Važnost prediktora u logističkom regresijskom modelu: primjena u istraživanju percepcije turizma lokalnog stanovništva

Vodanović, Jakov

PRIRODOSLOVNO–MATEMATIČKI FAKULTET
SVEUČILIŠTA U SPLITU

JAKOV VODANOVIĆ

Važnost prediktora u logističkom regresijskom modelu: primjena u istraživanju percepcije turizma lokalnog stanovništva

DIPLOMSKI RAD

Split, lipanj 2024.

PRIRODOSLOVNO–MATEMATIČKI FAKULTET
SVEUČILIŠTA U SPLITU

ODJEL ZA MATEMATIKU

Važnost prediktora u logističkom regresijskom modelu: primjena u istraživanju percepcije turizma lokalnog stanovništva

DIPLOMSKI RAD

Neposredna voditeljica:
dr. sc. Ana Perišić

Student:
Jakov Vodanović

Mentorica:
doc. dr. sc. Vesna Gotovac Đogaš

Split, lipanj 2024.

TEMELJNA DOKUMENTACIJSKA KARTICA

PRIRODOSLOVNO–MATEMATIČKI FAKULTET
SVEUČILIŠTA U SPLITU
ODJEL ZA MATEMATIKU

DIPLOMSKI RAD

Važnost prediktora u logističkom regresijskom modelu: primjena u istraživanju percepcije turizma lokalnog stanovništva

Jakov Vodanović

Sažetak:

Ovaj diplomski rad istražuje percepciju lokalnog stanovništva o utjecaju turizma na njihove živote i analizira relativnu važnost različitih varijabli vezanih uz utjecaj turizma na život lokalnog stanovništva. Relativna važnost varijabli ocjenjivat će se svojim utjecajem na ukupnu percepciju turizma. Koriste se anketni podatci prikupljeni u gradu Splitu. Statističke metode koje su korištene u ovom radu su izgradnja univarijatnih i multivarijatnih modela logističke regresije, provođenje „stepwise“ procedure za izgradnju modela, metoda analize dominance i metoda regresije koreliranih komponenti. Prvo se daje teorijski pregled navedenih metoda nakon čega slijedi njihova primjena. Kao ključne varijable koje najviše utječu na percepciju turizma identificirane su izgled grada, apartmanizacija i autentičnost grada. S druge strane, varijable iseljavanja iz centra i čistoće grada su se pokazale najmanje važnim pojavama.

Ključne riječi:

logistička regresija, analiza dominance, CCR

Podatci o radu:

TEMELJNA DOKUMENTACIJSKA KARTICA

broj stranica 42, broj slika 4, broj tablica 13, broj literaturnih navoda 9, jezik izvornika hrvatski

Mentorica: *doc. dr. sc. Vesna Gotovac Đogaš*

Neposredna voditeljica: dr. sc. Ana Perišić

Članovi povjerenstva:

doc. dr. sc. Vesna Gotovac Đogaš

dr. sc. Ana Perišić

dr. sc. Jelena Pleština

Povjerenstvo za diplomski rad je prihvatilo ovaj rad *1. srpnja 2024.*

BASIC DOCUMENTATION CARD

FACULTY OF SCIENCE, UNIVERSITY OF SPLIT
DEPARTMENT OF MATHEMATICS

MASTER'S THESIS

The Importance of Predictors in a Logistic Regression Model: Application in Analysis of Local Residents' Perception of Tourism

Jakov Vodanović

Abstract:

This thesis investigates the perceptions of local residents on the impact of tourism on their lives and analyzes the relative importance of variables related to the booming tourism sector. The relative importance of variables will be evaluated by their impact on the overall perception of tourism. Survey data was collected in the city of Split. The statistical methods employed in the thesis include the construction of univariate and multivariate logistic regression models, the application of the stepwise procedure for model building, dominance analysis, and the correlated component regression method. The study first provides a theoretical overview of these methods followed by their application. Key variables identified as having the greatest impact on the perception of tourism are city appearance, apartmentization and city authenticity. On the other hand, variables such as city center exodus and city cleanliness were found to be the least important.

Key words:

logistic regression, dominance analysis, CCR

Specifications:

42 pages, 4 images, 13 tables, 9 references, original in: Croatian

BASIC DOCUMENTATION CARD

Mentor: *professor Vesna Gotovac Đogaš*

Immediate mentor: *Ana Perišić, PhD*

Committee:

professor Vesna Gotovac Đogaš

Ana Perišić, PhD

assistant professor Jelena Pleština

This thesis was approved by a Thesis committee on 1^{st} *of July.*

Uvod

Turizam je jedna od najvažnijih gospodarskih grana u mnogim zemljama pa tako i u Hrvatskoj. Razvoj turizma donosi brojne ekonomske prednosti, ali također može imati značajne negativne utjecaje na lokalne zajednice, okoliš i društvo. Tema diplomskog rada je procjena relativne važnosti učinaka turizma iz perspektive lokalne zajednice kako bi se utvrdili učinci turizma koji predstavljaju faktore rizika za dobrobit lokalnog stanovništva. Takva analiza omogućuje prioritizaciju pokazatelja za praćenje društvene održivosti, te učinkovitije planiranje preventivnih ili korekcijskih mjera.

Statističke metode koje će se primijeniti u radu su izgradnja univarijatnih i multivarijatnih modela logističke regresije, provođenje „stepwise“ procedure za izgradnju modela, metoda analize dominance i metoda regresije koreliranih komponenti (eng. CCR). To su neki od razmatranih pristupa u ocjeni važnosti varijabli u logističkoj regresiji što je i glavni cilj teorijskog dijela rada.

U prvom, teorijskom dijelu rada obrazložit će se svaka od navedenih metoda. Nakon toga u drugom, empirijskom dijelu rada metode će se upotrijebiti za analizu relativne važnosti pojava vezanih uz razvoj turizma na lokalno stanovništvo. Relativna važnost pojava ocjenjivat će se svojim utjecajem na ukupnu percepciju turizma.

BASIC DOCUMENTATION CARD

Diplomski rad izrađen je na temelju rada Sever, I. i Perišić, A. (2024) u sklopu projekta Survey+ čiji je nositelj Institut za turizam.

Sadržaj

BASIC DOCUMENTATION CARD

Poglavlje 1

Logistička regresija

1.1 Uvodne definicije

Definicija 1.1. *Neka je $(\Omega, \mathcal{F})$ izmjeriv prostor i $\mathcal{P}$ skup vjerojatnosnih mjera na $(\Omega, \mathcal{F})$. Tada je uređena trojka $(\Omega, \mathcal{F}, \mathcal{P})$ statistička struktura.*

Definicija 1.2. *Slučajni uzorak na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je niz $X_1, X_2, \ldots, X_n$ slučajnih varijabli (vektora ili elemenata) na $(\Omega, \mathcal{F})$ takvih da su nezavisne i jednako distribuirane u odnosu na svaku vjerojatnost $P \in \mathcal{P}$.*

Definicija 1.3. *Statistika na statističkoj strukturi $(\Omega, \mathcal{F}, \mathcal{P})$ je svaka slučajna varijabla (vektor) $T : \Omega \to \mathbb{R}^d$ takva da postoji $n \in \mathbb{N}$ i slučajni uzorak $(X_1, \ldots, X_n)$ na $(\Omega, \mathcal{F}, \mathcal{P})$ te izmjerivo preslikavanje $t : \mathbb{R}^n \to \mathbb{R}^d$ takvo da je $T = t(X_1, \ldots, X_n)$.*

Može se pokazati da svaka funkcija gustoća inducira vjerojatnosnu mjeru pa je statističke strukture moguće zadati i skupom funkcija gustoće.

1.2 Univarijatna logistička regresija

Razni problemi, naročito u matematičkoj primjeni, mogu se modelirati regresijskim metodama u kojima postoji jedna zavisna varijabla i jedna ili više nezavisnih koje se koriste za predviđanje vrijednosti zavisne. Logistička regresija je tip regresije u kojem je zavisna varijabla dihotomna, tj. poprima vrijednosti 1 ili 0. Za početak valja se upoznati s univarijatnim modelom, tj. modelom s jednom nezavisnom varijablom koji će se kasnije proširiti na više varijabli. Uz pretpostavku dihotomnosti zavisne varijable, ona se može modelirati kao Bernoullijeva slučajna varijabla s vjerojatnošću uspjeha, tj. vjerojatnošću da je zavisna varijabla jednaka 1, definiranom na sljedeći način:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}, \tag{1.1}$$

gdje su β_0 i β_1 parametri koji se procjenjuju u logističkoj regresiji. Strogo govoreći, da bi definicija (1.1) imala smisla zavisnu varijablu Y treba promatrati kao slučajnu varijablu uvjetno na realizaciju x slučajne varijable X, odnosno $\mathbb{E}[Y|X = x] = \pi(x)$. To neće biti strogo naznačeno dalje u radu nego će se zavisna varijabla uvjetno na realizaciju označavati samo s Y. Nadalje, zasad u univarijatnom modelu nije dozvoljena nedihotomna kategorijalna varijabla koja nije intervalna, razlog za to bit će objašnjen u konstrukciji multivarijatnog modela.

Ponekad će biti od koristi promatrati logit transformaciju, koja je afinog oblika. Slijedi definicija.

Definicija 1.4 (Logit transformacija za univarijatni model)**.** *Logit transformacija definirana je na sljedeći način:*

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x.$$

1.3 Procjena parametara modela univarijatne logističke regresije

Za procjenu parametara modela logističke regresije najčešće se koristi procjenitelj maksimalne vjerodostojnosti (eng. Maximum Likelihood Estimator, akronim MLE). Slijedi definicija.

Definicija 1.5. *Neka je* $\mathbb{X} = (X_1, \ldots, X_n)$ *slučajni uzorak iz modela* $\mathcal{P} = \{f(\cdot;\theta) : \theta \in \Theta\}$. *Ako je* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *realizacija slučajnog uzorka* $\mathbb{X}$, *tada je vjerodostojnost funkcija* $L : \Theta \to \mathbb{R}$ *definirana kao*

$$L(\theta \mid \boldsymbol{x}) = L(\theta) := f_{\mathbb{X}}(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

Definicija 1.6. *Statistika* $\hat{\theta} = \hat{\theta}(\mathbb{X})$ *je procjenitelj maksimalne vjerodostojnosti za* θ *(MLE) ako vrijedi*

$$L(\hat{\theta} \mid \mathbb{X}) = \max_{\theta \in \Theta} L(\theta \mid \mathbb{X}).$$

Za dani slučajni uzorak $\mathbb{X} = (X_1, \ldots, X_n)$ neka je dana funkcija $\ell_n : \Theta \to \mathbb{R}$:

$$\ell_n(\theta) = \ln L(\theta \mid \mathbf{x}) := \sum_{i=1}^{n} \ln f(x_i;\theta),$$

Ova funkcija se naziva log-vjerodostojnost i često je lakše raditi s njom nego s funkcijom vjerodostojnosti, a budući da je logaritam strogo rastuća diferencijabilna funkcija, traženje globalnog ekstrema funkcije vjerodostojnosti može se svesti na traženje globalnog ekstrema funkcije log-vjerodostojnosti.

Primjerice za propoziciju koja slijedi korisno je promotriti log-vjerodostojnost za parametre logističke regresije (β_0, β_1). Funkcija gustoće Bernoullijeve

slučajne varijable Y za koju je $\mathbb{E}[Y|X=x]=\pi(x)$ je

$$f_Y(y)=\begin{cases}\pi(x)^y\left[1-\pi(x)\right]^{1-y},\ y\in\{0,1\}\\ 0,\ \text{inače.}\end{cases}$$

Tada je funkcija vjerodostojnosti zadana s $L(\beta_0,\beta_1|(\mathbf{x},\mathbf{y}))=\prod_{i=1}^n\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}$ jer su varijable Y_i, $i=1,\ldots,n$ međusobno nezavisne.

Sada se može promotriti log-vjerodostojnost:

$$\begin{aligned}l(\beta_0,\beta_1)&=\ln(L(\beta_0,\beta_1\mid\mathbf{x}))=\sum_{i=1}^n y_i\ln(\pi(x_i))+(1-y_i)\ln(1-\pi(x_i))\\&=\sum_{i=1}^n y_i\ln\left(\frac{e^{\beta_0+\beta_1x}}{1+e^{\beta_0+\beta_1x}}\right)+(1-y_i)\ln\left(1-\frac{e^{\beta_0+\beta_1x}}{1+e^{\beta_0+\beta_1x}}\right)\\&=\sum_{i=1}^n y_i\left((\beta_0+\beta_1x)-\ln(1+e^{\beta_0+\beta_1x})\right)+(1-y_i)\ln\left(\frac{1}{1+e^{\beta_0+\beta_1x}}\right)\\&=\sum_{i=1}^n y_i\left((\beta_0+\beta_1x)-\ln(1+e^{\beta_0+\beta_1x})\right)+(1-y_i)\left(\ln(1)-\ln(1+e^{\beta_0+\beta_1x})\right)\\&=\sum_{i=1}^n y_i\left((\beta_0+\beta_1x)-\ln(1+e^{\beta_0+\beta_1x})\right)-(1-y_i)\ln(1+e^{\beta_0+\beta_1x})\\&=\sum_{i=1}^n y_i(\beta_0+\beta_1x_i)-\ln(1+e^{\beta_0+\beta_1x_i}).\end{aligned}$$

Propozicija 1.7. *Neka je dan slučajni uzorak* $(\mathbb{X},\mathbb{Y})=((X_1,Y_1),\ldots,(X_n,Y_n))$ *i neka njegova realizacija* $(\boldsymbol{x},\ \boldsymbol{y})=((x_1,y_1),\ldots,(x_n,y_n))$. *Procjenitelji maksimalne vjerodostojnosti* $(\hat\beta_0,\hat\beta_1)$ *za parametre* (β_0,β_1) *u modelu logističke regresije ispunjavaju jednadžbe*

$$\sum_{i=1}^n\left[y_i-\pi(x_i)\right]=0 \tag{1.2}$$

i

$$\sum_{i=1}^n x_i\left[y_i-\pi(x_i)\right]=0. \tag{1.3}$$

Dokaz. Kao što je pokazano, log-vjerodostojnost je:

$$l(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0+\beta_1 x_i})$$

Parcijalne derivacije log-vjerodostojnosti su:

$$\frac{\partial l}{\partial \beta_0}(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \frac{1}{1 + e^{\beta_0+\beta_1 x_i}} \cdot e^{\beta_0+\beta_1 x_i} \right) = \sum_{i=1}^{n} (y_i - \pi(x_i))$$

$$\frac{\partial l}{\partial \beta_1}(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i \cdot x_i - \frac{1}{1 + e^{\beta_0+\beta_1 x_i}} \cdot e^{\beta_0+\beta_1 x_i} \cdot x_i \right) = \sum_{i=1}^{n} x_i \left(y_i - \pi(x_i) \right)$$

Kako je $(\hat{\beta}_0, \hat{\beta}_1)$ MLE za (β_0, β_1) to je on globalni maksimum pa tako i stacionarna točka, odnosno za $(\hat{\beta}_0, \hat{\beta}_1)$ parcijalne derivacije su jednake nula. Izjednačavanje parcijalnih derivacija s nulom daje jednadžbe (1.2) i (1.3), tj. MLE ispunjava te jednadžbe. QED

Rješavanje sustava jednadžbi (1.2) i (1.3) za dane realizacije slučajnih uzoraka zahtijeva posebne iterativne metode zbog definicije funkcije π. Rješavanje ta dva sustava daje procjene parametara preko MLE metode. Za dani niz realizacija neka su MLE procjene parametara (β_0, β_1) označene s $(\hat{\beta}_0, \hat{\beta}_1)$ te neka je procjena parametra $\pi(x)$ označena s $\hat{\pi}(x)$.

1.4 Test značajnosti nezavisne varijable u univarijatnom modelu

Najčešće se prije ocjene prilagodbe modela promatra značajnost nezavisnih varijabli u modelu uz izračunate procjene parametara. Za pravilno razmatranje modela prikazat će se statistički testovi kojima će se ispitati značajnost nezavisne varijable u modelu. Test će se, zasad, prikazati za

univarijatni model, ali će se u daljnjim poglavljima metode proširiti na multivarijatni model.

Jedna od metoda testiranja je izgradnja modela preko metode MLE-a uz uključenu nezavisnu varijablu i izgradnja modela bez te varijable te usporedba tih dvaju modela. Da bi se modeli uopće mogli usporediti potrebno je imati metodu ocjene prilagođenosti kojom bi se usporedili modeli. Važno je naglasiti da ova ocjena nije ocjena prilagođenosti modela kao takvog (ne radi se o „goodness-of-fit“ testu), nego služi samo za usporedbu ovih dvaju modela. Model koji će biti mjerilo usporedbe naziva se zasićeni model.

Zasićeni model je model u kojem je broj parametara koji se procjenjuju jednak broju realizacija, primjerice u linearnoj regresiji ako su dane samo dvije realizacije model će biti pravac koji prolazi tim točkama zbog čega će model biti savršeno prilagođen. Na sličan način vrijedi da se zasićeni model logističke regresije najbolje moguće prilagođava danim podatcima, tj. vjerodostojnost zasićenog modela je maksimalna (u ovom slučaju jednaka 1 zbog dihotomnosti zavisne varijable). Sa zasićenim modelom uspoređuju se modeli sa uključenom nezavisnom varijablom i onaj bez nje kako bi se dobila ocjena njihove prilagođenosti te se te ocjene onda uspoređuju međusobno. Da bi se ta ocjena prilagođenosti kvantificirala i iz nje konstruirao test koristi se devijanca modela koja je definirana na sljedeći način:

$$D = -2\ln\left[\frac{(\text{vjerodostojnost procijenjenog modela})}{(\text{vjerodostojnost zasićenog modela})}\right]. \tag{1.4}$$

Izraz u zagradi u formuli (1.4) naziva se omjer vjerodostojnosti. Valja napomenuti da se u ovom radu promatra samo dihotomni model gdje vrijednosti zavisne varijable mogu biti ili 0 ili 1 zbog čega je vjerodostojnost za zasićeni model koji savršeno predviđa jednaka 1. Devijanca se interpretira na

sljedeći način: Ako je „blizu“ 0 to znači da je vjerodostojnost procijenjenog modela blizu maksimalne vjerodostojnosti, tj. da su procijenjene vrijednosti zavisne varijable „blizu“ realizacija. Za procijenjeni model s nezavisnom varijablom devijanca je

$$\begin{aligned} D &= -2\ln(L(\hat{\beta}_0, \hat{\beta}_1 \mid \mathbf{x})) = \\ &= -2\sum_{i=1}^{n} \left[y_i \ln(\hat{\pi}(x_i)) + (1-y_i)\ln(1-\hat{\pi}(x_i)) \right]. \end{aligned}$$

Neka je $n_1 = \sum_{i=1}^{n} y_i$ i $n_0 = \sum_{i=1}^{n}(1-y_i)$ te $n = n_0 + n_1$, tj. n_1 je broj jedinica zavisne varijable za danu realizaciju, a n_0 broj nula. Model bez nezavisne varijable će biti Bernoullijev model s vjerojatnošću uspjeha $\frac{n_1}{n}$ što je i MLE za parametar vjerojatnosti uspjeha. Vjerodostojnost takvog modela je $\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}$.

Sada je moguće usporediti devijance modela sa zavisnom varijablom i bez nje kako bi se ocijenila važnost nezavisne varijable. Način na koji se vrši ta usporedba je promjena u vrijednosti devijance koja je uzrokovana uključivanjem nezavisne varijable u model. Ta promjena koristi se kao ocjena važnosti nezavisne varijable, označava se s T i definira se na sljedeći način:

$$T = D(\text{model bez varijable}) - D(\text{model s varijablom}),$$

tj.

$$\begin{aligned} T &= -2\ln\left[\frac{\text{vjerodostojnost modela bez varijable}}{\text{vjerodostojnost modela s varijablom}}\right] = \\ &= -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n}(\hat{\pi}(x_i))^{y_i}(1-\hat{\pi}(x_i))^{1-y_i}}\right] = \\ &= 2\left\{\sum_{i=1}^{n}\left[y_i\ln(\hat{\pi}(x_i)) + (1-y_i)\ln(1-\hat{\pi}(x_i))\right] - \left[n_1\ln(n_1) + n_0\ln(n_0) - n\ln(n)\right]\right\}. \end{aligned}$$

Uz pretpostavku istinitosti nulte hipoteze koja pretpostavlja da je β_1 jednaka nuli, tj. da je nezavisna varijabla potpuno nevažna, statistika T (valja naglasiti da ako se T promatra kao statistika onda se na mjestima realizacija nalaze slučajne varijable) se ponaša kao χ^2 distribucija s jednim stupnjem slobode. Postoje i druge pretpostavke, ali sve su ispunjene asimptotski, odnosno uz dovoljno velik uzorak. Time je dovršena izgradnja testa značajnosti nezavisne varijable za univarijatni model. Ostaje proširiti model na multivarijatni.

1.5 Proširenje na multivarijatni model

Proširenje na multivarijatni model će se napraviti promjenama u definicijama univarijatnog modela, ali interpretacija tih definicija će najčešće ostati ista. Ideja multivarijatnog modela je da se omogući veći broj nezavisnih varijabli koje predviđaju zavisnu. Neka je $\mathbf{x} = (x_1, \ldots, x_p)$ vektor realizacija nezavisnih varijabli koje predviđaju zavisnu Y. Kao i u slučaju univarijatnog modela zavisna varijabla može se modelirati kao Bernoullijeva slučajna varijabla s vjerojatnošću uspjeha definiranom na sljedeći način:

$$\pi(\mathbf{x}) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p}}, \tag{1.5}$$

gdje su $\beta_0, \beta_1, \ldots, \beta_p$ parametri koji se procjenjuju u logističkoj regresiji. Kao i u univarijatnom slučaju da bi definicija (1.5) imala smisla zavisnu varijablu Y treba promatrati kao slučajnu varijablu uvjetno na realizaciju $\mathbf{x}$ slučajnog vektora $\mathbf{X}$, odnosno $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}] = \pi(\mathbf{x})$. Kao i ranije to neće biti strogo naznačeno dalje u radu nego će se zavisna varijabla uvjetno na realizaciju označavati samo s Y.

Definicija 1.8 (Logit transformacija za multivarijatni model)**.** *Logit tran-*

sformacija definirana je na sljedeći način:

$$g(x) = \ln\left[\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right] = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

Kao i u univarijatnom slučaju, funkcija $\pi : \mathbb{R}^n \to \mathbb{R}$ je uvjetna vjerojatnost da je zavisna varijabla jednaka 1 uz danu realizaciju nezavisnih varijabli

$$\pi(\mathbf{x}) = \mathbb{P}(Y|\mathbb{X} = \mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}.$$

Multivarijatni model je omogućio promatranje kategorijalnih varijabli koje nisu intervalne ili dihotomne. To prije nije bilo moguće jer takve varijable nema smisla kodirati u jednoj varijabli. Razlog za to je što je onda nametnuta intervalnost varijabli koja nije intervalna. Primjerice, ukoliko se promatra kategorijalna varijabla na nominalnoj skali (kao što je rasa, narodnost, marka proizvoda itd.) ili ordinalna neintervalna varijabla (kao što je stupanj obrazovanja), takve varijable moguće je kodirati brojevima, ali nema smisla promatrati razlike među tim brojevima. Zbog toga je potrebno definirati tzv. „dummy“ varijable. Neka je nezavisna varijabla X_j kategorijalna za neki $j = 1, 2, ..., p$. Neka varijabla X_j poprima $k \in \mathbb{N}$ različitih vrijednosti: $L_1, L_2, ..., L_k$. U tom slučaju potrebno je uvesti $k-1$ „dummy“ varijabli. U ovom slučaju je logit multivarijatnog modela dan izrazom

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \sum_{l=1}^{k-1} \beta_{jl} D_{jl} + \ldots + \beta_p x_p$$

pri čemu su D_{jl} „dummy“ varijable konstruirane na sljedeći način:

$$D_{jl} = \begin{cases} 1, & x_j = L_l \text{ za } l = 2, \ldots, k \\ 0, & \text{inače} \end{cases}$$

Procjene parametara će se vršiti metodom maksimalne vjerodostojnosti.

Propozicija 1.9. *Neka je dan slučajni uzorak* $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$ *te neka njegova realizacija* $(x_{i1}, \ldots, x_{ip}, y_i)$, $i = 1, \ldots, n$. *Procjenitelji maksimalne vjerodostojnosti* $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ *ispunjavaju niz jednadžbi*

$$\sum_{i=1}^{n} [y_i - \pi(\boldsymbol{x}_i)] = 0 \tag{1.6}$$

i

$$\sum_{i=1}^{n} x_{ij} [y_i - \pi(\boldsymbol{x}_i)] = 0. \tag{1.7}$$

za $j = 1, \ldots, p$.

Ova propozicija proširenje je Propozicije 1.7 i dokazuje se analogno. Za dani niz realizacija neka su MLE procjene parametara $(\beta_0, \beta_1, \ldots, \beta_p)$ označene s $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ te neka je procjena parametra $\pi(\mathbf{x})$ označena s $\hat{\pi}(\mathbf{x})$. U multivarijatnom modelu značajnost modela moguće je ispitati usporedbom devijanci modela sa svim uključenim varijablama i modela bez varijabli. Značajnost pojedine nezavisne varijable može se ispitati usporedbom modela sa uključenom varijablom od interesa i modela bez te varijable.

1.6 Ocjene prilagodbe modela

1.6.1 Pearsonova statistika i devijacija

Nakon ispitivanja značajnosti nezavisnih varijabli ima smisla pitati se koliko je model uspješan u predviđanju zavisne varijable. U svrhu ispitivanja prilagodbe modela koriste se testovi prilagodbe (eng. goodness-of-fit).

Za ocjenu prilagodbe modela i sastavljanje relevantnih testova prilagodbe važan je broj različitih realizacija nezavisne varijable (eng. covariate pattern). Neka je broj različitih realizacija jednak J i neka je frekvencija j-te

realizacije za sve $j = 1, \ldots, J$ označena s m_j. Tada vrijedi $\sum_{j=1}^{J} m_j = n$ gdje je n duljina uzorka (ukupan broj realizacija).

Ako u modelu postoje neprekidne varijable zbog čega je broj realizacija jako velik onda vrijedi $n \approx J$, odnosno ponavljanje rezultata će biti rijetko zato što je u model uključena neprekidna varijabla. Statistički rezultati u ovom slučaju će biti n-asimptotski. S druge strane, ako se u modelu promatraju samo kategorijalne varijable broj različitih realizacija neće biti velik i za jako velik broj realizacija broj m_j će također znatno rasti. Statistički rezultati u tom slučaju će biti m-asimptotski (Za detaljniji prikaz vidjeti Hosmer, D. W. i Lemeshow, S. i Sturdivant, R. (2013)).

Neka je $\hat{y}_j$ zbroj svih predviđenih vrijednosti, a y_j zbroj svih stvarnih realiziranih vrijednosti zavisne varijable za realizaciju nezavisnih varijabli $\mathbf{x}_j$. Testiranje će se provoditi promatranjem razlika stvarnih vrijednosti i predviđenih vrijednosti zavisne varijable za j-tu realizaciju $(y_j - \hat{y}_j)$. Konkretnije,

$$\hat{y}_j = m_j \hat{\pi}_j = m_j \frac{e^{\hat{g}(x_j)}}{1 + e^{\hat{g}(x_j)}}.$$

Pearsonov rezidual tada je jednak:

$$r(y_j, \hat{\pi}_j) = \frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}.$$

Pomoću tog reziduala definira se Pearsonova χ^2 statistika:

$$X^2 = \sum_{j=1}^{J} r(y_j, \hat{\pi}_j)^2.$$

Ako je y_j promatrana kao slučajna varijabla onda je Pearsonov rezidual asimptotski normalno distribuirana slučajna varijabla uz pretpostavku binomne distribucije varijable Y_j. Uz pretpostavku ispravnosti modela onda je

X^2 slučajna varijabla koja prati χ^2 distribuciju s $J-(p+1)$ stupnjeva slobode.

Osim Pearsonovog reziduala često se koristi i rezidual devijance:

$$d(y_j, \hat{\pi}_j) = \pm\sqrt{2\left[y_j \ln\left(\frac{y_j}{m_j\hat{\pi}_j}\right) + (m_j - y_j)\ln\left(\frac{m_j - y_j}{m_j(1-\hat{\pi}_j)}\right)\right]},$$

gdje je predznak jednak predznaku izraza $(y_j - m_j\hat{\pi}_j)$. Kada je $y_j = 0$, rezidual je definiran s:

$$d(y_j, \hat{\pi}_j) = -\sqrt{2m_j \ln(1-\hat{\pi}_j)},$$

a kada je $y_j = m_j$ rezidual je definiran s:

$$d(y_j, \hat{\pi}_j) = \sqrt{2m_j \ln(\hat{\pi}_j)}.$$

Sada je moguće definirati i statistiku:

$$D = \sum_{j=1}^{J} d(y_j, \hat{\pi}_j)^2,$$

koja, kao i Pearsonova χ^2 statistika, uz pretpostavku ispravnosti modela prati χ^2 distribuciju s J - p - 1 stupnjeva slobode.

U slučaju kada je $J \approx n$, promatrane statistike neće biti $\chi^2(J-p-1)$ distribuirane i stoga ih se ne može koristiti za ocjenu prilagodbe modela. U takvim slučajevima primjenjuju se modificirani testovi kao što je npr. Hosmer-Lemenshow test. Takvi modificirani testovi većinom se provode uz prethodno grupiranje podataka. Na primjer, u Hosmer-Lemenshow testu podaci se grupiraju prema procijenjenim vjerojatnostima. Na istraživaču je odrediti broj grupa, a računalni programi često grupiranje provode uz automatski postavljeno g = 10 grupa (za više informacija vidjeti Hosmer, D. W. i Lemeshow, S. i Sturdivant, R. (2013)).

1.6.2 R^2 za logističku regresiju

U ocjeni prilagodbe modela linearne regresije koeficijent determinacije R^2 je poznata i često korištena ocjena. Međutim, model logističke regresije zahtijeva drugačiju definiciju te ocjene zbog specifičnosti modela. Kako bi R^2 bila prikladno definirana postavljena su četiri zahtjeva koja osiguravaju kvalitetu ocjene:

1) ograničenost: ocjena treba biti ograničena nulom odozdo i jedinicom odozgo gdje 0 označava potpunu neprilagođenost modela, a 1 savršenu prilagođenost;

2) invarijatnost na linearne transformacije: ocjena treba biti invarijantna s obzirom na nesingularne (primjer singularne bilo bi množenje s nulom) linearne transformacije zavisne ili nezavisnih varijabli;

3) monotonost: dodavanje nezavisnih varijabli ne bi trebalo smanjivati ocjenu prilagodbe;

4) intuitivna interpretabilnost: ocjena mora biti interpretabilna (na način na koji je to R^2 u slučaju linearne regresije).

Slijede primjeri četiri ocjene prilagodbe logističkog modela koje su analogne R^2 u linearnoj regresiji, a svaka od njih ispunjava barem tri zahtjeva. Neka je L_M vjerodostojnost procijenjenog modela, a L_0 vjerodostojnost modela bez nezavisnih varijabli. Ocjene su sljedeće:
McFadden (1974.):

$$R^2_M = \frac{\ln(L_0) - \ln(L_M)}{\ln(L_0)} = 1 - \frac{\ln(L_M)}{\ln(L_0)}$$

Nagelkerke (1991.):

$$R_N^2 = \frac{1 - \left(\frac{L_0}{L_M}\right)^{\frac{2}{n}}}{1 - (L_0)^{\frac{2}{n}}}$$

Estrella (1998.):

$$R_E^2 = 1 - \left[\frac{\ln(L_M)}{\ln(L_0)}\right]^{-\frac{2\ln(L_0)}{n}}$$

Zheng & Agresti (2000.): ova ocjena je kvadrirana korelacija predviđenih i stvarnih vrijednosti zavisne varijable i označava se sa $R^2_{y,\hat{y}}$.

Veće vrijednosti ovih pokazatelja ukazuju na bolju prilagodbu modela.

1.6.3 Druge ocjene prilagođenosti modela

U ovom odjeljku prikazat će se i nekoliko indikatora prediktivne moći modela. Budući da je zavisna varijabla dihotomna, lako je tablično prikazati usporedbu stvarnih i predviđenih vrijednosti kao što je prikazano u tablici 1.1. Uobičajeno je podijeliti podatke na skup podataka za treniranje modela („training set“) i skup podataka za ocjenu prediktivne moći („test set“). Za ocjenu prediktivne moći modela koristit će se indikatori prikazani u tablici 1.2.

Pozitivni rezultati u idućoj tablici označavaju uspjeh, tj. jedinicu kao vrijednost zavisne varijable, a negativni neuspjeh, tj. nulu.

Ovih pet ocjena ima različite ciljeve. Točnost daje informaciju o udjelu ispravno prepoznatih instanci u cijelom skupu podataka, međutim indikator točnosti osjetljiv je na nebalansirane skupove podataka, primjerice na varijable koje su iznimno često pozitivne ili iznimno često negativne: tako će trivijalni model koji za bilo koje vrijednosti zavisnih varijabli predviđa

	Stvarno pozitivni (P)	Stvarno negativni (N)
Predviđeni pozitivni rezultati	Točni pozitivni (TP)	Pogrešni pozitivni (FP)
Predviđeni negativni rezultati	Pogrešni negativni (FN)	Točni negativni (TN)

Tablica 1.1: Konfuzijska matrica

Ocjena	Formula
Točnost	$\frac{TP+TN}{P+N}$
Osjetljivost	$\frac{TP}{P}$
FP stopa	$\frac{FP}{N}$
Specifičnost	$\frac{TN}{N}$
Preciznost	$\frac{TP}{TP+FP}$

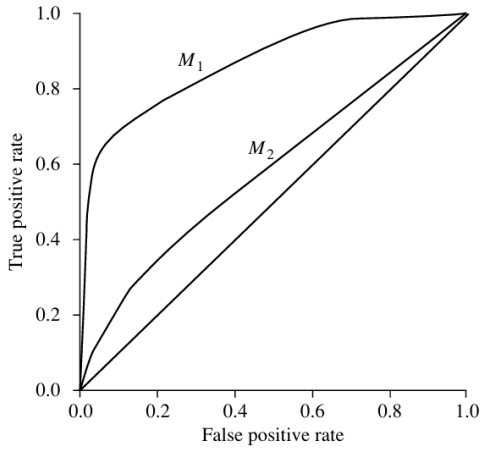
Tablica 1.2: Indikatori prediktivne moći

negativan rezultat imati jako visoku točnost za varijablu u kojoj je negativan rezultat iznimno čest. Stoga se definiraju osjetljivost i specifičnost koje rješavaju taj problem jer one mjere točnost isključivo u kontekstu pozitivnih ili negativnih vrijednosti zavisne varijable. Preciznost mjeri koliko je, od svih predviđenih pozitivnih rezultata, zapravo pozitivnih.

Još jedna često korištena mjera, u oznaci AUC (eng. Area Under the Curve), je površina ispod ROC (eng. Receiver operating characteristic) krivulje. Kao što je već komentirano, u dihotomnoj logističkoj regresiji modeli zapravo procjenjuju vjerojatnosti da je zavisna varijabla, za dane vrijednosti nezavisne, jednaka 1. Za konačnu klasifikaciju uzima se unaprijed određeni prag i najčešći odabir je 0,5, što ima najviše smisla. Dakle ako je modelom procijenjena vrijednost zavisne varijable jednaka 0,6 onda se konačna

predviđena vrijednost zavisne varijable uzima kao 1. Ali taj prag može biti varijabilan i to utječe na broj TP, FP, TN i FN, a time i na ocjene „prediktivne moći“ modela.

ROC krivulja dobiva se grafičkim prikazom osjetljivosti i FP stope kao funkcije varijabilnog praga od 0 do 1. Ako je prag 0 onda se sve vrijednosti klasificiraju kao jedinice i osjetljivost i FP stopa su 1 jer je sve klasificirano kao pozitivno, analogno je jasno da su obe vrijednosti 0 ako je prag 1. Primjer dviju ROC krivulja za dva modela, M_1 i M_2, prikazan je na slici 1.1.



Slika 1.1: ROC krivulja

Što je površina ispod krivulje veća to je model bolji. Prednost ove ocjene nad drugima je što uzima sve pragove u obzir.

Prema Hosmer, D. W. i Lemeshow, S. i Sturdivant, R. (2013):

- Ako je AUC ≤ 0.5:
  to sugerira da model predviđa lošije ili isto od modela koji predviđa na slučajan način.

- Ako je $0.7 \leq \text{AUC} < 0.8$:
  to sugerira da model prihvatljivo predviđa.

- Ako je $0.8 \leq \text{AUC} < 0.9$:
  to sugerira da model izvrsno predviđa.

- Ako je $\text{AUC} \geq 0.9$:
  to sugerira da model izvanredno predviđa.

Obrađene su često korištene ocjene prilagođenosti modela. Sada će se posvetiti pažnja ocjeni važnosti varijabli u modelu multivarijatne logističke regresije. Posebno će se promotriti dvije metode: analiza dominance i standardizacija koeficijenata.

Poglavlje 2

Ocjena važnosti varijabli u modelu multivarijatne logističke regresije

U ovom odjeljku pokazat će se nekoliko pristupa u ocjeni važnosti varijabli u logističkoj regresiji što je i glavni cilj teorijskog dijela rada.

2.1 Analiza dominance

Analiza dominance metoda je koja se temelji na usporedbi koeficijenta determinacije R^2 modela sa i bez uključene nezavisne varijable od interesa, odnosno varijable čiju važnost treba ocijeniti. Pri tome se promatra promjena koeficijenta determinacije u svim mogućim podmodelima. Podmodel glavnog modela je svaki model koji je izgrađen pomoću proizvoljnog pravog podskupa nezavisnih varijabli glavnog modela (primjerice, ako je glavni model izgrađen pomoću nezavisnih varijabli X_1, X_2 i X_3, jedan od podmodela je model izgrađen pomoću nezavisnih varijabli X_1 i X_3). Level podmodela je

broj nezavisnih varijabli uključenih u podmodel.

Neka je s $R^2_{X_j}$ označen koeficijent determinacije modela u kojem je X_j nezavisna varijabla. Oznaka se prirodno proširuje ako je nezavisnih varijabli u modelu više (primjerice za model s nez. var. X_1 i X_2 oznaka je $R^2_{X_1X_2}$). Za potrebe analize dominance nebitno je je li koeficijent determinacije koji se koristi R^2_M, R^2_N ili R^2_E jer je pokazano da svi koeficijenti dovode do istih zaključaka. U praktičnom dijelu rada koristit će se McFaddenov koeficijent determinacije.

Jedna varijabla potpuno dominira (eng. complete dominance) drugu ako dodavanjem te varijable koeficijent determinacije modela naraste više nego dodavanjem druge u svim podmodelima glavnog modela. Primjerice, ako je model sastavljen od četiri nezavisne varijable X_1, X_2, X_3 i X_4 onda se u ispitivanju dominira li varijabla X_3 ili X_4 gleda koji je od sljedećih brojeva veći: $R^2_{X_1,X_2,X_3} - R^2_{X_1,X_2}$ ili $R^2_{X_1,X_2,X_4} - R^2_{X_1,X_2}$; na analogan način bi to trebalo napraviti za sve podmodele. Ako jedna varijabla ne dominira potpuno drugu, moguće je promatrati prosjeke povećanja koeficijenta determinacije posebno za modele različitih levela. Ako su svi ti prosjeci veći za varijablu X_i nego što su za X_j tada varijabla X_i uvjetno dominira (eng. conditional dominance) X_j. Ako se ni takva dominacija ne može uspostaviti, moguće je promatrati prosjek povećanja koeficijenta determinacija u svim podmodelima i kaže se da jedna varijabla općenito dominira (eng. general dominance) nad drugom u slučaju većeg prosjeka.

Kako bi se donio zaključak o najvažnijim varijablama, promatra se koliko pojedina varijabla dominira nad drugim varijablama i o kojoj se dominaciji

radi. Ako jedna varijabla potpuno dominira nad svim ostalima i koeficijent determinacije je visok, može se zaključiti da se radi o iznimno važnoj varijabli, a ako neku varijablu potpuno dominiraju sve ostale onda su te ostale varijable relevantnije od nje.

2.2 Standardizirani koeficijenti logističke regresije

Važnost varijabli moguće je ocijeniti i usporediti i usporedbom vrijednosti procijenjenih koeficijenta. Međutim, problem direktne usporedbe procijenjenih koeficijenata kako bi se usporedila važnost nezavisnih varijabli je što su različite varijable najčešće u neusporedivim mjernim jedinicama. Ideja standardizacije koeficijenata je da se to ispravi kako bi se mogla vršiti direktna usporedba koeficijenata koji su standardizirani, odnosno u istom mjerilu. U obzir se uzimaju standardne devijacije varijabli budući da se promatra utjecaj promjene nezavisne varijable na ukupnu promjenu u zavisnoj. Standardizirani koeficijenti koji u izračun uzimaju standardnu devijaciju pripadne nezavisne varijable nazivaju se djelomično standardizirani, a oni koji koriste i standardnu devijaciju zavisne varijable nazivaju se potpuno standardiziranim koeficijentima.

Djelomično standardizirana su sljedeća tri koeficijenta:

Agresti (1996.):

$$b_A^* = (b)(s_X),$$

gdje je b procjena koeficijenta za nezavisnu varijablu X, a s_X je standardna

devijacija te nezavisne varijable. Druga procjena:

$$b_S^* = (b)\left(\frac{s_X}{\sqrt{\pi/3}}\right).$$

Long (1997.):

$$b_L^* = (b)\left(\frac{s_X}{\pi/\sqrt{3}+1}\right).$$

Uz empirijsku procjenu standardne devijacije dobiva se idući koeficijent (koji je potpuno standardiziran).

Menard (1995.):

$$b_M^* = (b)\,(s_X)\,(R)\,/s_{\hat{g}(\mathbf{x}))},$$

gdje je R korijen koeficijenta determinacije, a $s_{\hat{g}(\mathbf{x}))}$ je standardna devijacija procijenjenog logita.

2.3 CCR

Osnovna ideja CCR ili „Correlated Component Regression“ metode je da se originalne nezavisne varijable zamijene komponentama koje su linearne kombinacije originalnih nezavisnih varijabli. Metoda se prvo pokazuje za linearnu regresiju nakon čega će se proširiti na logističku.

Prva komponenta S_1 gradi se na sljedeći način: Neka je $\lambda_g^{(1)}$ za svaku nezavisnu varijablu $g = 1, 2, \ldots, P$ regresijski koeficijent u univarijatnoj linearnoj regresiji za zavisnu varijablu Y i nezavisnu X_g. Zatim se S_1 definira kao ponderirani prosjek svih nezavisnih varijabli gdje su ponderi $\lambda_g^{(1)}$:

$$S_1 = \frac{1}{P}\sum_{g=1}^{P}\lambda_g^{(1)}X_g$$

2.3. CCR

Metodom najmanjih kvadrata procjenjuju se linearni regresijski modeli za zavisnu varijablu Y i nezavisne S_1 i X_g za svaki $g = 1, 2, \ldots, P$ te se definira $\lambda_g^{(2)}$ kao vrijednost regresijskog koeficijenta varijable X_g. Tada je

$$S_2 = \frac{1}{P} \sum_{g=1}^{P} \lambda_g^{(2)} X_g.$$

Za izgradnju treće komponente, S_3, procjenjuju se linearni regresijski modeli za zavisnu varijablu Y i nezavisne S_1 i S_2 i X_g za svaki $g = 1, 2, \ldots, P$ te se definira $\lambda_g^{(3)}$ kao vrijednost regresijskog koeficijenta varijable X_g. Tada je

$$S_3 = \frac{1}{P} \sum_{g=1}^{P} \lambda_g^{(3)} X_g.$$

Na isti način grade se iduće komponente. Komponenata može maksimalno biti P, odnosno ne može prijeći broj nezavisnih varijabli jer u slučaju $K = P$ model postaje multivarijatni regresijski model sa svim uključenim nezavisnim varijablama (za dokaz vidjeti Magidson (2013)). Varijable koje su značajne u univarijatnim modelima (odnosno u prvoj komponenti) zovu se primarne varijable (eng. prime variables), a one koje nisu značajne u univarijatnom modelu, ali postaju značajne u sljedećim komponentama nazivaju se supresorske varijable (eng. suppressor variables). Valja naglasiti da postoje različiti pristupi definiciji supresorskih varijabli. Ali, u osnovi, supresorska varijabla je varijabla koja povećava prediktivnu sposobnost modela, odnosno druge varijable (ili varijabli) kada je uključena u regresijski model.

Nakon što se odabere broj komponenti K i svaka komponenta se izračuna, provodi se linearna regresija za zavisnu varijablu Y i nezavisne $S_1, \ldots, S_L$ i tako se dobivaju regresijski koeficijenti $b_k^{(K)}$ za sve $k = 1, \ldots, K$ za varijablu S_k. Konačni koeficijent za varijablu X_g je $\beta_g = \frac{1}{P} \sum_{k=1}^{K} b_k^{(K)} \lambda_g^{(k)}$. Dodatno,

prethodnoj procjeni modela može se odrediti optimalan broj nezavisnih varijabli u modelu, a to se radi tako da u svakom koraku uklanja najmanje važnu varijablu, gdje je važnost definirana apsolutnom vrijednošću standardiziranog koeficijenta $\beta_g^* = \left(\frac{\sigma_g}{\sigma_y}\right)\beta_g$, gdje σ označava standardnu devijaciju. Metoda M-presječne provjere (CV) koristi se za određivanje dva parametra algoritma: broj komponenata K i broj nezavisnih varijabli P.

Proširenje na logističku napravi se zamjenom Y kao zavisne varijable s logit transformacijom $g(Y)$.

Sada kada je teorijski okvir postavljen, može se prijeći na empirijski dio diplomskog rada.

Poglavlje 3

Problem ocjene neodrživog razvoja turizma

Turizam u Hrvatskoj je vrlo relevantna gospodarska grana koja uz gospodarski rast donosi i određene rizike. Problem mjerenja i ocjene važnosti različitih rizika je ono čime se bavi ovaj dio rada. Neodrživi razvoj turizma uključuje sve negativne utjecaje na ekološki, socijalni, ekonomski, psihološki ili politički život. Velik broj istraživanja bavi se procjenom nosivog kapaciteta destinacije, tj. pronalaženju praga poput maksimalnog broja posjetitelja koji bi se mogao smatrati prihvatljivim u kontekstu održivosti turizma. Međutim, zbog vrlo različitih percepcija turizma i njegovih posljedica u različitim kulturama i kontekstima, ne postoji standardizirani metodološki pristup ili dobro definirane mjere za procjenu tog praga. Primjerice, gužva, kao jedna od najočitijih posljedica turizma, smatra se psihološkim konstruktom koji je snažno pod utjecajem osobnih karakteristika domaćeg stanovništva, a ne objektivnom mjerom gustoće posjetitelja, odnosno nešto što je nekome nepodnošljiva gužva, nekom drugom nije i ne postoji objektivna mjera koja to može razlikovati.

United Nations World Tourism Organization (2023) predlaže alternativno rješenje u kojem se neodrživi razvoj turizma razmatra iz perspektive rizika koje identificira lokalno stanovništvo; to je metoda kojom će se ocjenjivati neodrživi razvoj turizma u ovom radu. Identifikacija čimbenika rizika i njihova relativna važnost temeljit će se na anketi koju je ispunilo domaće stanovništvo čiji je cilj shvatiti najvažnije čimbenike u formiranju stavova vezanih za turizam.

Metodološki okvir prikazan u ovom radu primijenjen je za evaluaciju percepcija stanovnika o utjecajima turizma u drugom najvećem gradu Hrvatske, Splitu. Prema popisu stanovništva iz 2021. godine (Državni zavod za statistiku (2022)) u njemu živi približno 160.000 stanovnika. Povijesna jezgra Splita upisana je na Popis svjetske baštine 1979. godine, a uključuje ruševine Dioklecijanove palače izgrađene između 295. i 305. godine te niz srednjovjekovnih građevina. Grad privlači veliki broj turista s ostvarenih 2.620.705 noćenja u komercijalnim smještajnim objektima u 2022. godini. Intenzivni rast turizma tijekom posljednjeg desetljeća doveo je do promjena u lokalnoj zajednici zbog iseljavanja lokalnog stanovništva iz centra grada gdje su stambeni prostori pretvoreni u turističke objekte za kratkoročni najam, kao i do povećanja cijena nekretnina i promjena u karakteru mjesta (Matečić et al., 2022.). Turistička aktivnost je snažno koncentrirana u povijesnom gradskom središtu i podložna sezonalnosti.

Anketa među lokalnim stanovnicima provedena je u lipnju 2022. na uzorku od 385 stalnih stanovnika Splita. Kao metoda prikupljanja podataka korišteno je telefonsko anketiranje uz asistenciju računala (CATI). Podatke je

prikupila agencija za istraživanje tržišta IPSOS. Kao instrument istraživanja korišten je strukturirani upitnik, a uključivao je podatke o sociodemografskim karakteristikama, percepcijama utjecaja turizma i preferencijama za daljnji razvoj turizma. Uzorak je reprezentativan na razini grada prema spolu i dobnoj skupini stanovnika.

Ispitanici su zamoljeni da na skali 1 do 5 iskažu stav o utjecaju turizma na njihov svakodnevni život. Postavljeno im je pitanje: „Razmislite kako sve turizam utječe na Vaš svakodnevni život, lokalno gospodarstvo, okoliš, sigurnost u mjestu, cijene i slično. Uzimajući u obzir i dobre i loše strane turizma, smatrate li da je život u mjestu lošiji ili bolji zbog turizma?“. Odgovor na ovo pitanje će biti zavisna varijabla u ovom radu. Za potrebe ovog rada ocjene 1, 2 i 3 smatrat će se negativnom ocjenom, odnosno nulom, a ocjene 4 i 5 smatrat će se pozitivnom ocjenom, odnosno jedinicom.

U prikazu rezultata istraživanja prvo su dani rezultati eksploratorne analize i njezinih rezultata kako bi se dobio općeniti uvid u podatke koji se istražuju. Nakon toga prikazuje se univarijatna logistička regresija za sve nezavisne varijable i standardizacija koeficijenata te prvi zaključci o važnosti varijabli. Potom se pokazuju rezultati potpunog multivarijatnog modela sa svim nezavisnim varijablama i „stepwise“ modela. Za kraj provedena je i prikazana analiza dominance i CCR metoda.

3.1 Eksploratorna analiza podataka

U tablici 3.1 prikazane su varijable koje je anketa ispitivala i postotak negativnih odgovora za danu kategoriju, odnosno varijablu. Sve varijable su mjerene na Likertovoj skali s 5 stupnjeva.

Varijable	**Opis varijable**	**%**	**Rang**
Problemi vezani uz gužve:			
buka	Buka	33.83	16
promet	Gužva u prometu	70.03	7
gužve	Gužve na ulicama/javnim površinama	37.69	15
prijevoz	Gužve u javnom prijevozu	33.53	18
otpad	Neprimjereno odloženo smeće	71.22	6
neugodni mirisi	Neugodni mirisi (iz kontejnera, kanalizacije, ventilacije)	55.19	10
ponašanje	Neprimjereno ponašanje turista	47.48	12
parkiranje	Problemi s parkiranjem (nedostatak slobodnih mjesta)	84.87	3
Povećane cijene:			
nekretnine	Nekretnine	90.21	1
najam	Cijene najma	88.43	2
komunalije	Komunalije (struja, voda, plin)	43.92	14
namirnice	Hrana i piće u trgovinama	64.09	8
restorani	Cijene u restoranima/kafićima	83.68	5
Ostali negativni utjecaji turizma:			
izgled	Izgled grada (neprivlačan, neugodan)	25.52	20
apartmanizacija	Apartmanizacija	58.46	9
autentičnost	Gubitak autentičnosti/identiteta mjesta	49.55	11
iseljavanje	Iseljavanje lokalnog stanovništva	29.38	19
prostor	Neadekvatno korištenje javnog prostora	33.83	17

usluge	Smanjena dostupnost javnih usluga	46.88	13
stanovanje	Nedostatak pristupačnih stambenih mogućnosti	84.47	4

Tablica 3.1: Tablica negativnih percepcija

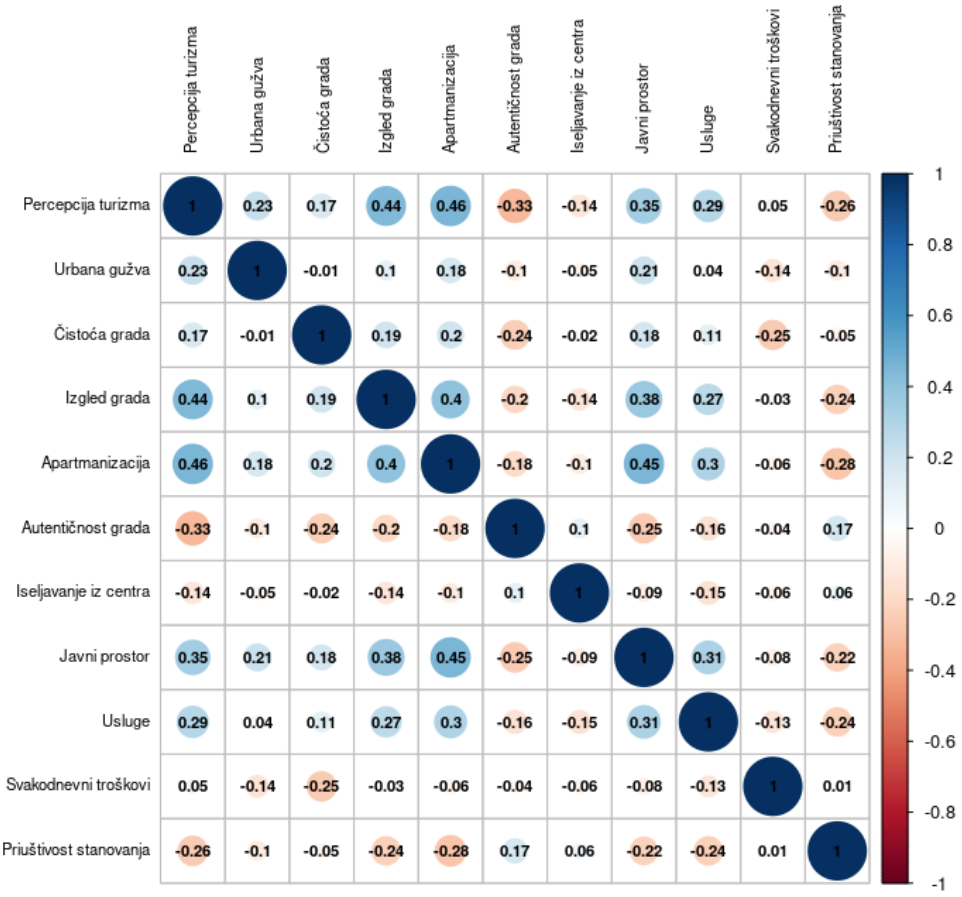
Rangiranje u tablici 3.1 provedeno je prema zastupljenosti negativne percepcije. Najnegativnije percepcije utjecaja turizma povezane su s povećanjem cijena - stanovnici većinski vjeruju da je razvoj turizma snažno utjecao na povećanje cijena nekretnina (90 %), dugoročnog najma smještaja (88 %) i restorana/kafića (84%). Nadalje, 85% smatra da su problemi s parkiranjem tijekom turističke sezone ozbiljan svakodnevni problem, dok 71% vidi odlaganje otpada kao ozbiljan problem. Većina stanovnika (58%) doživljava apartmanizaciju kao faktor koji negativno utječe na život u gradu, a svaki drugi stanovnik smatra da Split gubi svoj karakter, svoju autentičnost. Pitanje o iseljavanju lokalnog stanovništva iz povijesnog centra grada u predgrađa, koje za razliku od drugih pitanja odražava stvarno ponašanje, a ne percepcije, otkriva da je 29% stanovnika ili se preselilo iz centra grada ili su to učinili njihovi obitelji/prijatelji.

Prije provodenja multivarijatne logističke regresije, provedena je PCA (eng. Principal Component Analysis) među visoko koreliranim varijablama kako bi se umanjio utjecaj visoke korelacije među njima (visoka korelacija među nezavisnim varijablama također se naziva multikolinearnost). Potrebno ju je umanjiti jer otežava (u nekim slučajevima čak i onemogućuje) ocjenu važnosti pojedinih nezavisnih varijabli u modelu. Studije o stavovima lokalnog stanovništva obično uključuju brojne međusobno povezane varijable, od kojih se neke mogu smatrati integralnim dijelovima istog višedimenzionalnog

konstrukta. Provedbom PCA na probleme vezane za gužve dobivena su dva faktora. Faktor „urbana gužva“ određen je varijablama: buka, promet, gužve, javni prijevoz i parking. Drugi faktor „čistoća grada“ određen je varijablama: otpad, neugodni mirisi i ponašanje turista. PCA je provedena i na probleme vezane za povećanje cijena te su dobivena dva faktora. Faktor „svakodnevni troškovi“ određen je varijablama: komunalije, cijena namirnica te cijene u restoranima i kafićima. Drugi faktor „priuštivost stanovanja“ određen je varijablama: dostupnost nekretnina, cijena nekretnina i cijena najma. Više detalja o konstrukciji komponenata dostupno je u radu Sever, I. i Perišić, A. (2024). Konačno, sve nezavisne varijable koje se promatraju i na kojima se nastavlja analiza su: urbana gužva, čistoća grada, izgled grada, apartmanizacija, autentičnost, iseljavanje, javni prostor, usluge, svakodnevni troškovi i priuštivost stanovanja. Nadalje, varijable nisu rekodirane, odnosno visoka ocjena za nezavisnu varijablu ne mora označavati ono što bi bilo pozitivno za percepciju turizma.

Slijedi tablica s korelacijama varijabli.

3.1. Eksploratorna analiza podataka



Slika 3.1: Korelacije među nezavisnim varijablama

Valja primijetiti da nijedna od nezavisnih varijabli koje će se uključiti u model nije međusobno iznimno visoko pozitivno ili negativno korelirana što će olakšati analizu važnosti varijabli.

Podatci su podijeljeni na podatke za treniranje (80% podataka) i podatke za testiranje (20% podataka). Podjela podataka omogućuje ocjenu prediktivne moći modela pomoću pokazatelja osjetljivosti, točnosti, specifičnosti, preciznosti i AUC-a na testnim podatcima.

3.2 Testiranje značajnosti raznim modelima

3.2.1 Univarijatni modeli

U tablici 3.2 prikazani su rezultati univarijatnih modela. Prikazana je p-vrijednost varijable gdje visoka p-vrijednost ukazuje na nisku značajnost nezavisne varijable, prikazana je AUC ocjena i standardizirani regresijski koeficijent (djelomično standardizirani koeficijent).

Nezavisna varijabla	**p-vrijednost**	**AUC (treniranje)**	**Polustan. koef.**
Urbana gužva	0.002	0.589	0.38
Čistoća grada	0.008	0.598	0.35
Izgled grada	< 0.001	0.754	1.12
Apartmanizacija	< 0.001	0.759	1.15
Autentičnost grada	< 0.001	0.684	-0.67
Iseljavanje iz centra	0.004	0.571	-0.37
Javni prostor	< 0.001	0.700	0.85
Usluge	< 0.001	0.660	0.58
Svakodnevni troškovi	0.279	0.535	0.13
Priuštivost stanovanja	< 0.001	0.648	-0.57

Tablica 3.2: Univarijatni modeli s pripadajućim p-vrijednostima, AUC na podatcima za treniranje i standardiziranim koeficijentima

Iz tablice 3.2 vidi se da nezavisna varijabla „Svakodnevni troškovi“ za univarijatni model nije značajna. To se vidi i iz svakog od prikazanih pokazatelja: p-vrijednost je velika, također to je varijabla s minimalnim AUC-om i s najmanjim (po apsolutnoj vrijednosti) standardiziranim regresijskim koeficijentom.

Nez. var.	Osjet.	Toč.	Spec.	Prec.	AUC (testni)
Urbana gužva	0.588	0.642	0.697	0.667	0.743
Čistoća grada	0.647	0.567	0.485	0.564	0.595
Izgled grada	0.559	0.627	0.697	0.655	0.709
Apartmanizacija	0.588	0.701	0.818	0.769	0.762
Autentičnost grada	0.706	0.701	0.697	0.706	0.733
Iseljavanje iz centra	0.765	0.493	0.212	0.5	0.488
Javni prostor	0.794	0.701	0.606	0.675	0.682
Usluge	0.765	0.672	0.576	0.65	0.709
Svakodnevni troškovi	0.882	0.478	0.061	0.492	0.503
Priuštivost stanovanja	0.588	0.642	0.697	0.667	0.697

Tablica 3.3: Univarijatni modeli s pripadajućim pokazateljima osjetljivosti, točnosti, specifičnosti, preciznosti i AUC na testnim podatcima

U tablici 3.3 prikazane su ocjene osjetljivosti (udio točnih pozitivnih predviđanja), točnosti (udio točnih predviđanja), specifičnosti (udio točnih negativnih predviđanja), preciznosti (udio točnih pozitivnih predviđanja u svim pozitivnim predviđanjima) i AUC-a izračunatog s testnim podatcima (površina ispod ROC krivulje). Što su ove ocjene veće to je veća prediktivna moć modela. U tablici 3.3 se vidi da varijable kao što su „Iseljavanje iz centra“ i „Svakodnevni troškovi“ imaju točnost nižu od 50 posto što ih čini lošijima od slučajnog modela; to je još jedna indikacija da te varijable samostalno ne predviđaju dobro zavisnu varijablu. Samo varijable „Urbana gužva“, „Izgled grada“, „Apartmanizacija“, „Autentičnost grada“ i „Usluge“ imaju prihvatljivu razinu prilagođenosti po AUC-u. Nijedan model nije izvrsno ili izvanredno prilagođen što ukazuje na potrebu za izgradnjom multivarijatnih

modela.

3.2.2 Potpuni multivarijatni model

U tablici 3.4 prikazani su rezultati potpunog multivarijatnog modela. Potpuni multivarijatni model uključuje sve nezavisne varijable. Prikazane su procijenjene vrijednosti regresijskih koeficijenata, standardne pogreške regresijskih koeficijenata te p-vrijednosti.

Nezavisna varijabla	**Procjena koef.**	**St. pogreška**	**p-vr.**
Urbana gužva	0.366	0.229	0.110
Čistoća grada	0.244	0.231	0.290
Izgled grada	0.747	0.163	< 0.001
Apartmanizacija	0.741	0.174	< 0.001
Autentičnost grada	-0.499	0.155	0.001
Iseljavanje iz centra	-0.582	0.429	0.174
Javni prostor	0.171	0.188	0.362
Usluge	0.316	0.157	0.045
Svakodnevni troškovi	0.430	0.178	0.015
Priuštivost stanovanja	-0.357	0.176	0.042

Tablica 3.4: Procijenjeni koeficijenti, standardne pogreške i p-vrijednosti za multivarijatni model

Po potpunom multivarijatnom modelu nisu značajne varijable „Urbana gužva“, „Čistoća grada“, „Iseljavanje iz centra“ te „Javni prostor“, a na granici značajnosti su „Priuštivost stanovanja“ i „Usluge“. U usporedbi s univarijatnim modelima rezultati su drugačiji što pokazuje važnost razmatranja različitih pristupa kako bi se došlo do valjanih zaključaka. Točnost modela je 0.731, a AUC je 0.824 što pokazuje izvrsnu prilagođenost modela.

Također, po p-vrijednosti 0.1188 Hosmer-Lemeshow testa nema dovoljno dokaza koji upućuju na lošu prilagođenost modela.

Dalje u radu prikazuje se izgradnja modela „stepwise“ procedurom nakon čega su analizom dominance i CCR analizom dani zaključci o važnosti varijabli. To su sofisticiraniji modeli i pomoći će u donošenju snažnijih zaključaka vezanih za važnost varijabli.

3.2.3 „Stepwise“ model

Prije rezultata koje je iznjedrila „stepwise“ procedura daje se kratki prikaz same procedure. Za opis modela uvest će se novi indikator: AIC (eng. Akaike information criterion). AIC je ocjena modela definirana na sljedeći način:

$$AIC = 2k - 2l(\hat{\boldsymbol{\beta}}),$$

gdje je k broj nezavisnih varijabli u modelu, a $2l(\hat{\boldsymbol{\beta}})$ maksimalna vrijednost funkcije log-vjerodostojnosti. Ocjena istovremeno nastoji mjeriti prediktivnu moć modela i kažnjava velik broj nezavisnih varijabli u modelu. Na temelju ove ocjene moguće je uspoređivati ugniježđene modele gdje niže vrijednosti AIC-a ukazuju na bolji model.

„Stepwise“ procedura je procedura iterativne konstrukcije regresijskog modela koja uključuje odabir nezavisnih varijabli koje će se koristiti u konačnom modelu. Ocjena modela u „stepwise“ proceduri može se raditi na različite načine, ali u ovom radu izabrana je ocjena AIC, cilj je nju minimizirati. Tri su tipa iterativne metode: unaprijedna selekcija započinje s modelom koji ne sadrži nijednu varijablu, dodaje varijable te testira model s dodanom varijablom, a zatim zadržava one varijable koje minimiziraju AIC. Unazadna

eliminacija započinje s nizom svih nezavisnih varijabli, brišući jednu po jednu, zatim testirajući model pomoću AIC-a da vidi je li se smanjio ili povećao. Dvosmjerna eliminacija je kombinacija prve dvije metode. Rezultati u tablici 3.5 dobiveni su „stepwise“ modelom s dvosmjernom eliminacijom.

Nezavisna varijabla	**Procjena koef.**	**St. pogreška**	**p-vr.**
Urbana gužva	0.427	0.220	0.052
Izgled grada	0.795	0.160	< 0.001
Apartmanizacija	0.797	0.169	< 0.001
Autentičnost grada	-0.555	0.149	< 0.001
Usluge	0.342	0.151	0.024
Svakodnevni troškovi	0.378	0.169	0.025
Priuštivost stanovanja	-0.348	0.176	0.049

Tablica 3.5: „Stepwise“ model s procijenjenim koeficijentima, standardnim pogreškama i p-vrijednostima

Rezultati u „stepwise“ modelu u skladu su s rezultatima potpunog multivarijatnog modela. Izbačene su varijable „Čistoća grada“, „Iseljavanje iz centra“te „Javni prostor“, što su baš varijable kojima je potpuni multivarijatni model dao visoke p-vrijednosti. Točnost modela je 0.761, a AUC je 0.842 što su bolji pokazatelji od pokazatelja potpunog multivarijatnog modela. To je posebno pozitivno s obzirom na to da „stepwise“ model ima manje varijabli. AIC potpunog modela je 258.58, a AIC „stepwise“ modela je 256.37.

Hosmer-Lemeshow test za „stepwise“ model ima iznimno visoku (0.8765) p-vrijednost što ukazuje na to da nema dovoljno dokaza koji upućuju na lošu

prilagođenost modela.

3.2.4 Analiza dominance

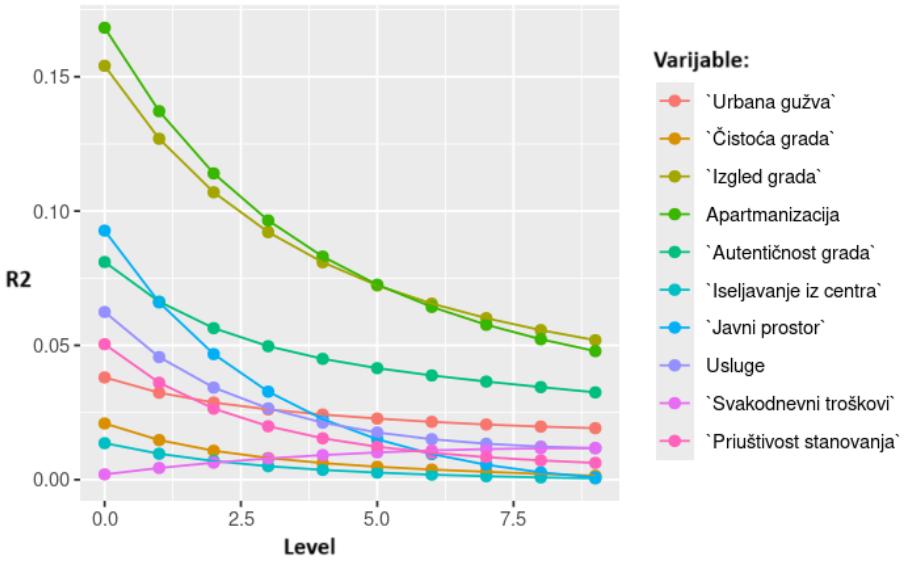
Sada će se relativna važnost varijabli procijeniti korištenjem analize dominance. Sve nezavisne varijable koje su promatrane za model potpune multivarijatne regresije bit će uključene u analizu jer su sve bile statistički značajne u barem jednom modelu. Promatrat će se važnost po svim dominacijama, odnosno po potpunoj dominaciji (3.6), uvjetnoj dominaciji (3.2) i općenitoj dominaciji (3.3).

	‘Urbana gužva‘	‘Čistoća grada‘	‘Izgled grada‘	Apartmanizacija	‘Autentičnost grada‘	‘Iseljavanje iz centra‘	‘Javni prostor‘	Usluge	‘Svakodnevni troškovi‘	‘Priuštivost stanovanja‘
Urbana gužva	-	D	ND	ND	ND	D	-	-	D	-
Čistoća grada	ND	-	ND	ND	ND	-	-	ND	-	-
Izgled grada	D	D	-	-	D	D	D	D	D	D
Apartmanizacija	D	D	-	-	D	D	D	D	D	D
Autentičnost grada	D	D	ND	ND	-	D	-	-	D	D
Iseljavanje iz centra	ND	-	ND	ND	ND	-	ND	ND	-	ND
Javni prostor	-	-	ND	ND	-	D	-	-	-	-
Usluge	-	D	ND	ND	-	D	-	-	-	-

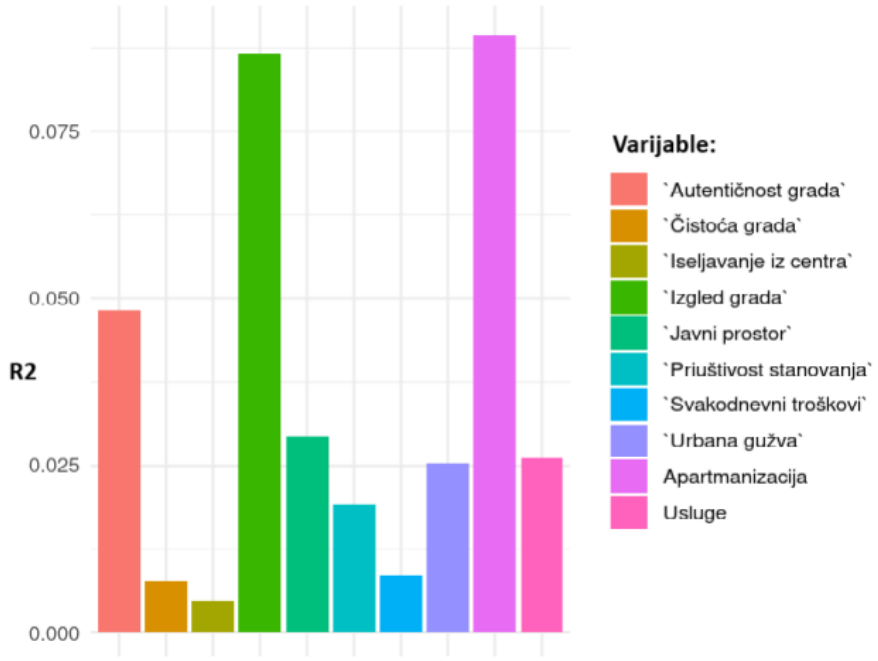
Svakodnevni troškovi	ND	-	ND	ND	ND	-	-	-	-	-
Priuštivost stanovanja	-	-	ND	ND	ND	D	-	-	-	-

Tablica 3.6: Matrica potpune dominacije: D označava potpunu dominaciju varijable u retku nad varijablom u stupcu, ND potpunu dominaciju varijable u stupcu nad varijablom u retku, a - označava da se potpuna dominacija ne može uspostaviti

U 3.6 vidi se da varijable „Apartmanizacija“ i „Izgled grada“ potpuno dominiraju sve ostale varijable (osim međusobno jedna drugu), a isto se pokazalo u uvjetnoj dominaciji prikazanoj na slici 3.2. Po općenitoj dominaciji u grafu 3.3 nešto se važnijom pokazala varijabla „Apartmanizacija“. Varijabla „Autentičnost“ treća je po broju varijabli koje potpuno dominira, uvjetno dominira sve osim „Javni prostor“ i dvije najdominantnije te je treća u općenitoj dominaciji.



Slika 3.2: Uvjetna dominacija



Slika 3.3: Općenita dominacija

Nešto što se vidjelo i iz prethodnih analiza je da varijable „Iseljavanje iz centra“i „Čistoća grada“ nisu važne: ne dominiraju niti jednu varijablu potpuno, uvjetno „Iseljavanje iz centra“ ne dominira nijednu, a „Čistoća grada“ jedino uvjetno dominira „Iseljavanje iz centra“. U općenitoj dominaciji dvije su najniže po poretku.

Za razliku od prošlih modela („stepwise“ i multivarijatni) varijabla „Javni prostor“ se pokazala važnom u analizi dominance.

3.2.5 CCR

Prethodno izgradnji CCR modela krosvalidacijom je ispitan optimalan broj komponenata te je po kriteriju AUC zaključeno da su dvije komponente optimalan broj komponenti koje trebaju biti uključene u model. Nakon računanja dviju komponenti i provedbe regresije na podacima za treniranje

s tim komponentama dobiveni su sljedeći rezultati.

Nezavisna varijabla	**Procjena koef.**	**St. pogreška**	**p-vr.**
S1	7.032	0.881	< 0.001
S2	8.26	2.764	0.003

Tablica 3.7: Procijenjeni koeficijenti, standardne pogreške i p-vrijednosti za CCR model s dvije komponente

Model iz tablice 3.7 ima AUC 0.821 na testnim podatcima i točnost 0.7313 što je lošije od „stepwise“ modela i vrlo slično kao multivarijatni model.

U sljedećoj tablici bit će prikazane p-vrijednosti nezavisnih varijabli u univarijatnim modelima i u modelima s jednom komponentom S_1. Ti podatci će pokazati koje varijable su primarne, a koje supresorske.

Nezavisna varijabla	**p-vr. univar. modela**	**p-vr. modela s komp.** S_1
Urbana gužva	0.002	0.864
Čistoća grada	0.008	0.593
Izgled grada	< 0.001	0.383
Apartmanizacija	< 0.001	0.809
Autentičnost grada	< 0.001	0.338
Iseljavanje iz centra	0.004	0.903

Javni prostor	< 0.001	0.053
Usluge	< 0.001	0.509
Svakodnevni troškovi	0.279	0.025
Priuštivost stanovanja	< 0.001	0.835

Tablica 3.8: Tablica p-vrijednosti u univarijatnom modelu i u modelu s prvom CCR komponentnom S_1

Iz tablice 3.8 vidi se da su sve varijable primarne osim varijable „Svakodnevni troškovi“ koja je supresorska varijabla, jedino ona ima visoku p-vrijednost u univarijatnom modelu i nisku u modelu s prvom komponentom. Također je zanimljivo da varijabla „Javni prostor“ se ponaša i kao primarna i kao supresorska varijabla.

U sljedećim tablicama prikazani su standardizirani koeficijenti CCR, univarijatnog i potpunog multivarijatnog modela poredani po apsolutnoj vrijednosti. Takav pregled omogućava jednostavnu usporedbu važnosti pojedinih varijabli u različitim modelima.

Nezavisna varijabla	**Stand. koef. CCR modela**
Izgled grada	0.940
Apartmanizacija	0.853
Autentičnost grada	-0.607
Svakodnevni troškovi	0.397
Usluge	0.313
Priuštivost stanovanja	-0.368
Iseljavanje iz centra	-0.279
Urbana gužva	0.245

Javni prostor	0.226
Čistoća grada	0.166

Tablica 3.9: Standardizirani koeficijenti za CCR model sortirani po apsolutnoj vrijednosti

Nezavisna varijabla	**Stand. koef. univar. modela**
Apartmanizacija	1.149
Izgled grada	1.121
Autentičnost grada	-0.667
Javni prostor	0.850
Usluge	0.585
Priuštivost stanovanja	-0.568
Urbana gužva	0.383
Čistoća grada	0.345
Iseljavanje iz centra	-0.372
Svakodnevni troškovi	0.127

Tablica 3.10: Standardizirani koeficijenti za univarijatni model sortirani po apsolutnoj vrijednosti

Nezavisna varijabla	**Stand. koef. multivar. modela**
Izgled grada	0.912
Apartmanizacija	0.899
Autentičnost grada	-0.582
Svakodnevni troškovi	0.432
Usluge	0.360

Priuštivost stanovanja	-0.355
Urbana gužva	0.286
Iseljavanje iz centra	-0.233
Čistoća grada	0.192
Javni prostor	0.186

Tablica 3.11: Standardizirani koeficijenti za multivarijatni model sortirani po apsolutnoj vrijednosti

U tablicama 3.9, 3.10 i 3.11 vidi se da su varijable „Izgled grada“, „Apartmanizacija“i „Autentičnost grada“uvijek među tri najviše varijable po apsolutnoj vrijednosti standardiziranog koeficijenta, odnosno imaju najveću značajnost. S druge strane varijable „Iseljavanje iz centra“, „Čistoća grada“i „Urbana gužva“ uvijek su među četiri varijable s najnižom vrijednošću, odnosno njihova značajnost je najmanja u prikazanim modelima.

Zaključak

Razvoj turizma donosi brojne ekonomske prednosti, ali može imati značajne negativne utjecaje na lokalne zajednice, okoliš i društvo. U radu se istraživala percepcija lokalnog stanovništva o utjecaju turizma na grad. Korištenjem anketnih podataka i statističkih metoda, utvrdili su se najvažniji i najmanje važni prediktori percepcije turizma.

Primijenjene su statističke metode poput izgradnje univarijatnih i multivarijatnih modela logističke regresije, provođenja „stepwise“ procedure, analize dominance i metode regresije koreliranih komponenti (CCR).

Većina modela testiranjem pokazala se dobro prilagođenima uz relativno visoku prediktivnu moć, naročito potpuni multivarijatni model, „stepwise“ model i CCR model. Iako postoje razlike u rezultatima različitih modela, gotovo svi modeli dosljedno ukazuju na to da percepciju turizma najviše oblikuju percepcije utjecaja turizma na izgled grada, apartmanizaciju i autentičnost grada. S druge strane, faktori poput iseljavanja iz centra i čistoće grada pokazali su se kao najmanje važni. Ove rezultate potvrdila je metoda analize dominance i metoda regresije koreliranih komponenti.

Literatura

Azen, R. i Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34(3):319–347.

Državni zavod za statistiku (2022). *Popis stanovništva 2021.*

Gotovac Đogaš V. (2023.). *Statistika - skripta.* Prirodoslovno-matematički fakultet.

Han, J. i Kamber, M. i Pei, J. (2012). *Data Mining Concepts and Techniques*, stranice 364–377. Elsevier, 3rd edition.

Hosmer, D. W. i Lemeshow, S. i Sturdivant, R. (2013). *Applied Logistic Regression*, stranice 1–17, 31–40, 143–147. John Wiley & Sons, 3rd edition.

Magidson, J. (2013). *New Perspectives in Partial Least Squares and Related Methods*, poglavlje Correlated Component Regression: Re-thinking Regre-ssion in the Presence of Near Collinearity, stranice 65–78. Springer New York.

Menard S. (2004). Six approaches to calculating standardized logistic regression coefficients. *American Statistical Association*, 58(3):218–223.

Literatura

Sever, I. i Perišić, A. (2024). From thresholds to risk factors: Prioritization of socio-economic risks to sustainable tourism development. 63rd ERSA congress.

United Nations World Tourism Organization (2023). *Statistical Framework for Measuring the Sustainability of Tourism.*