_{12} D_{12} \quad (41)$$

gdje je

$Y$  : zavisna varijabla

$D_1$  do  $D_{12}$  : sezonske *dummy* varijable.

Rezultati regresijske analize prikazani su na slikama 18., 19. i 20.

Na modelu sa slike 18. vidljivo kako kretnje ukupnih prihoda pokazuje znakove sezonalnosti. Svi regresijski koeficijenti značajni su na 1%. Najviše prihoda ostvaruje u lipnju, srpnju, kolovozu i rujnu odnosno u vrijeme trajanja turističke sezone na obali Jadrana. Od navedenih mjeseci, daleko najveći prihodi se ostvaruju u kolovozu te oni iznose u prosjeku 212 135 €. Najmanji se prihodi ostvaruju u zimskom periodu, točnije u siječnju i ožujku. Usporedba prosječnih prihoda u kolovozu s prosječnim prihodima u siječnju, ukazuje kako se u siječnju ostvaruje u prosjeku svega 22% prihoda u odnosu na kolovoz. U kolovozu se u prosjeku ostvaruje i do pet puta više prihoda nego li u siječnju. Standardna pogreška regresije ukazuje kako razlika između stvarnih i procijenjenih prihoda iznosi 29 678,92 €.

Model 4: OLS, using observations 2017:01-2022:12 (T = 60)  
 Missing or incomplete observations dropped: 12  
 Dependent variable: UPE

	coefficient	std. error	t-ratio	p-value	
dm1	47723,0	13272,8	3,596	0,0008	***
dm2	61000,5	14839,5	4,111	0,0002	***
dm3	45011,3	12116,4	3,715	0,0005	***
dm4	78281,4	13272,8	5,898	3,59e-07	***
dm5	81960,8	13272,8	6,175	1,35e-07	***
dm6	118423	13272,8	8,922	9,23e-012	***
dm7	164192	13272,8	12,37	1,54e-016	***
dm8	212135	13272,8	15,98	7,49e-021	***
dm9	134844	13272,8	10,16	1,51e-013	***
dm10	86467,6	13272,8	6,515	4,08e-08	***
dm11	49290,0	13272,8	3,714	0,0005	***
dm12	53176,0	13272,8	4,006	0,0002	***
Mean dependent var	94108,88	S.D. dependent var	57938,00		
Sum squared resid	4,23e+10	S.E. of regression	29678,92		
R-squared	0,786519	Adjusted R-squared	0,737597		
F(11, 48)	16,07679	P-value (F)	1,44e-12		
Log-likelihood	-696,3336	Akaike criterion	1416,667		
Schwarz criterion	1441,799	Hannan-Quinn	1426,498		

Slika 18.: Rezultati regresijske analize, Y=UPE (Izvor: Gretl)

Model 2: OLS, using observations 2017:01-2022:12 (T = 60)  
 Missing or incomplete observations dropped: 12  
 Dependent variable: N

	coefficient	std. error	t-ratio	p-value	
dm1	912,000	232,548	3,922	0,0003	***
dm2	1229,25	259,997	4,728	2,02e-05	***
dm3	880,833	212,287	4,149	0,0001	***
dm4	1380,60	232,548	5,937	3,13e-07	***
dm5	1178,60	232,548	5,068	6,39e-06	***
dm6	1619,80	232,548	6,965	8,31e-09	***
dm7	2063,80	232,548	8,875	1,08e-011	***
dm8	2543,40	232,548	10,94	1,24e-014	***
dm9	1858,40	232,548	7,991	2,27e-010	***
dm10	1411,40	232,548	6,069	1,97e-07	***
dm11	845,200	232,548	3,635	0,0007	***
dm12	894,600	232,548	3,847	0,0004	***
Mean dependent var	1395,683	S.D. dependent var	701,0681		
Sum squared resid	12978895	S.E. of regression	519,9939		
R-squared	0,552425	Adjusted R-squared	0,449856		
F(11, 48)	5,385883	P-value (F)	0,000017		
Log-likelihood	-453,6710	Akaike criterion	931,3421		
Schwarz criterion	956,4742	Hannan-Quinn	941,1726		

Slika 19.: Rezultati regresijske analize, Y=N (Izvor: Gretl)



Slika 19. prikazuje rezultate analize sezonalnosti i kretanju broja noćenja po pojedinim mjesecima.

Kao i u kretanju prihoda, kod ostvarenoga broja noćenja, može se zamijetiti statistički značajna sezonalnost. Najveći broj noćenja ostvaruje se također za vrijeme turističke sezone, a prosječno se najmanje ostvaruje noćenja u studenom. Kod noćenja, razlika između prosječno najlošijeg i prosječno najboljeg mjeseca je nešto manje drastična nego li kod prosječnih prihoda. U najlošijem mjesecu prosječno se ostvaruje 33% od noćenja koja se prosječno ostvare u kolovozu. Moguće je da prihodi pokazuju veću oscilaciju zbog utjecaja povećanja cijena za vrijeme turističke sezone. Drugim riječima, potražnja u vrijeme turističke sezone raste stoga se može postaviti veća cijena smještaja, a s druge strane ne može se povećati prihvatni kapacitet.

Regresijski rezultati analize sezonalnosti u kretanju cijene prikazani su na slici 20.

```

Model 5: OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: PCPN

```

	coefficient	std. error	t-ratio	p-value
dm1	60,4000	5,18167	11,66	1,33e-015 ***
dm2	50,5000	5,79329	8,717	1,86e-011 ***
dm3	50,3333	4,73020	10,64	3,19e-014 ***
dm4	57,8000	5,18167	11,15	6,28e-015 ***
dm5	69,8000	5,18167	13,47	6,32e-018 ***
dm6	75,2000	5,18167	14,51	3,53e-019 ***
dm7	81,6000	5,18167	15,75	1,36e-020 ***
dm8	83,8000	5,18167	16,17	4,64e-021 ***
dm9	72,4000	5,18167	13,97	1,55e-018 ***
dm10	62,6000	5,18167	12,08	3,66e-016 ***
dm11	62,2000	5,18167	12,00	4,62e-016 ***
dm12	67,2000	5,18167	12,97	2,66e-017 ***

Mean dependent var	66,15000	S.D. dependent var	14,84819
Sum squared resid	6443,933	S.E. of regression	11,58657
R-squared	0,504604	Adjusted R-squared	0,391076
F(11, 48)	4,444750	P-value(F)	0,000133
Log-likelihood	-225,4328	Akaike criterion	474,8656
Schwarz criterion	499,9977	Hannan-Quinn	484,6962

Slika 20.: Rezultati regresijske analize, Y=PCPN (Izvor: Gretl)

Kao i kod prethodne dvije varijable, najviša se cijena postiže u mjesecima od lipnja do rujna. Najviša prosječna cijena ostvaruje se u kolovozu te iznosi 83,80 € po noćenju. Razlika u najvišoj i najnižoj prosječnoj cijeni je svega 66%. Od triju korištenih varijabli, cijene pokazuju najmanje oscilacija tijekom jedne godine.



## 4.2. Ekonometrijsko modeliranje ostvarenih ukupnih prihoda

Prije postavljanja modela logaritmirani su originalni podaci za sve tri varijable kako bi se postigla stacionarnost varijabli te je za modeliranje prihoda korišten dvostruko logaritamski regresijski model oblika:

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} \quad (32)$$

gdje je:

- $Y_i$ : mjesečni prihodi u eurima (UPE, u €)
- $X_{1i}$ : broj noćenja (N)
- $X_{2i}$ : prosječna cijena po noćenju (PCPN, €).

Na slici 21. prikazani su rezultati procjene dvostruko logaritamskoga modela.

```
Model 1: OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: l_UPE
```

	coefficient	std. error	t-ratio	p-value
const	-0,00407035	0,0119484	-0,3407	0,7346
l_N	0,999780	0,000669238	1494	1,34e-132 ***
l_PCPN	1,00132	0,00275235	363,8	1,23e-097 ***

Mean dependent var	11,17193	S.D. dependent var	0,956715
Sum squared resid	0,001229	S.E. of regression	0,004644
R-squared	0,999977	Adjusted R-squared	0,999976
F(2, 57)	1251889	P-value (F)	4,8e-133
Log-likelihood	238,7314	Akaike criterion	-471,4627
Schwarz criterion	-465,1797	Hannan-Quinn	-469,0051

Log-likelihood for UPE = -431,584

Slika 21.: Rezultati regresijskog modela (Izvor: Gretl)

Rezultati testiranja statističke značajnosti regresijskih koeficijenata uz pomoć rezultata dvosmjernog t-testa (t-ratio) te p-vrijednosti (p-value) upućuju na značajnost regresijskih koeficijenata na razini od 1%. Može se stoga odbaciti nulta hipoteza o neznačajnosti statističkih parametara i zaključiti kako odabrani regresori statistički značajno utječu na kretanje ukupnih mjesečnih prihoda.

Procijenjena funkcija prihoda glasi:

$$\ln Y_i = -0,00407 + 0,99978 \ln X_{1i} + 1,00132 \ln X_{2i} \quad (43)$$

Iz danog modela može se vidjeti kako elastičnost ukupnih prihoda u odnosu na ostvarena noćenja iznosi 0,99% uz nepromijenjenu prosječnu cijenu. Drugim riječima, ukoliko se broj noćenja poveća za 1% uz nepromijenjenu prosječnu cijenu po noćenju, ukupni prihodi povećati će se za 0,99% dok s druge strane, povećanje prosječne cijene po noćenju uz nepromijenjen broj noćenja za jedan posto rezultirati će povećanjem prihoda za 1,00132%.

Antilogaritmiranjem procijenjene funkcije dobiven je procijenjeni model potencijske regresije:

$$Y_i = 0,99594 \cdot X_{1i}^{0,99978} \cdot X_{2i}^{1,00132} \quad (44)$$

Korigirani koeficijent determinacije (*Adjusted R-squared*) dobivenog regresijskog modela iznosi 0,999976 što znači da se 0,99% promjena u kretanju prihoda može objasniti kretanjem broja noćenja i prosječne cijene po noćenju. Rezultati testiranja značajnosti koeficijenta determinacije F testom ukazuju na reprezentativnost procijenjenog modela.

Standardna pogreška regresije 0.004644% što je iznimno niska pogreška.

Testiranje modela na ispunjavanje postavki klasičnog linearnog regresijskog modela prikazano je u nastavku.

Normalnost reziduala testirana je Jarque-Bera testom, a rezultati su prikazani na slici 22.

```
Test for normality of residual -  
Null hypothesis: error is normally distributed  
Test statistic: Chi-square(2) = 2,77453  
with p-value = 0,249757
```

Slika 22.: Test normalnosti reziduala (Izvor: Gretl)

Nulta hipoteza pretpostavlja normalnu distribuciju rezidualnih odstupanja. Rezultati testiranja normalnosti grešaka relacije Jarque-Bera testom sa slike 5. pokazuju kako su reziduali normalno distribuirani jer je  $JB = 2,77453 < X_{\alpha}^2(2) = 5,99$  odnosno  $p = 0,249757 > \alpha = 0,05$ . Nulta se hipoteza prihvaća kao istinita.

Homoskedastičnost varijance reziduala testirana je White testom.

```
White's test for heteroskedasticity
OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: uhat^2
```

	coefficient	std. error	t-ratio	p-value
const	-0,000796028	0,00104534	-0,7615	0,4497
l_N	4,92773e-05	4,68773e-05	1,051	0,2978
l_PCPN	0,000334404	0,000481246	0,6949	0,4901
sq_l_N	-4,18469e-07	2,17395e-06	-0,1925	0,8481
X2_X3	-1,03118e-05	1,37294e-05	-0,7511	0,4559
sq_l_PCPN	-3,44608e-05	5,59015e-05	-0,6165	0,5402

Unadjusted R-squared = 0,096806

Test statistic:  $TR^2 = 5,808370$ ,  
with p-value =  $P(\text{Chi-square}(5) > 5,808370) = 0,325314$

Slika 23.: rezultat testa heteroskedastičnosti (Izvor: Gretl)

Empirijska vrijednost veća je od kritične vrijednosti  $X_5^2 = 11,070 > 5,808370$ , što je vidljivo po rezultatima sa slike 23. Shodno tome prihvaća se nulta hipoteza o homoskedastičnosti varijance reziduala.

Slika 24. pokazuje rezultate analize postojanja problema multikolinearnosti između nezavisnih varijabli.

```
Variance Inflation Factors
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem
```

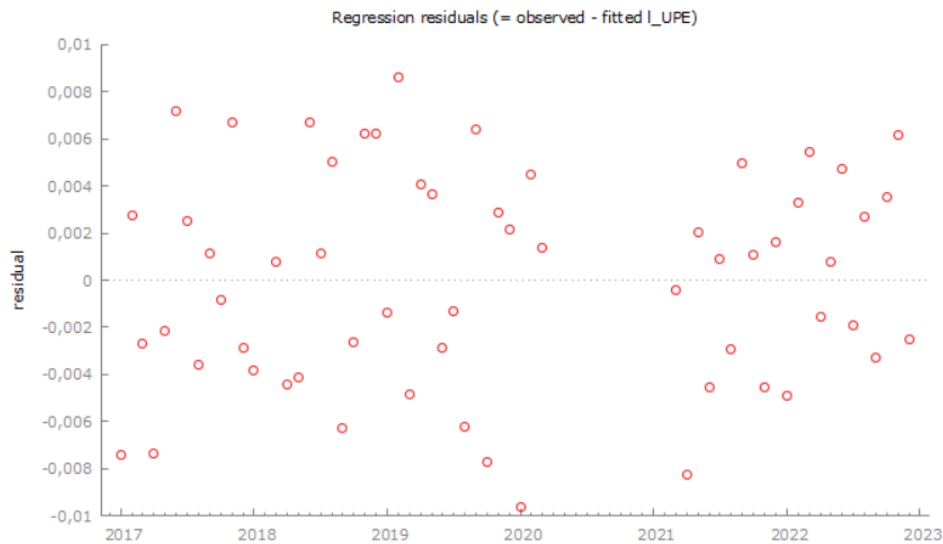
l_N	1,011
l_PCPN	1,011

VIF(j) =  $1/(1 - R(j)^2)$ , where R(j) is the multiple correlation coefficient between variable j and the other independent variables

Slika 24. Rezultat analize varijance (Izvor: Gretl)

Vrijednost VIF pokazatelja je niska i iznosi 1,011 što upućuje da u modelu ne postoji ozbiljan problem multikolinearnosti među regresorima.

S obzirom da među prikupljenim podacima postoji više opažanja s nedostajućim vrijednostima nije moguće provesti test na autokorelaciju. No, odsutnost autokorelacije može se grafički prikazati uz pomoć slučajnih odstupanja koja su potpuno nasumična što je prikazano na slici 25.



Slika 25.: Dijagram raspršenosti (Izvor: Gretl)

Na slici 26. prikazani su rezultati testa nelinearnosti, a nulta pretpostavka testa ukazuje na linearnu vezu između zavisne i nezavisne varijable.

```

Auxiliary regression for non-linearity test (squared terms)
OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: uhat

```

	coefficient	std. error	t-ratio	p-value
const	-0,255266	0,214926	-1,188	0,2401
l_N	0,000937057	0,00467278	0,2005	0,8418
l_PCPN	0,121309	0,103458	1,173	0,2460
sq_l_N	-9,31287e-05	0,000391549	-0,2378	0,8129
sq_l_PCPN	-0,0144814	0,0123494	-1,173	0,2460

Unadjusted R-squared = 0,026061

Test statistic:  $TR^2 = 1,56367$ ,  
with p-value =  $P(\text{Chi-square}(2) > 1,56367) = 0,457565$

Slika 26.: Rezultati testa linearnosti (Izvor: Gretl)

S obzirom da je empirijska p-vrijednost testa ( $p = 0,457656$ ), veća od  $\alpha = (0,05)$  nulta se hipoteza prihvaća kao istinita odnosno varijable su u linearnom odnosu.

Proveden je i Ramsey RESET test. Rezultati su prikazani na slici 27.

Rezultati Ramsey-evog RESET testa ukazuju na značajnost logaritmiranih regresijskih koeficijenata te s druge strane pokazuje kako potencirane varijable nisu značajne. RESET test pokazuje je li dobro postavljen odabrani model, a s obzirom da je p vrijednost logaritmiranih

regresorskih koeficijenata puno niža od vrijednosti  $\alpha$  na razini pouzdanosti od 99% zaključuje se kako je korišten model s logaritmiranim vrijednostima varijabli dobro postavljen.

```

Auxiliary regression for RESET specification test
OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: l_UPE

```

	coefficient	std. error	t-ratio	p-value
const	-0,0617617	0,207988	-0,2969	0,7676
l_N	1,01930	0,0689345	14,79	8,21e-021 ***
l_PCPN	1,02038	0,0675618	15,10	3,21e-021 ***
yhat^2	-0,00206891	0,00723995	-0,2858	0,7761
yhat^3	7,14490e-05	0,000248501	0,2875	0,7748

Test statistic: F = 0,041862,  
with p-value = P(F(2,55) > 0,0418624) = 0,959

Slika 27.: Rezultati Ramsey RESET-a (Izvor: Gretl)

U nastavku se nastoji analizirati utjecaj pandemije na ostvarene prihode u hotelu procjenom ANOVA regresijskoga modela oblika:

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta_1 X_{i1} + \beta_2 X_{i2} \quad (45)$$

gdje je:

- $Y_i$ : mjesečni prihodi u eurima (UPE, u €)
- $X_{1i}$ : broj noćenja (N)
- $X_{2i}$ : prosječna cijena po noćenju (PCPN, €)
- *covid\_dummy*; kvalitativna

kvalitativna varijabla  $\begin{cases} 0 = \text{za razdoblje od 2017:01 do 2020:02 godine} \\ 1 = \text{za razdoblje od 2020:03 do 2022:12 godine} \end{cases}$

Rezultati procijenjenoga modela prikazani su na slici 28.

```

Model 4: OLS, using observations 2017:01-2022:12 (T = 60)
Missing or incomplete observations dropped: 12
Dependent variable: l_UPE

      coefficient      std. error      t-ratio      p-value
-----
const      -0,00351926      0,0123007      -0,2861      0,7759
l_N         0,999802          0,000682473    1465         4,88e-130 ***
l_PCPN      1,00112           0,00291324     343,6        8,86e-095 ***
covid_dummy 0,000292582       0,00131345     0,2228       0,8245

Mean dependent var      11,17193      S.D. dependent var      0,956715
Sum squared resid       0,001228      S.E. of regression       0,004683
R-squared                0,999977      Adjusted R-squared       0,999976
F(3, 56)                820677,2      P-value (F)              6,0e-130
Log-likelihood           238,7579      Akaike criterion         -469,5159
Schwarz criterion        -461,1385      Hannan-Quinn             -466,2390

```

Slika 28.: rezultati procijenjenog modela (Izvor: Gretl)

Postavljenim modelom nastoji se utvrditi da li postoji statistički značajna razlika u ostvarenim prihodima u zavisnosti o noćenjima i prosječnoj cijeni po noćenju u razdoblju prije i poslije pandemije.

Model na slici 28. ponovo pokazuje statistički značajne rezultate. Može se primijetiti značajnost varijabli noćenja i prosječne cijene po noćenju gdje je p vrijednost niža od 0,05. Koeficijent determinacije je također vrlo visok. P vrijednost *dummy* varijable je veća od  $\alpha = 0,05$  što znači da varijabla nije statistički značajna, što ukazuje da ne postoji statistički značajna razlika u ostvarenim prihodima u dva razmatrana razdoblja.

U nastavku su testirane pretpostavke klasičnoga linearnoga regresijskoga modela.

```

Test for normality of residual -
Null hypothesis: error is normally distributed
Test statistic: Chi-square(2) = 2,48924
with p-value = 0,28805

White's test for heteroskedasticity -
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 7,48017
with p-value = P(Chi-square(8) > 7,48017) = 0,485819

```

Slika 29.: Rezultati testa normalnosti reziduala i heteroskedastičnosti (Izvor: Gretl)

Na slici 29. prikazani su rezultati testa normalnosti reziduala te testa heteroskedastičnosti. Kritična vrijednost iznosi  $X_2^2 = 5,99$  što je veće od empirijske vrijednosti koja iznosi  $JB = 2,48924$  što znači da se prihvaća  $H_0$  odnosno reziduali su normalno distribuirani. Isto se može



prikazati i uz pomoć p vrijednosti testa, gdje je  $p = 0,28805 > \alpha = 0,05$ , dakle prihvaća se nulta hipoteza.

Rezultati White-ovog testa heteroskedastičnosti. Ukazuju na homoskedastičnost varijance reziduala. I u ovome slučaju kritična vrijednost je veća od empirijske vrijednosti,  $X_3^2 = 7,48017 > W = 6,48788$  što znači da je varijanca reziduala homoskedastična.

```
Variance Inflation Factors
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

      1_N      1,022
     1_PCPN    1,155
    covid_dummy 1,147

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation coefficient
between variable j and the other independent variables
```

Slika 30.: Rezultat analize varijance (Izvor: Gretl)

Rezultati testiranja multikolinearnosti sa slike 30. također ukazuju na ne postojanje ozbiljnoga problema kolinearnosti među regresorima, jer je faktor inflacije varijance (*Variance Inflation Factor - VIF*) za svaku varijablu manji od 10.

### 4.3. Prognoziranje ukupnih prihoda

Prognoziranje na temelju svih prikupljenih podataka nije bilo moguće jer su postojala opažanja s nedostajućim vrijednostima. S obzirom da rezultati procjene ukupnih prihoda uvođenjem *dummy* varijabli nisu ukazali na statistički značajnu razliku u ukupnim prihodima prije i nakon pojave pandemije uzrokovane koronavirusom, za prognozu korišteni podaci nakon pandemije odnosno podaci od ožujka 2021. godine do prosinca 2022. godine. Prognoziranje je temeljeno na procjeni dvostruko logaritamskog modela.

U svrhu prognoziranja ukupnih mjesečnih prihoda za 2023. godinu potrebno je odrediti vrijednosti regresorskih varijabli na temelju određenih pretpostavki. S obzirom da menadžer hotela određuje cijene za narednu godinu, prosječna cijena po noćenju se može približno izračunati ukoliko je poznat i planirani broj ostvarenih noćenja. U hotelu su cijene za 2023. godinu u prosjeku 20% veće nego li 2022. godine stoga su za prognozu korišteni prošlogodišnja opažanja čije su vrijednosti uvećane za 20%. Također se pretpostavlja da će broj noćenja porasti u prosjeku za 5,99%.

Rezultati dobivenih prognoza prikazani su na slici 31.

Rezultati prognoze prikazani na slici pokazuju vrlo nisku vrijednost srednje apsolutne postotne pogreške koja iznosi svega 0,029% što govori o visokoj prognostičkoj pouzdanosti procijenjenoga modela.

2021:03	6,579251	6,577458	0,005181	6,562635 - 6,592282
2021:04	10,497808	10,506411	0,004071	10,494765 - 10,518058
2021:05	10,116742	10,113713	0,004113	10,101945 - 10,125481
2021:06	10,931856	10,935075	0,004168	10,923150 - 10,947001
2021:07	11,776189	11,774307	0,004194	11,762308 - 11,786307
2021:08	12,403114	12,405544	0,004220	12,393470 - 12,417618
2021:09	11,969515	11,964177	0,004111	11,952417 - 11,975938
2021:10	11,349736	11,348021	0,004075	11,336362 - 11,359681
2021:11	10,534307	10,537238	0,004220	10,525163 - 10,549312
2021:12	10,553962	10,551643	0,004061	10,540025 - 10,563261
2022:01	11,043594	11,048941	0,004029	11,037414 - 11,060469
2022:02	11,180148	11,178061	0,004127	11,166255 - 11,189867
2022:03	11,218823	11,215006	0,004215	11,202947 - 11,227066
2022:04	11,440570	11,443161	0,004078	11,431494 - 11,454827
2022:05	11,521687	11,521467	0,004029	11,509940 - 11,532993
2022:06	11,935058	11,930488	0,004055	11,918886 - 11,942090
2022:07	12,266278	12,268315	0,004108	12,256563 - 12,280068
2022:08	12,367302	12,364880	0,004113	12,353114 - 12,376645
2022:09	12,032106	12,035927	0,004057	12,024320 - 12,047534
2022:10	11,721288	11,719325	0,004158	11,707428 - 11,731222
2022:11	11,450795	11,446554	0,004275	11,434323 - 11,458785
2022:12	11,485338	11,489756	0,004255	11,477584 - 11,501928
2023:01		11,288962	0,004029	11,277434 - 11,300489
2023:02		11,418081	0,004025	11,406565 - 11,429597
2023:03		11,455027	0,004061	11,443407 - 11,466646
2023:04		11,683181	0,004028	11,671657 - 11,694705
2023:05		11,761487	0,004055	11,749885 - 11,773090
2023:06		12,170509	0,004183	12,158542 - 12,182475
2023:07		12,508336	0,004272	12,496114 - 12,520558
2023:08		12,604900	0,004265	12,592699 - 12,617102
2023:09		12,275948	0,004144	12,264091 - 12,287804
2023:10		11,959346	0,004063	11,947722 - 11,970969
2023:11		11,686574	0,004101	11,674842 - 11,698307
2023:12		11,729776	0,004092	11,718070 - 11,741483

Forecast evaluation statistics using 22 observations

Mean Error	-3,2297e-016
Root Mean Squared Error	0,0036478
Mean Absolute Error	0,0032178
Mean Percentage Error	0,00020062
Mean Absolute Percentage Error	0,028991
Theil's U2	0,0029532
Bias proportion, UM	0
Regression proportion, UR	0
Disturbance proportion, UD	1

Slika 31.: Rezultati prognoze (Izvor: Gretl)



Zbog interpretacije rezultata, logaritmirane vrijednosti prihoda su antilogaritmirane te su mjesečne vrijednosti prihoda u 2023. godini prikazane u izvornim jedinicama.

Dobivene vrijednosti prognoziranih mjesečnih prihoda prikazane su u tablici 3. Vidljivo je da se najveći prihodi očekuju tijekom ljetnih mjeseci, dok se najmanji prihodi očekuju u siječnju i veljači.

	log_UP (2023.)	UP, u € (2022.)	UP, u € (prognozirano, 2023.)	UP, u € (ostvareno, 2023.)
sij.	11,28896	62.542,00	79.934,43	
vlj.	11,41808	71.693,00	90.951,44	
ožu.	11,45503	74.520,00	94.374,58	
tra.	11,68318	93.020,00	118.560,78	139.677,00
svi.	11,76149	100.880,00	128.217,97	140.310,00
lip.	12,17051	152.521,00	193.012,26	
srp.	12,50834	212.411,00	270.583,50	
kol.	12,6049	234.991,00	298.015,27	
ruj.	12,27595	168.065,00	214.474,90	
lis.	11,95935	123.166,00	156.270,85	
stu.	11,68657	93.976,00	118.963,74	
pro.	11,72978	97.279,00	124.215,84	
Σ		1.485.064,00	1.887.575,56	

Tablica 3.: Ukupni mjesečni prihodi u hotelu *Nautica* u 2023. godini (Izvor: Samostalan rad autora)

Očekivani ukupni prihodi u 2023. godini iznose 1.887.576,00 € što je oko 27% više nego li prethodne godine kada su prihodi iznosili 1.485.064,00 €, a najveći prihodi se očekuju u kolovozu 2023. godine u iznosu od 298.015,27 €. S druge strane, ako promatramo prosječnu stopu rasta u kretanju prihoda koja iznosi oko 8,3%, očekivani godišnji prihodi iznosili bi 1.608.324,31 €.

Usporedba prognoziranih i ostvarenih prihoda za travanj 2023. godine ukazuje na podcjenjivanje prognoziranih prihoda, koji su manji za 21.116,22 € od ostvarenih. Nešto je manja razlika između ostvarenih i prognoziranih prihoda u svibnju, gdje ona iznosi 12.092,00 €. Za mjesec lipanj se očekuje još manja razlika između prognoziranih i ostvarenih prihoda, na što upućuje činjenica da je u trenutku pisanja diplomskoga rada planirani prihod od trenutnih rezervacija iznosio 166.752,00 €.

Razloga za odstupanje stvarnih vrijednosti od prognoziranih vrijednosti ukupnih prihoda ima mnogo, no vrlo je vjerojatno da je odstupanje većim djelom prouzročeno iznimno visokim porastom kuponske prodaje smještaja u mjesecu travnju i svibnju. U odnosu na prethodnu godinu, u mjesecu travnju 2023. godine zabilježeno je povećanje broja noćenja, od strane

gostiju koji su kupili kupon za smještaj, za 270%. U travnju 2023. godine prodano je 759 noćenja uz prihode od 42.834,00 €, a godinu ranije prodano je svega 282 noćenja s ostvarenim prihodima od 13.834,00 €.

Iz rezultata analize može se zaključiti kako broj noćenja i prosječna cijena po noćenju u velikoj mjeri objašnjavaju kretanje ukupnih mjesečnih prihoda u hotelu *Nautica*. No čak i uz sve ispunjene pretpostavke klasične linearne regresije, procijenjeni model koji se koristi za prognozu sklon je prognostičkim pogreškama koje je nemoguće predvidjeti.

U stvarnome svijetu na prihode hotela utječe mnogo faktora osim cijene i noćenja, ali i na te primarne faktore također utječu neki drugi faktori te ih je nemoguće sve uključiti u analizu. Ako se uzme u obzir ranije navedeni primjer povećanog broja noćenja ostvarenih od strane turista koji su smještaj kupili preko kupona, može se zaključiti da na broj noćenja utječe broj kupljenih kupona. Shodno tome, na kupnju kupona utjecati će neki drugi faktori, primjerice visina dohotka turista na emitivnom tržištu. Nadalje se može reći da visina dohotka ovisi o gospodarskim prilikama koje se mogu izraziti visinom BDP-a. Promjena u samo jednom od navedenih faktora rezultirati će promjenom u ostalim faktorima. Ako država ostvaruje manje prihoda, manje ostvaruju i građani. Ako građani ostvaruju manji dohodak, manje će se odvajati sa turistička putovanja. Naposljetku manje će ih posjetiti, na primjer, hotel *Nautica*.

Također treba napomenuti kako postoje i značajne nepredvidivosti na području prosječne cijene po noćenju u promatranom hotelu. Razlog tome jest visok udio *Booking.com* gostiju koji ostvaruju razne popuste i pogodnosti od kojih su neki van kontrole hotelijera. Primjer takvog popusta je *Genious booker* popust kojeg ostvaruju određeni korisnici te uz popust dobiju i povoljne uvjete otkazivanja koji ponekad dopuštaju i besplatno otkazivanje rezervacije i do dva dana prije dolaska.

Visoku pouzdanost u prognoziranju vrlo je teško postići, a s obzirom na količinu utjecajnih faktora u stvarnome svijetu. No, čak i donekle precizne prognoze mogu poslužiti kao temelj za donošenje valjanih poslovnih odluka. Kada je riječ o izvještavanju o planiranim prihodima, primjerice upravnom odboru ili dioničarima, često su dovoljne i znatno manje zahtjevne prognoze.

## ZAKLJUČAK

Svaki poslovni subjekt bilježi određene podatke koji nastaju kroz poslovne procese. Analiza i modeliranje takvih podataka može rezultirati određenim informacijama koje nadalje mogu poslužiti kao podloga za donošenje konkretnih poslovnih odluka. Štoviše, analitički pristup poslovanju, u modernome je svijetu neophodan za konkurentan performans poslovnog subjekta.

Na primjeru diplomskoga rada može se vidjeti kako modeliranje i analiza podataka iz poslovanja hotela *Nautica* uz pomoć nelinearne regresije pruža uvid u kretanje varijable koju želimo objasniti nekim drugim faktorima. Osim toga, moguće je i predvidjeti kretanje analizirane varijable.

Prema dobivenim rezultatima mogu se planirati razni poslovni procesi, nabava potrošne robe, stalna i tekuća imovina, potrebe za ljudskim resursima, efikasnost rada, i druge poslovne aktivnosti te očekivani pokazatelji uspješnosti poslovanja. Osim što rezultati analize mogu biti podloga poslovnim odlukama, oni se također prezentiraju upravnom odboru te dioničarima.

Poslovanje subjekta potrebno je analizirati kontinuirano te je potrebno analizama obuhvatiti što veći broj utjecajnih faktora. Kod hotelskog poslovanja, osim prihoda, mogu se analizirati i predviđati brojni drugi faktori poput troškova, dolazaka, broj prodanih obroka i dr. U diplomskome je radu prikazana samo jedna analitička metoda od mnogih koja je bila upotrijebljena za analizu i modeliranje prihoda.

No, kao je već ranije spomenuto, savršene je rezultate vrlo teško postići u praksi, tako je i prognoza vrlo često neprecizna. U stvarnome svijetu prisutno je suviše mnogo utjecaja koji oblikuju ponašanje bilo kojeg faktora u poslovanju hotela. Vrlo je mnogo faktora koje obični smrtnici jednostavno ne mogu predvidjeti, što se moglo vidjeti 2020. godine kada je paralizirana djelatnost turizma zbog globalne pandemije uzrokovane koronavirusom. Iako ponekad neefikasne, prognoze i analize podataka mogu i dalje rezultirati uspješnim poslovnim odlukama.

## Popis Slika:

• Slika 1.: Pozitivna korelacija.....	4
• Slika 2.: Negativna korelacija.....	5
• Slika 3.: Nekorelirane varijable.....	5
• Slika 4.: Autokorelacija.....	13
• Slika 5. Multikolinearnost.....	15
• Slika 6.: Udaljenost reziduala od pravca regresije.....	17
• Slika 7.: Logaritamska funkcija.....	21
• Slika 8: Objašnjena i neobjašnjena komponenta varijable $Y_i$ .....	24
• Slika 9.: ANOVA tablica.....	25
• Slika 10.: Koeficijent determinacije kao mjera uspješnosti prilagodbe linearne funkcije empirijskim podacima.....	27
• Slika 11.: Obim i struktura podataka.....	46
• Slika 12.: Deskriptivno – statistički pokazatelji korištenih varijabli.....	47
• Slika 13.: Kretanje ostvarenih noćenja u razmatranom razdoblju.....	47
• Slika 14.: Kretanje ostvarenih prihoda u razmatranom razdoblju.....	47
• Slika 15.: Kretanje prosječen cijene po noćenju u razmatranom razdoblju.....	48
• Slika 16.: Rezultati korelacijske analize.....	48
• Slika 17: Grafički prikaz povezanosti korištenih varijabli.....	49
• Slika 18.: Rezultati regresijske analize, $Y=UPE$ .....	50
• Slika 19.: Rezultati regresijske analize, $Y=N$ .....	50
• Slika 20.: Rezultati regresijske analize, $Y=PCPN$ .....	51
• Slika 21.: Rezultati regresijskog modela.....	52
• Slika 22.: Test normalnosti reziduala.....	53
• Slika 23.: rezultat testa heteroskedastičnosti.....	54
• Slika 24. Rezultat analize varijance.....	54
• Slika 25.: Dijagram raspršenosti.....	55
• Slika 26.: Rezultati testa linearnosti.....	55
• Slika 27.: Rezultati Ramsey RESET-a.....	56
• Slika 28.: rezultati procijenjenog modela.....	57
• Slika 29.: Rezultati testa normalnosti reziduala i heteroskedastičnosti.....	57
• Slika 30.: Rezultat analize varijance.....	58

- Slika 31.: Rezultati prognoze.....59

#### Popis Tablica:

- Tablica 1: Nelinearni i linearizirani regresijski modeli.....23
- Tablica 2.: Postupak analize i modeliranja podataka.....45
- Tablica 3.: Ukupni mjesečni prihodi u hotelu *Nautica* u 2023. godini.....60

#### Popis literature:

##### Knjige:

- Asteriou, D. and Hall, S. (2021): *Applied Econometrics 4th edn.* Bloomsbury Publishing
- Babu, M., Venkataramanaiah, M. and Jahnavi, M. (2013): *Forecasting with Econometric Methods 1st edn.*, Scholars' Press
- Baldigara, T. (2023.): *Ekonometrija: e-nastavni materijal*, Fakultet za menadžment u turizmu i ugostiteljstvu Opatija,
- Bernstein, Peter L. (1996.): *Against The Gods: The Remarkable Story Of Risk*, John Wiley & Sons, Inc
- Brooks, C. (2019): *Introductory Econometrics for Finance. 4th edn.*, Cambridge University Press
- Chatfield, C. (2000): *Time-Series Forecasting 1st edn.*, CRC Press
- Cohen, J. et al. (2013): *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences 3rd edn.*, Taylor and Francis
- Frechtling, D. (2012) *Forecasting Tourism Demand. 1st edn.* Taylor and Francis,
- Gujarati, D. (2017): *Econometrics by Example 2nd edn.*, Bloomsbury Publishing.
- Hayashi, F. (2011): *Econometrics*, Princeton University Press
- Hill, C., Griffiths, W. and Lim, G. (2018): *Principles of Econometrics 5th edn.*, Wiley
- Hoffmann, J. (2021) *Linear Regression Models 1st edn.*, CRC Press
- Miles, J. and Shevlin, M. (2000) *Applying Regression and Correlation 1st edn.* SAGE Publications
- Montgomery, D., Jennings, C. and Kulahci, M. (2015): *Introduction to Time Series Analysis and Forecasting 2nd edn.*, Wiley

- Sanders, N. (2016): *Forecasting Fundamentals*, Business Expert Press
- Weisberg, S. (2013): *Applied Linear Regression. 4th edn.*, Wiley
- Zax, J. (2011) *Introductory Econometrics. 1st edn.* Stanford University Press

Internet izvori:

- [https://bs.wikipedia.org/wiki/Koeficijent\\_korelacije#/media/Datoteka:Pearson\\_Correlation\\_Coefficient\\_and\\_associated\\_scatterplots.png](https://bs.wikipedia.org/wiki/Koeficijent_korelacije#/media/Datoteka:Pearson_Correlation_Coefficient_and_associated_scatterplots.png)
- <https://corporatefinanceinstitute.com/resources/data-science/autocorrelation/>

Popis priloga:

- Prilog 1.: Privola poslodavca .....66

MARKO GRČAR LOVRIC  
Anke Butorac 34  
52 440 Poreč  
OIB 66911031575

**PREDMET: Privola**

Kojom tvrtka SIGMA POSLOVODSTVO d.o.o., Sv. Antona 15, 52 466 Novigrad, OIB 21462841905, a u sklopu koje posluju Hotel i Marina Nautica, daje gornjem naslovu suglasnost u uvid i korištenje poslovnih podataka tvrtke, a u svrhu prikupljanja i korištenja istih za potrebe pisanja diplomskog rada.

Podaci koji će se prikupljati i obrađivati odnose se na:

- Broj gostiju
- Broj noćenja
- Dolaske/ odlaske

**\*\*NAPOMENA:** Privola se odnosi samo na navedene svrhe obrade i navedene kategorije poslovnih podataka te se obrada poslovnih podataka ne smije koristiti u druge svrhe.

SIGMA POSLOVODSTVO d.o.o.

  
\_\_\_\_\_

M.P.

SIGMA POSLOVODSTVO  
d.o.o.  
NOVIGRAD, Sv. Antona 15