

# Najemnine v Ljubljani

Domen Mohorčič

22. avgust 2020

## 1 Uvod

Ko se povprečen Slovenec odseli od staršev v svoje stanovanje, je star 28,2 leti (stat.si). Pred izselitvijo pa si mora stanovanje poiskati. Po navadi ljudje pri izbiri stanovanja gledajo na to, ali jim je stanovanje všeč in ali se jim zdi cena ustrezna stanovanju. Od česa pa sploh je odvisna cena stanovanja? Friškovec (2010)[1] ugotavlja, da je oglaševalska cena stanovanja pozitivno odvisna od površine, števila kopalnic in ali gre za mansardno stanovanje, negativno pa predvsem od višine nadstropja. Repič (2014)[2] pa je v magistrski nalogi pokazala, da je prodajna cena stanovanja pozitivno odvisna od prisotnosti dvigala, parkirnega mesta, opremljenosti stanovanja, bližine središča Ljubljane in števila sob v stanovanju, negativno pa na ceno vplivajo starost stanovanja, površina in trajanje, ko je nepremičnina na voljo za prodajo.

V Sloveniji v najemniških stanovanjih živi le 2,4 gospodinjstev (cekin.si). Kljub temu je trg najema nepremičnin kar velik, še posebej v Ljubljani (na nepremicnine.net je od 1500 oglasov za najem, od tega kar 1000 v Ljubljani), kjer pa so najbolj zaželeni študentje ali posamezniki. Nikjer pa nisem zasledil raziskave, ki bi ugotavljala, kaj vpliva na ceno najema.

Namen seminarske naloge je ugotoviti, kateri dejavniki najbolj vplivajo na ceno najemnin v Ljubljanskima predeloma Vič in Rudnik.

## 2 Podatki

Podatke sem pridobil iz slovenske spletne strani nepremicnine.net dne 8.8.2020. Iskal sem stanovanja v Ljubljani v predelih Vič in Rudnik. Pri pregledovanju oglasov sem se osredotočil na naslednje podatke: nadstropje, v katerem se stanovanje nahaja, število vseh nadstropij v zgradbi, leto gradnje stavbe, leto prenove stanovanja, število sob v stanovanju, ali ima stanovanje shrambo/klet, ali je stanovanje opremljeno, število pripadajočih parkirišč, velikost

bivalne površine, zunanje površine (balkon, vrt, ...), mesečni stroški bivanja in cena najema. Ker pa sem hotel ugotoviti, ali na ceno najema vpliva tudi lokacija stanovanja, sem poiskal še oddaljenost do središča Ljubljane (v mojem primeru Prešernov trg). Na prej omenjeni spletni strani pa v večini primerov ni napisanega točnega naslova, zato sem iskal samo približne lokacije (ulica ali naselje).

Pri določanju razdalje sem si pomagal z orodjem `distance.to`. Za analizo podatkov sem uporabil program RStudio.

## 2.1 Opis spremenljivk

Zbrane podatke sem označil z naslednjimi spremenljivkami:

spremenljivka	opis
nadstropje	V katerem nadstropju se stanovanje nahaja
vsaNadstropja	Število vseh nadstropij v stavbi
letoGradnje	Leto, v katerem je bilo stanovanje zgrajeno
letoPrenove	Leto, v katerem je bilo stanovanje prenovljeno
stSob	Število sob v stanovanju
stParkirisc	Število parkirnih mest, ki pripadajo stanovanju
parkirisce	Ali stanovanju pripada lastno parkirišče
opremljenost	Kako je stanovanje opremljeno (polno, delno ali nič)
shramba	Ali stanovanju pripada zunanja soba za shranjevanje
zunanjePovrsine	Vsota zunanjih površin stanovanja
povrsina	Velikost bivalne površine v stanovanju
oddaljenost	Oddaljenost stanovanja od Prešernovega trga
cena	Cena mesečne najemnine stanovanja
stroski	Cena mesečnih stroškov bivanja
skCena	Seštevek cene in mesečnih stroškov bivanja

Za model napovedi mesečne najemnine sem izbral spremenljivke *skCena*, *parkirisce*, *povrsina*, *letoGradnje* in *oddaljenost*. *skCena* je odvisna spremenljivka, ostale štiri pa so neodvisne.

Spremenljivko *letoPrenove* sem odstranil iz modela, ker za večino stanovanj podatka nisem našel. *nadstropje* in *vsaNadstropja* sem izvzel, ker pri pregledu oglasov nisem dobil občutka, da bi ti dve spremenljivki pomembno vplivali na ceno najemnine. Spremenljivko *opremljenost* sem odstranil, ker je bilo 89.3% stanovanj opremljenih in tako ni bilo dovolj raznolikosti. Pri pregledu korelacijske matrike sem ugotovil, da sta spremenljivki *povrsina* in *stSob* povezani s korelacijskim koeficientom 0.811 zato sem

obdržal spremenljivko *povrsina*. Namesto *stParkirisc* sem uporabil spremenljivko *parkirisce*, saj je tako bolj predstavljeno, ali ga stanovanje ima. Spremenljivko *zunanjePovrsine* pa sem odstranil iz modela, ker je imela večina stanovanj samo balkon, redko pa so se pojavili velikimi vrtovi.

## 2.2 Analiza podatkov

Izmed petih spremenljivk so *letoGradnje*, *povrsina*, *oddaljenost* in *skCena* zvezne, *parkirisce* pa je diskretna. Zbral sem podatke o 103 različnih stanovanjih ( $N = 103$ ).

Za vsako zvezno spremenljivko sem izračunal povprečje, standardni odklon, mediano absolutnih odstopanj od mediane, asimetričnost in sploščenost. Povprečje se izračuna po formuli

$$\mu = \frac{\sum_{i=1}^N x_i}{N} \quad (1)$$

Standardni odklon se izračuna po formuli

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (2)$$

Mediano absolutnih odstopanj od mediane se izračuna tako, da se izračunajo vse absolutne razlike med podatki in mediano in se poišče mediano razlik.

Asimetričnost se izračuna po naslednji formuli

$$skew = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N * \sigma^3} \quad (3)$$

Sploščenost pa se izračuna po formuli

$$kurtosis = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N * \sigma^4} - 3 \quad (4)$$

V spodnji tabeli je so predstavljene prej omenjene lastnosti zveznih spremenljivk: minimum (min), maksimum (max), povprečje (avg), mediana (median), standardni odklon (sd), mediana absolutnih odstopanj od mediane (MAD), test asimetričnosti (skew) in test sploščenosti (kurt):

	min	max	avg	sd	median	MAD	skew	kurt
letoGradnje	1895	2020	1986,16	28,50	1995	25,20	-1,22	1,27
povrsina	10	150	67,75	38,40	65	41,22	0,42	-0,67
oddaljenost	0,97	5,10	2,53	1	2,18	0,67	0,77	-0,26
skCena	160	3700	940,13	700,20	800	333,58	2,17	5,28

Ker pa se vrednosti testa asimetričnosti in sploščenosti zelo razlikujejo od 0, to kaže na nenormalno porazdelitev, in tako nam vrednosti o povprečju ali standardnem odklonu povesta bolj malo. Naredil sem še test normalnosti z ukazom `shapiro.test()` (Shapiro-Wilk) in test simetričnosti z `symmetry.test()` (MGG):

	Shapiro-Wilk	MGG
letoGradnje	2e-07	1,32e-05
povrsina	0,002	0,344
oddaljenost	8e-06	3,82e-07
skCena	1e-11	4,22e-04

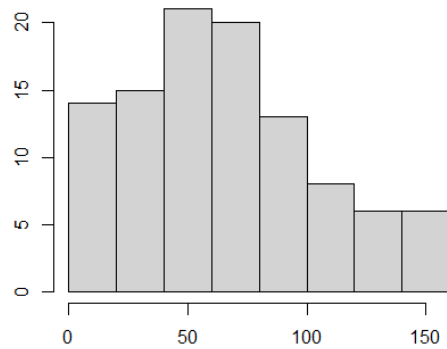
Shapiro-Wilk-ov test nam pove, s kakšno verjetnostjo je naš vzorec normalno porazdeljen, če je normalno porazdeljena populacija. V našem primeru ima najvišjo verjetnost spremenljivka *povrsina* z 0,2%, kar pa je premalo, da bi lahko sklepali o normalni porazdelitvi. Minimalni prag za sprejem hipoteze o normalnosti je namreč 5%.

MGG (Miao, Gel, and Gastwirth) test simetričnosti pa nam testira hipotezo, da je naš vzorec simetričen proti alternativni hipotezi, da vzorec ni simetričen. Vse spremenljivke razen *povrsina* imajo p vrednost testa zelo majhno, zato so nesimetrične. Spremenljivka *povrsina* pa je simetrično porazdeljena.

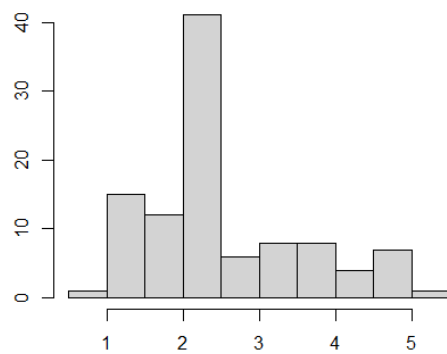
Tudi ko pogledamo histograme, vidimo, da so vse spremenljivke nenormalno porazdeljene in nesimetrične. Najbližje normalni porazdelitvi je *povrsina*, vendar je bolj nagnjena v desno.



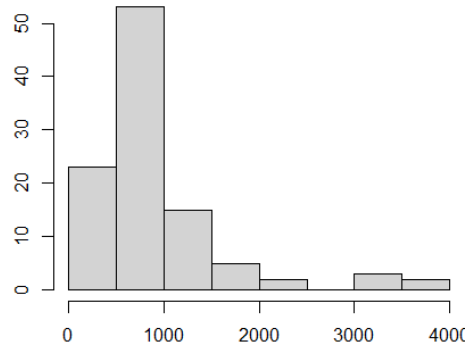
Slika 1: Leto gradnje



Slika 2: Povrsina



Slika 3: Oddaljenost



Slika 4: Skupna cena

Spremenljivka *parkirisce* pa je diskretna, zato jo lahko predstavimo z vzorčnim deležem:

	da	ne
parkirisce	0,544	0,456

## 3 Večkratna regresija

### 3.1 Koeficienti korelacije

Pri modelu večkratne regresije je najprej potrebno preveriti, kako so neodvisne spremenljivke povezane med sabo. Če so povezane preveč, lahko z neko spremenljivko opišemo drugo, in tako iz druge ne izvemo nič novega ali pa zelo malo o odvisni spremenljivki. Korelacijski koeficient med dvema spremenljivkama lahko izračunamo po naslednji formuli:

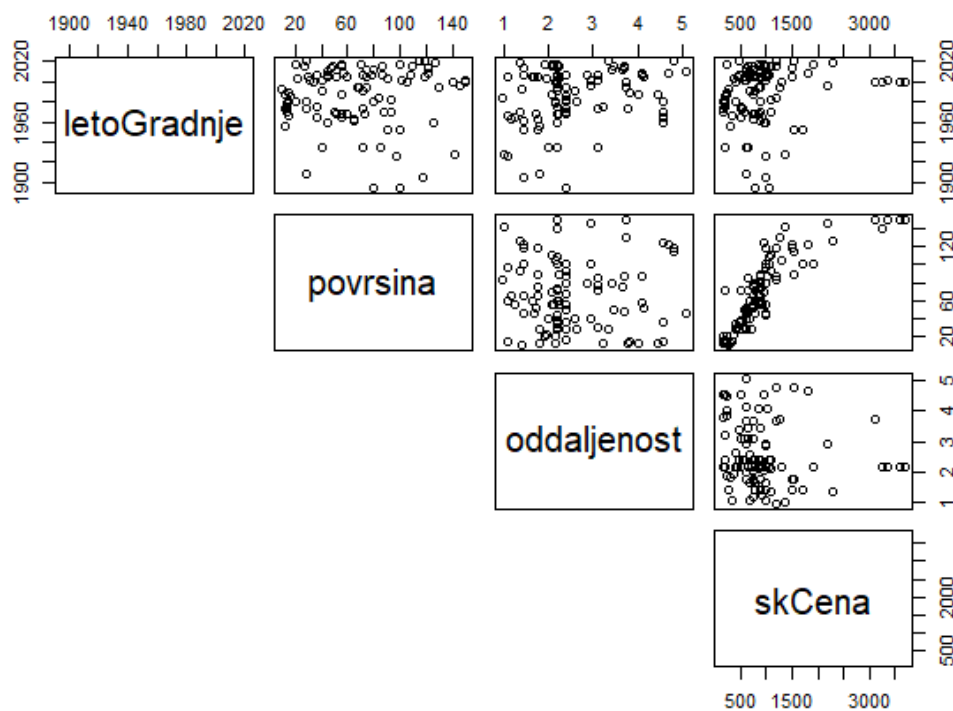
$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - E(X))(Y - E(Y)))}{\sigma_X \sigma_Y} \quad (5)$$

Koeficient korelacije ima vrednost med  $-1$  in  $1$ . Če ima vrednost blizu  $1$ , sta spremenljivki pozitivno povezani, če pa ima vrednost blizu  $-1$ , sta negativno povezani. Če vrednost znaša okoli  $0$ , spremenljivki nista povezani.

V spodnji tabeli so predstavljeni korelacijski koeficienti za zvezne spremenljivke:

	letoGradnje	povrsina	oddaljenost	skCena
letoGradnje	1.000	0.035	0.252	0.168
povrsina	0.035	1.000	-0.058	0.829
oddaljenost	0.252	-0.058	1.000	-0.100
skCena	0.168	0.829	-0.100	1.000

Neodvisne spremenljivke *letoGradnje*, *povrsina* in *oddaljenost* imajo medsebojne koeficiente skoraj 0 z izjemo *letoGradnje* in *oddaljenost*, ki imata pozitiven koeficient 0.252. Spremenljivki sta pozitivno povezani, kar pomeni, da se novejša stanovanja nahajajo malo bolj izven središča Ljubljane.



Slika 5: Grafični prikaz korelacijske matrike

### 3.2 Večkratna regresija

Linearna regresija je analiza, pri kateri ugotavljamo funkcijsko zvezo med dvema spremenljivkama, pri večkratni regresiji pa funkcijsko zvezo med več spremenljivkami, kjer je ena odvisna, ostale pa neodvisne. Cilj analize je najti linearno funkcijo, ki najbolje opiše obnašanje odvisne spremenljivke v odvisnosti od ostalih spremenljivk.

Neodvisne spremenljivke so *letoGradnje*, *povrsina*, *oddaljenost*, *parkirisce*.

Odvisna spremenljivka je *skCena*. Funkcija bo izgledala tako:

$$skCena = a + b * letoGradnje + c * površina + d * oddaljenost + e * parkirisce + \epsilon \quad (6)$$

$b, c, d$  in  $e$  so koeficienti posameznih neodvisnih spremenljivk.  $a$  predstavlja začetno vrednost, sam po sebi pa nima smisla (stanovanje z 0 v pri vseh neodvisnih spremenljivkah bi imelo  $a$  mesečne najemnine, tako stanovanje pa ne obstaja).  $\epsilon$  predstavlja naključno napako.

Večkratno regresijo sem določil z ukazom

`lm(skCena~letoGradnje+povrsina+oddaljenost+parkirisce, data=data)`.

Dobil sem naslednje podatke:

**Call:**

`lm(formula = skCena ~ letoGradnje + površina + oddaljenost + parkirisce, data = data)`

Residuals:

Min	1Q	Median	3Q	Max
-765.48	-203.91	-9.88	159.84	1283.37

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-8806.067	2660.228	-3.310	0.00131	**
letoGradnje	4.499	1.353	3.325	0.00124	**
povrsina	15.714	1.020	15.410	< 2e-16	***
oddaljenost	-60.497	38.036	-1.591	0.11494	
parkirisce	-184.771	79.226	-2.332	0.02174	*

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

Residual standard error: 371.6 on 98 degrees of freedom

Multiple R-squared: 0.7293, Adjusted R-squared: 0.7183

F-statistic: 66.01 on 4 and 98 DF, p-value: < 2.2e-16

## 4 Komentar

## 5 Literatura

[0] <https://www.distance.to/>

[1] Analiza dejavnikov oglaševanih cen rabljenih stanovanj v Ljubljani in njeni okolici, Sonja Friškovec, Aleksander Janež, Univerza na Primorskem, jesen 2010



- [2] Dejavniki oblikovanja prodajnih cen stanovanj, Mojca Repič, Ljubljana, oktober 2014
- [3] enačba 11