

ACADEMIA DE STUDII ECONOMICE

Facultatea: Cibernetică, Statistică și Informatică Economică

Specializarea: Informatică economică

Lucrare de cercetare

Analiza factorilor care influențează prețul caselor din București

Cadrul didactic coordonator:

Manta Eduard Mihai

Studenti:

Fătu Mihaela-Adelina

Enescu Roxana-Andreea-Daniela

Dună Vlad-Dragoș-Ionuț

București

2022 – 2023

Cuprins

Introducere.....	2
Aplicația 1. Modele de regresie.....	3
1.1 Literature review	3
1.2 Metodologia cercetării	7
2.1 Date utilizate	8
2.2 Rezultate empirice	11
Limitări	33
Concluzii.....	33
Bibliografie.....	34
Anexa.....	35
Tabele	35
Figuri	35

Introducere

Piața imobiliară reprezintă mecanismul economicosocial principal prin care se orientează activitatea operatorilor și se prezintă ca un sistem complex, în care interacționează și se condiționează reciproc cererea cu oferta. În acest sens, tranzacțiile imobiliare sunt acele transferuri permanente sau temporare ale unui drept de la o parte la cealaltă în schimbul unei recompense care, de obicei, este o sumă de bani.

Evaluarea proprietăților imobiliare este procesul de elaborare a unei opinii asupra valorii pentru proprietatea imobiliară. Tranzacțiile imobiliare, menționate și definite mai sus, necesită adesea evaluări datorită apariției lor rare și fiecare proprietate este unică spre deosebire de acțiunile corporative, ce sunt tranzacționate zilnic și sunt identice. Astfel, există mai multe tipuri de valori (de piață, echitabilă, de investiție, sinergie, de lichidare), abordări ale evaluării care contribuie la evoluția pieței imobiliare.

În România, piața imobiliară nu contribuie foarte mult la evoluția economiei precum în alte țări la fel sau mai dezvoltate. Astfel, din punct de vedere economic, constatăm că această piață este “decuplata” de la economie în general. Procentul mai mic în PIB constituit de sectorul construcțiilor (ca infrastructură, nu doar cel imobiliar) în România este explicația care ne este oferită de specialiști. De la prima criză imobiliară din România, lucrurile au schimbat radical comportamentul participanților din piața imobiliară. Atunci, cumpărătorii erau preocupați de prețul și structura imobilelor, aveau o încredere exagerată în promisiunile dezvoltatorilor imobiliari și nu luau prea des în seamă potențialele riscuri ale proiectelor.

Echipa noastră a ales această temă întrucât este o temă de actualitate, care se regăsește în mod constant în viața de zi cu zi, fiind o necesitate a oamenilor care trebuie să fie atenți în procesul de vânzare-cumpărare, având în vedere numeroși factori precum locația și starea în care se află imobilul, prețul acestuia cu fluctuațiile sale în continuă schimbare, cât și alți factori socio-economici din spațiul pieței imobiliare care prezintă interes. În situația dată, a fost analizată retrospectiva rezidențială pentru anul 2019, care potrivit siteului Imobiliare.ro, ritmul de creștere a prețurilor de pe piața rezidențială autohtonă a accelerat față de anul precedent. La nivelul celui de-al patrulea trimestru din 2019, locuințele (apartamentele și casele) sunt mai scumpe cu 7% decât în 2018. Comparativ cu anul trecut, diferența de preț la 12 luni se situa la 5,3%.

Aplicația 1. Modele de regresie

1.1 Literature review

Terenul este suportul esențial în cadrul vieții și a existenței noastre. Juriștii, geografi și economiștii își îndreaptă atenția spre acest lucru, întrucât studiind utilizările terenurilor, sunt influențate națiuni și societăți. Într-o activitate economică, proprietățile imobiliare devin un contribuitor important în circuitul economic prin crearea unei valori importante, care în urma unor evaluări amănunțite, în funcție de cerințele clienților, au devenit o investiție foarte importantă a oamenilor în societate. (CREȚU, 2019, pp. 93-101)

Oferta și cererea constituie procesul de interacțiune pentru a stabili valoarea de piață a imobilelor. Procesul de dezvoltare socio-economică reflectă interdependența cererii prin nevoia de teren și de furnizarea acestora prin intermediul pieței limitate a terenurilor. (LEȘAN, 2020, pg. 65-69) Astfel, există mulți factori ce influențează cererea și oferta, unde valoare de piață devine afectată și duce la împărțirea pieței funciare în segmentele ei diferite. Astfel, acești factori sunt împărțiți în patru mari categorii:

- Categoria factorilor economici
- Categoria factorilor sociali și economici
- Categoria factorilor de reglementare a statului a pieței funciare
- Categoria stării fizice a terenului, precum și mediul înconjurător

Prima categorie prezintă o serie de factori care afectează nivelul cererii. Printre acestea, se numără nivelul ocupării populației, nivelul prețurilor, cheltuielile pentru căutarea și achiziționarea unui imobil bun, înregistrarea tranzacției și disponibilitatea finanțării. Factorii ce influențează nivelul ofertei este reprezentat de numărul bunurilor imobile oferite pe piață, cheltuielile de înființare a unității de pământ și cheltuielile de construcție.

A doua categorie de factori surprinde numărul populației, densitatea acesteia, compoziția etnică, vârsta și nivelul de studii, precum migrația populației.

A treia categorie include factori precum reglementarea legată a pieței bunurilor imobile la nivel de stat și la nivelul administrației locale, precum și politica statului privind impozitarea, inflația, împrumuturile și investițiile.

A patra categorie cuprinde factori precum mărimea terenului, forma acestuia, îmbunătățirile, disponibilitatea comunicațiilor, gradul de dezvoltare a infrastructurii, ș.a.m.d.

Toți factorii menționați influențează valoarea de piață a terenurilor, care la urmă, determină unicitatea pieței locale. Acești factori afectează direct piața terenurilor în condiții de formare a prețurilor. Totodată, este destul de problematic să evaluezi unul sau alt factor în condițiile evaluării terenurilor pentru expropriere pentru utilitate publică. În cazul exproprierii, prețul de piață a bunurilor imobile se formează conform propriilor legi și, de regulă, în mare parte au un caracter special datorită condițiilor speculative formate din relațiile dintre proprietari cât și dintre alți participanți la procesul de expropriere.

Comportamentul consumatorului se schimbă constant, iar nevoile individului se înmulțesc datorită progresului tehnologiei și a științei care conduce la inventarea de noi nevoi reale și artificial. De obicei, în piața imobiliară nu se intervine atunci când vine vorba de economiile pieței, însă guvernele sunt cele care stabilesc infrastructura în care se desfășoară piața imobilelor, ceea ce asigură exact acele servicii de care are nevoie sectorul privat pentru a prospera. (ROTARU, 2021, pg. 262-269) Principalii participanți în această piață sunt:

Proprietarul/Utilizatorul – cumpără casele ca o investiție și/ sau locuiesc în ele.

- Proprietarul este investitorul pur, cel ce nu consumă bunul imobiliar pe care îl cumpără și care de obicei închiriaza proprietatea altui individ.
- Rentierul este consumator pur.
- Dezvoltatorul (Developer) este cel care pregătește terenul pentru construcție, producând un nou bun pe piață.
- Renovatorul este cel care furnizează pieței clădiri remobilate.
- Intermediarii sunt băncile, avocații, brokerii imobiliari, etc. ce întreprind actul de cumpărare și vânzare al unei proprietăți imobiliare.

Durabilitatea. Imobilul, prin definiție, este durabil. O clădire poate exista timp de decenii sau chiar secole, iar pământul este indestructibil prin esența sa. Din această cauză piețele imobiliare sunt de tip „stock/flow”. Teoretic vorbind, aproximativ 95% din ofertă constă în stocul caselor existente, în timp ce doar 5% reprezintă fluxul noilor dezvoltări. Stocul ofertei imobiliare în orice perioadă este determinat de stocul existent în perioada anterioară, rata deteriorării stocului existent, rata renovării stocului existent, și fluxul noilor dezvoltări în perioada curentă. Efectul ajustărilor pieței imobiliare tinde să fie diminuat de către stocul relativ mare de clădiri existente, plus noile construcții care încep să apară tot mai multe în cadrul pieței imobiliare.

Eterogenitatea. Fiecare bun imobiliar este unic, atât în termeni de locație, de specific al clădirii, cât și în termeni de finanțare. Această caracteristică face dificilă stabilirea prețului, crează asimetrie de informații și restricții de substituibilitate. Stocul imobiliar se depreciază, având astfel o calitate diferită în raport cu o nouă clădire. Din acest motiv procesul de echilibrare a pieței operează cu nivel de calitate diferite. Astfel, piața imobiliară poate fi divizată în segmente – rezidențial, comercial și industrial. De asemenea, mai poate fi divizată în subcategorii, ca de exemplu – imobile destinate recreerii, imobile generatoare de venituri, zone istorice, zone protejate, piața primară și secundară etc.

Costuri de tranzacție ridicate. Cumpărarea și/sau transferul într-o casă, costă mult mai mult decât majoritatea altor tipuri de tranzacții. Aceste costuri includ costurile de căutare, taxele imobiliare, costurile de transfer, taxele legale, taxele de transfer, și taxele de înregistrare a contractului. De regulă, costurile de tranzacție pentru vânzător reprezintă 8 – 10% din prețul de cumpărare. Procesul de ajustare a pieții poate fi subiectul unor întârzieri, datorită duratei de timp necesară atât finanțării, design-ului și construirii noii oferte, cât și datorită ratei scăzute de modificare a cererii. Datorită acestor decalaje de timp există o mare probabilitate de dezechilibrare a pieței pe termen scurt. Astfel, mecanismele de ajustare pe piața imobiliară tind să fie foarte lente și relative în raport cu alte piețe mai fluide.

Bunul imobiliar este atât un bun de investiție, cât și un bun de consum. Un bun imobiliar poate fi cumpărat cu scopul obținerii unui venit (bun de investiție), cu intenția de a-l utiliza în folos propriu (bun de consum), sau, chiar cu ambele intenții. Aceste funcții ale bunului imobiliar pot fi separate (participanții pe piață se concentrează asupra unei singure funcții) sau pot fi combinate (de exemplu, în situația persoanelor care locuiesc în casă, dar care o și închiriază parțial în același timp). Această natură duală a bunului imobiliar conduce la fenomenul de supra – investiție (engl. „over – invest”) în bunul imobiliar, adică, de a investi mai mulți bani într-un activ decât valorează pe piața liberă.

Imobilitatea. Pe piața imobiliară, consumatorii sunt cei care vin la produs, iar din această cauză nu putem vorbi de un spațiu fizic pentru piață. De exemplu, dacă există preferința de a locui în zone suburbane, oamenii sunt cei care trebuie să se mute în zone suburbane, evident, fiind imposibil să-și mute casa în suburbii. În acest caz, se diminuează la maxim posibilitatea distrugerii / avarierii prin transferarea/ transportarea produsului.

Pornind de la idea că locuințele din România, printre care se numără și cele din București sunt în mare majoritate clădiri vechi, construite înainte de anul 1977, și că durata de viață a unei clădiri este de 50 de ani, vârstă pe care multe blocuri au depășit-o acum ceva timp, prezintă grave degradări, astfel oamenii sunt în căutare de locuințe noi în ultimul deceniu, atât în ceea ce privește cererea, cât și în ceea ce privește oferta. Majoritatea investițiilor în imobiliare s-au efectuat în București-Ilfov datorită numărului mare al populației pe o suprafață destul de mică, a veniturilor medii ridicate, cât și fluxului investițiilor din străinătate directe către anumite sectoare. Conform evoluției costurilor, dar a prețurilor din acest sector economic, instabilitatea economică îl “încurajează” pe consumator să aleagă între a-și renova imobilul sau a-și construi propria casă în regie proprie, ce presupune ridicarea unui imobil cu forțele proprii, fără a apela la serviciile unei companii de construcții. (Drd. Elena IONAȘCU, 2019, pg. 212-236)

Din punct de vedere economic, piața imobiliară reprezintă totalitatea tranzacțiilor în urma cărora sunt implicate drepturile de proprietate sau folosință asupra terenurilor și clădirilor, iar schimbul este efectuat pe baza unei recompense, de obicei a unei sume de bani. Ea este cu mult mai complexă, necesită un grad ridicat de resurse financiare la derularea unei tranzacții, este necesară implicarea obligatorie a statului, iar frecvența unei asemenea operațiuni comerciale, raportate la un individ, este foarte scăzută.

În urma analizei evoluției pieței imobiliare, a pieței imobiliare la nivel național, ce reprezintă și situația de actualitate din București, sunt ușor de depistat părțile tari și părțile slabe în activitatea acestor domenii. Faptul trebuie să motiveze administrațiile publice centrale, a celor locale și a factorilor de decizie să nu permită sau să preîntâmpine impactul factorilor negativi ce au avut loc pe piețele respective, dar totodată să introducă mecanismele și instrumentele cele mai eficiente, aplicate pe piețele date. (TUREȚCHI, pg. 435-439)

1.2 Metodologia cercetării

Aplicația studiază influențele asupra prețului imobilelor din București la nivelul anului 2019. Este folosit software-ul RStudio și limbajul și mediul R pentru calcul statistic și grafică.

$$Y = \alpha + \beta X$$

unde:

Y – variabila dependentă, prețul imobilului

X – variabila independentă, suprafața imobilului

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3$$

unde:

X₁ – suprafața imobilului

X₂ – totalul etajelor clădirii în care se află imobilul

X₃ – etajul la care se află imobilul

Pentru testarea validității modelului am calculat indicatorii de bonitate, aceștia fiind coeficientul de determinație, coeficientul de nedeterminație, coeficientul de determinație ajustat, raportul de corelație. Totodată, am testat ipotezele modelului de regresie liniară și semnificația parametrilor din model.

În cadrul aplicației sunt utilizate următoarele teste:

- Pentru Homoschedasticitate:
 - Testul Breusch-Pagan
 - Testul White
- Pentru autocorelare:
 - Testul Durbin-Watson
 - Testul Breusch-Godfrey
- Pentru normalitate:
 - Testul Jarque-Bera
 - Testul Shapiro-Wilk

2.1 Date utilizate

Fișierul csv conține date referitoare la prețul de vânzare al imobilelor din București, România în martie 2019. Setul de date este compus din 6 variabile independente: numărul de camere, suprafață, etaj, numărul total de etaje din clădire, sectorul locuinței și un scor al locației. Variabila dependentă este reprezentată de prețul fiecărei locuințe în parte.

Modelul economic care descrie factorii determinanți ai prețului unei locuințe din București este:

$$Pret = f(NrCamere, Suprafata, Etaj, TotalEtaje, Sector, Scor)$$

- Prin variabila *NrCamere* se înțelege numărul total de camere din locuință. Garsonierele au un preț mai mic la achiziție din cauza spațiului insuficient de depozitare, de desfășurare și de compartimentare. Cu cât numărul camerelor este mai mare, cu atât crește și prețul imobilului. Un număr mare de camere oferă spațiu optim de locuit pentru cupluri, familii și tineri, camere mai spațioase, loc pentru animalele de companie și mai ales intimitatea.
- Variabila *Suprafata* reprezintă spațiul total măsurat în m² ai locuinței. După mulți ani în care locația a jucat rolul principal în luarea deciziei de cumpărare, carantina, mutarea biroului și a școlii acasă, dar și limitarea severă a călătoriilor, a activităților sociale și de relaxare au înclinat balanța către locuințe mai spațioase și mai confortabile, cu accent pe compartimentare eficientă și pe optimizarea spațiului, dar și pe spațiile exterioare precum terase sau grădini. O suprafață cât mai mare influențează creșterea prețului deoarece există mai mult spațiu pentru lucruri, pasiuni, familie, depozitare, muncă și familie.
- Variabila *Etaj* indică etajul la care imobilul este situat. Apartamentele aflate la parter sau la ultimul etaj au un preț mai scăzut decât cele aflate la etajele intermediare. Mansardele și subsolurile fiind cele mai evitate amplasări, pot fi ușor mai ieftine.

Apartamentele de la ultimul etaj sunt în contact direct permanent cu vremea de afară și condițiile meteo, iar în cazul în care blocul nu este izolat bine, este posibil ca vara să se resimtă în casă aproximativ tot atâtea grade câte sunt și afară, chiar dacă este pornit aerul condiționat. De asemenea, iarna este posibil să fie nevoie de soluții suplimentare de încălzire. Chiar dacă blocul a fost remediat și izolat de curând, este posibil ca lucrările să nu fie făcute la standardele corespunzătoare. Izolația precară poate duce la infiltrații și igrasie, pentru că între perete și terasa blocului se creează condens sau apa se infiltrează

prin crăpături. Totodată, în blocurile vechi, apa caldă poate ajunge mult mai greu la etajele superioare.

Un apartament la parter are un preț mai mic, în comparație cu un apartament asemanator la etaj intermediar. Acesta suferă de lipsa de intimitate, lipsa de lumină naturală, supunerea la un nivel mai mare de zgomot și pierderile de caldură frecvente. Parterul este mai friguros decât apartamentele de la etaj intermediar (5 grade este aproximativ diferența de temperatură dintre parter și un etaj intermediar). Pierderile de caldură aduc implicit costuri mai mari, mai ales în perioadele de trecere dintre anotimpuri.

- Variabila *TotalEtaje* arată numărul total de etaje care formează imobilul. Un imobil cu un singur nivel limitează o bună parte din priveliște și din lumina naturală. Dacă nu se acordă o atenție sporită izolației, s-ar putea resimți temperaturile extreme. Astfel, vara e posibil fie prea cald, iar iarna prea frig.

Prețul crește o dată cu creșterea numărului de etaje deoarece există mai mult spațiu locuibil pe o suprafață de teren mai mică, există mai multă lumină naturală și mai mult loc pentru grădină. Un imobil cu etaj permite o organizare mai bună a spațiilor locuinței. Cel mai adesea parterul este dedicat zonei de zi în timp ce etajul adăpostește dormitoare.

- Variabila *Sector* indică sectorul în care este poziționat imobilul, în cadrul Municipiului București. Prețul este influențat în mare măsură de poziționarea imobilului deoarece, ca în orice alt oraș, există zone mai îngrijite, mai liniștite, mai curate, mai bogate în mijloace de transport importante și lista poate continua.

Vom alege ca exemplu sectorul 5. La ora actuală, Rahova este un cartier al contrastelor, extinzându-se pe o suprafață mare și cuprinzând atât părți mai sărace, care nu sunt văzute mereu cu ochi buni de bucureșteni, cât și zone frumoase și destul de scumpe, precum cea aflată în apropierea Parcului Sebastian.

Un avantaj al zonei Rahova este acela că ea se întinde până în apropierea centrului Capitalei și este bine conectată cu acesta, cât și cu alte părți ale orașului prin numeroase mijloace de transport în comun. Aglomerația de pe principalele artere de circulație, numărul limitat de locuri de parcare, lipsa spațiilor verzi de mari dimensiuni și lipsa unui metrou reprezintă câteva dintre dezavantajele cartierului.

- Prin variabila *Scor* se înțelege un indice de apreciere generală a locuinței de la 1 la 5, valori întregi, 1 însemnând părere foarte proastă / nerecomandare, iar 5 părere foarte bună / recomandare. Prelucrarea adusă este reprezentată de transformarea variabilei *Scor* într-o variabilă de tip dummy, astfel: considerăm scorul peste pragul de 3.5 un scor bun, care primește valoarea 1, iar cel sub 3.5 valoarea 0.
- Variabila *Preț* reprezintă suma totală de bani în mii euro pe care ar trebui să o plătească cumpărătorul pentru achiziționarea unui imobil. Cu cât prețul este mai mare, cu atât imobilul satisface mai multe nevoi și dorințe.

Sursa de colectare a datelor este <https://www.kaggle.com/datasets/denisadutca/bucharest-house-price-dataset>, ce au fost colectate de pe www.imobiliare.ro, cel mai popular site imobiliar din România.

2.2 Rezultate empirice

Regresia simplă

Evoluția economică a variabilelor

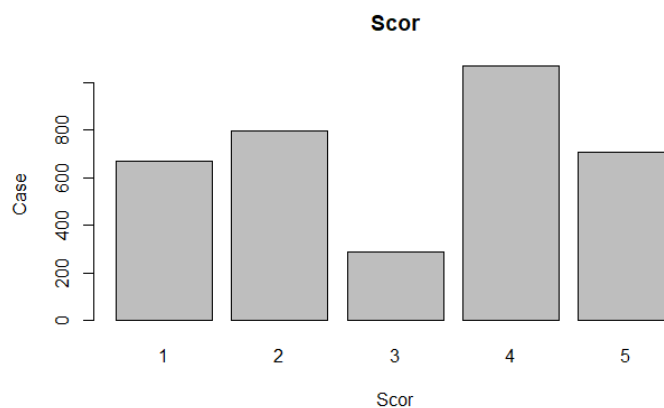


Figure 1: Graficul scorului imobilelor

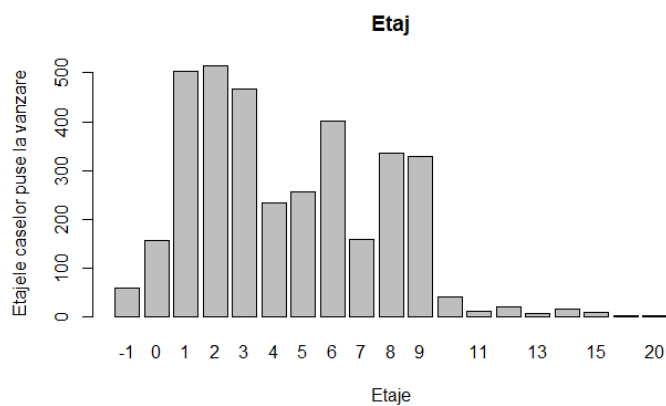


Figure 2: Graficul etajelor imobilelor

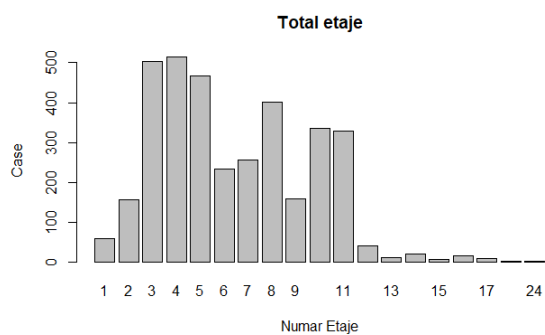


Figure 3: Graficul numărului total de etaje al clădirilor în care se află imobilele

Estimarea și interpretarea parametrilor

Dintre modelele cu coeficienți semnificativi, acela cu cel mai mare R^2 este cel care conține suprafața drept variabilă independentă.

```

Residuals:
    Min       1Q   Median       3Q      Max
-193.045  -20.529   -2.952   16.281   211.529

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.93933    1.62326  -14.75  <2e-16 ***
Suprafata    1.74447    0.02146   81.28  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.08 on 3527 degrees of freedom
Multiple R-squared:  0.6519,    Adjusted R-squared:  0.6518
F-statistic: 6606 on 1 and 3527 DF,  p-value: < 2.2e-16

```

Table 1: Tabelul modelului de regresie liniară unifactorială

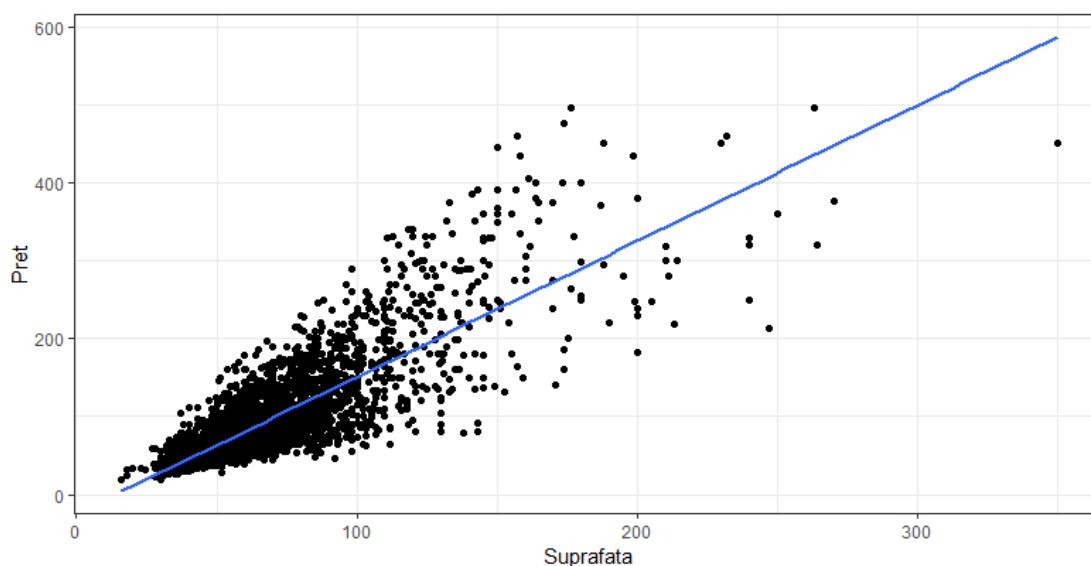


Figure 4: Dreapta de regresie pentru modelul unifactorial

Dreapta de regresie estimată este:

$$Pret = -23.93933 + 1.74447 * Suprafata$$

De aici reiese că:

- $\alpha = -23.93933$ exprimă nivelul prețului imobilelor din București atunci când suprafața acestuia tinde către 0.

- $\beta = 1.74447$ arată că prețul unui imobil din București se modifică în medie cu 1.74447 mii euro atunci când suprafața crește cu un m^2 .

Verificarea semnificației parametrilor

- $H_0: \alpha = 0, \beta = 0 \Rightarrow$ parametri sunt semnificativi, modelul nu este valid
- $H_1: \alpha \neq 0, \beta \neq 0 \Rightarrow$ parametri sunt semnificativi, modelul este valid

$$Pr = 2.2 * e^{-16} = 0.0000000000000022 < 0.01$$

Din tabel rezultă că ambii coeficienți sunt semnificativi la un nivel de semnificație de 99%, deoarece probabilitatea ambilor coeficienți are o valoare de $2.2 * e^{-16}$, care este mai mică decât 0.01, deci respingem ipoteza nulă și o acceptăm pe cea alternativă.

Testarea validității modelului

- H_0 : modelul nu este valid dpdv statistic
- H_1 : modelul este valid dpdv statistic

$$F - \text{statistic} = 6606$$

Deoarece F-statistic este semnificativ mai mare decât F-critic, este respinsă ipoteza nulă și acceptată cea alternativă, conform căreia modelul este valid pentru un nivel de semnificație $p - \text{value} = 2.2 * e^{-16} < 0.01$, deci la 99%.

Indicatori de bonitate

- Coeficientul de determinație

$R^2 = R - \text{squared} = 0.6519 \Rightarrow$ suprafața unui imobil explică aproximativ 65.19% din variația prețului acestuia

- Coeficientul de nedeterminație

$K^2 = 1 - R^2 = 0.3481 \Rightarrow$ acțiunea altor factori explică restul de 34.81% din variația prețului acestuia

- Coeficientul de determinație ajustat

$Adjusted R^2 = Adjusted R - \text{squared} = 0.6518 \Rightarrow$ 65.18% din variația prețului unui imobil este explicată de suprafața acestuia

- Raportul de corelație

$R = 0.8074 \Rightarrow$ între cele două variabile există o legătură directă, destul de puternică

Verificarea îndeplinirii ipotezelor modelului de regresie liniară

Am filtrat datele, astfel încât acestea să fie cât mai omogene, respectiv am eliminat valorile de tip outlier, păstrând doar imobilele cu suprafețe mai mici decât 90 mp și mai mari de 40 mp, iar marja prețului este de 30 până la 150 mii de euro.

- Verificarea homoscedasticității erorilor aleatoare

- Graficul Residuals vs Fitted

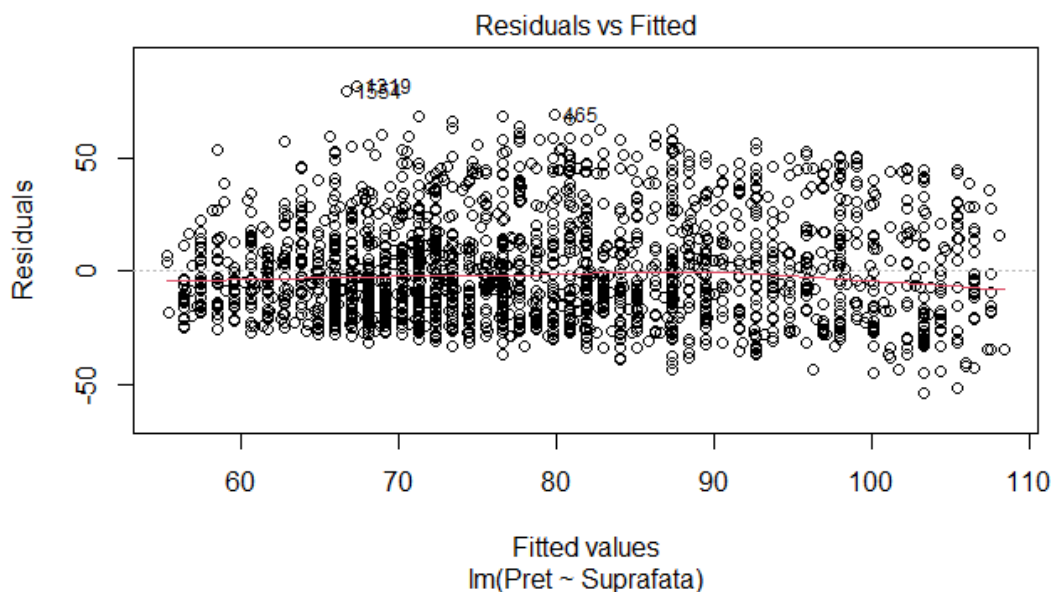


Figure 5: Graficul Residuals vs Fitted pentru regresia simplă

Graficul arată dispersia reziduurilor pe axa Oy și valoarea ajustată pe axa Ox. Se poate observa că reziduurile nu se răspândesc aleatoriu în jurul valorii 0, fiind mai multe în partea de sus, deci este un indiciu că avem heteroscedasticitate.

- Graficul Q-Q plot

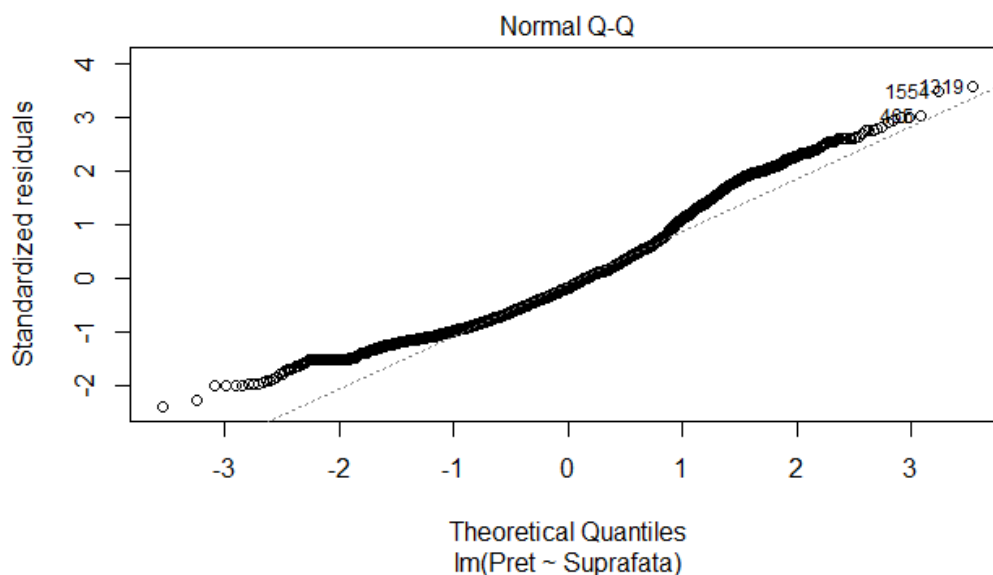


Figure 6: Graficul Q-Q plot pentru regresia simplă

Graficul Q-Q plot este un grafic de dispersie care prezintă două seturi de cuantile. Pe Ox sunt trasate cuantilele distribuției normale, iar pe Oy cele ale distribuției reziduale. Diagrama nu formează o linie diagonală și se pot observa mai multe valori extreme, sau care nu se află pe linie.

În urma testelor Breusch-Pagan și White valorile p-value au fost sub 0.05, deci s-a confirmat ipoteza cum că reziduurile modelului sunt heteroscedastice. Am corectat modelul folosind log-lin, iar următoarele sunt noile rezultate ale testelor:

- TESTUL BREUSCH-PAGAN

```
studentized Breusch-Pagan test

data:  model_loglog
BP = 0.015617, df = 1, p-value = 0.9005
```

Table 2: Testul Breusch-Pagan pentru regresia simplă

- H_0 : erorile sunt homoscedastice (reziduriile sunt distribuite egal)
- H_1 : erorile sunt heteroscedastice (reziduriile nu sunt distribuite egal)

$p - value = 0.9005 > 0.1 \Rightarrow$ se acceptă ipoteza nulă, cum că erorile sunt homoscedastice.

○ TESTUL WHITE

white's test results

Null hypothesis: Homoskedasticity of the residuals

Alternative hypothesis: Heteroskedasticity of the residuals

Test Statistic: 2.44

P-value: 0.294971

Table 3: Testul White pentru regresia simplă

- H_0 : erorile sunt homoscedastice (reziduurile sunt distribuite egal)
- H_1 : erorile sunt heteroscedastice (reziduurile nu sunt distribuite egal)

$p - value = 0.295 > 0.1 \Rightarrow$ se acceptă ipoteza nulă

Conform ambelor teste, erorile sunt homoscedastice, pentru un nivel de semnificație de 99%.

b) Detectarea non-autocorelării

○ Graficul ACF

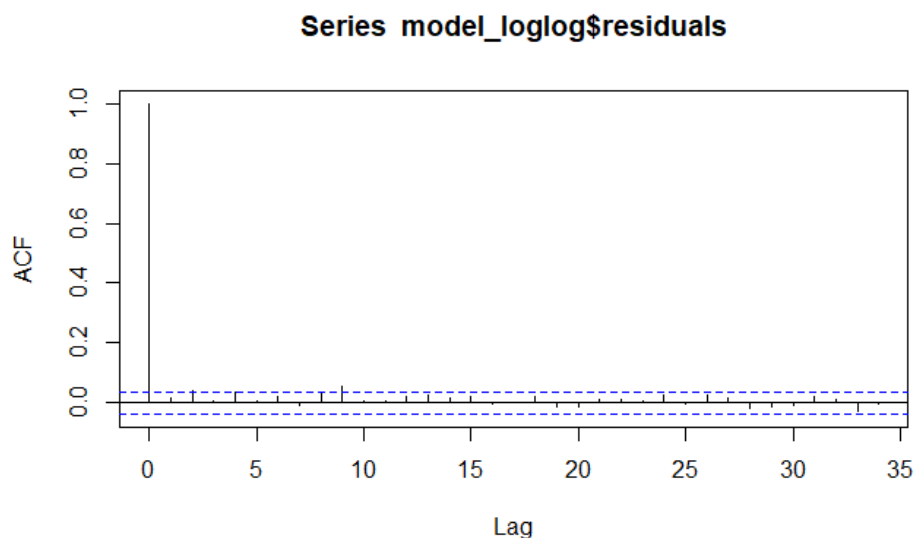


Figure 7: Graficul ACF pentru regresia simplă

Graficul dă de înțeles că modelul nu are reziduuri auto-corelate, deoarece există un singur lag care depășește intervalul punctat.

- TESTUL DURBIN-WATSON pentru auto-corelarea de gradul I

Durbin-Watson test

```
data: model_loglog
DW = 1.9644, p-value = 0.1788
alternative hypothesis: true autocorrelation is greater than 0
```

Table 4: Testul Durbin-Watson pentru regresia simplă

- H_0 : reziduurile nu sunt autocorelate
- H_1 : reziduurile sunt autocorelate

$p - value = 0.1788 > 0.1$ si $DW \approx 2 \Rightarrow$ acceptăm ipoteza nulă, modelul nu prezintă autocorelare de gradul I.

- TESTUL BREUSCH-GODFREY

Breusch-Godfrey test for serial correlation of order up to 2

```
data: model_loglog
LM test = 4.9198, df = 2, p-value = 0.08544
```

Table 5: Testul Breusch-Godfrey de ordinul II pentru regresia simplă

- H_0 : reziduurile nu sunt autocorelate
- H_1 : reziduurile sunt autocorelate

$p - value = 0.0854 > 0.05$ si $LM \approx 5 \Rightarrow$ acceptăm ipoteza nulă, modelul nu prezintă autocorelare de gradul II pentru o probabilitate de 95%.

Breusch-Godfrey test for serial correlation of order up to 3

```
data: model_loglog
LM test = 4.96, df = 3, p-value = 0.1748
```

Table 6: Testul Breusch-Godfrey de ordinul III pentru regresia simplă

- H_0 : reziduurile nu sunt autocorelate
- H_1 : reziduurile sunt autocorelate

$p - value = 0.1748 > 0.1$ și $LM \approx 5 \Rightarrow$ acceptăm ipoteza nulă, modelul nu prezintă autocorelare de gradul III.

c) Normalitatea distribuției erorilor aleatoare

○ Histograma reziduurilor

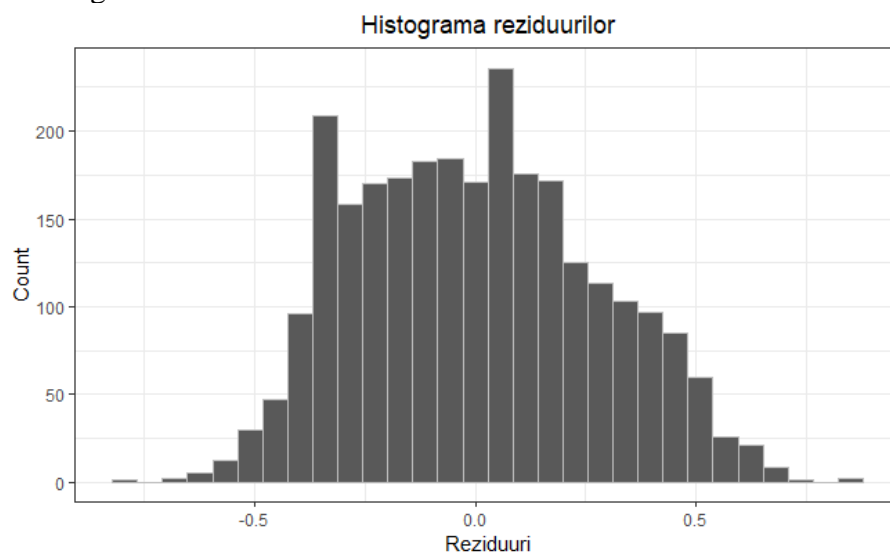


Figure 5: Histograma reziduurilor pentru regresia simplă

Din figura 5 se poate observa că distribuția reziduurilor nu este normală, neavând formă de clopot.

○ TESTUL JARQUE-BERA

Jarque Bera Test	
data:	data_cook\$uhat
X-squared	= 58.621, df = 2, p-value = 1.864e-13

Table 7: Testul Jarque-Bera pentru regresia simplă

- H_0 : erorile aleatoare au o distribuție normală
- H_1 : erorile aleatoare nu au o distribuție normală

$p - value = 1.864 * e^{-13} < 0.1 \Rightarrow$ respingem ipoteza nulă, o acceptăm pe cea alternativă, deci erorile nu au o distribuție normală.

○ TESTUL SHAPIRO-WILK

Shapiro-wilk normality test

```
data: data_cook$uhat  
W = 0.98786, p-value = 5.619e-14
```

Table 6: Testul Shapiro-Wilk pentru regresia simplă

- H_0 : erorile aleatoare au o distribuție normală
- H_1 : erorile aleatoare nu au o distribuție normală

$p - value = 5.619 * e^{-14} < 0.1 \Rightarrow$ respingem ipoteza nulă, o acceptăm pe cea alternativă, deci erorile nu au o distribuție normală.

Proгноze

Am împărțit setul de date într-un set de antrenare (aprox. 80%) și unul de testare (20% dintre observații). Modelul final corectat este de tipul:

$$\text{Log}(\text{Pret}) = f(\text{Log}(\text{Suprafata}))$$

Indicatorul RMSE măsoară acuratețea și arată cât de departe se încadrează predicțiile față de valorile reale măsurate folosind distanța euclidiană.

$$RMSE = 0.2733 < 1 \Rightarrow \text{predicția este bună}$$

Indicatorul MAE funcționează similar cu RMSE, cu excepția că toate diferențele individuale sunt ponderate în mod egal în medie.

$$MAE = 0.2272 < 1 \Rightarrow \text{predicția este bună}$$

Indicatorul MSE măsoară în unități care este pătratul variabilei țintă și penalizează mai mult și mai sever erorile.

$$MSE = 0.0747 > 0 \Rightarrow \text{predicția nu este bună}$$

Indicatorul MAPE măsoară ce indică media dispersiei dintre valorile previzionate și cele reale.

$$MAPE = 0.0523 < 1 \Rightarrow \text{predicția e bună}$$

1	2	3
63.29940	73.83227	103.96964

Table 7: Tabelul prognozei regresiei simple

Tabelul indică valorile în mii euro a prețurilor imobilelor cu suprafața de 50 m², 60 m², respectiv 90 m².

Regresia multiplă

Estimarea și interpretarea parametrilor

Am ales modelul de regresie multiplă pentru prețul caselor din București format din variabilele independente Suprafata, TotalEtaje și Etaj pentru că variabilele sunt semnificative din punct de vedere statistic, cu mențiunea că am filtrat din nou datele, astfel încât totalul etajelor să fie sub 10, fiindcă existau foarte multe valori de tip outlier.

```

Residuals:
    Min       1Q   Median       3Q      Max
-60.852 -17.761  -4.028   13.453   82.502

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.01829    2.23524   5.377 8.27e-08 ***
Suprafata     0.99994    0.03041  32.884 < 2e-16 ***
TotalEtaje    0.95659    0.17080   5.601 2.36e-08 ***
Etaj          -0.51724    0.19620  -2.636 0.00843 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.77 on 2562 degrees of freedom
Multiple R-squared:  0.3012,    Adjusted R-squared:  0.3004
F-statistic: 368.1 on 3 and 2562 DF,  p-value: < 2.2e-16

```

Table 8: Tabelul modelului de regresie multiplă

Dreapta de regresie estimată este:

$$Pret = 12.01829 + 0.99994 * Suprafata + 0.95659 * Total.Etaje - 0.5172 * Etaj$$

De aici reiese că:

- $\alpha = 12.01829$ – exprimă nivelul prețului imobilelor din București atunci când suprafața, etajul acestuia și numărul total de etaje al clădirii tind către 0.

- $\beta_1 = 0.99994$ – arată că prețul unui imobil din București se modifică în medie cu 0.99994 mii euro atunci când suprafața crește cu un m^2 .
- $\beta_2 = 0.95659$ – arată că prețul unui imobil din București se modifică în medie cu 0.95659 mii euro atunci când numărul total de etaje crește cu un etaj.
- $\beta_3 = -0.5172$ – arată că prețul unui imobil din București se modifică în medie cu -0.5172 mii euro atunci când etajul se modifică la o unitate superioară.

Verificarea semnificației parametrilor

Din tabel rezultă că toți coeficienții sunt semnificativi la un nivel de semnificație de 99%, deoarece probabilitatea lor au valori mai mici decât 0.01, deci respingem ipoteza nulă și o acceptăm pe cea alternativă \Rightarrow parametrii sunt semnificativi, modelul este valid.

Testarea validității modelului

$$F - statistic = 368.1$$

Deoarece F-statistic este semnificativ mai mare decât F-critic, este respinsă ipoteza nulă și acceptată cea alternativă modelul este valid pentru un nivel de semnificație $p - value = 2.2 * e^{-16} < 0.01$, deci la 99%.

Indicatori de bonitate

- Coeficientul de determinație

$R^2 = R - squared = 0.3012 \Rightarrow$ suprafața, numărul total de etaje și etajul unui imobil explică aproximativ 30,20% din variația prețului acestuia.

- Coeficientul de nedeterminație

$K^2 = 1 - R^2 = 0.6988 \Rightarrow$ acțiunea altor factori explică restul de 69.88% din variația prețului acestuia.

- Coeficientul de determinație ajustat

$Adjusted R^2 = Adjusted R - squared = 0.3004 \Rightarrow$ 30.04% din variația prețului unui imobil este explicată de suprafața, numărul total de etaje și etajul acestuia.

- Raportul de corelație

$R = 0.5488 \Rightarrow$ între variabilele modelului există o legătură directă, destul de puternică.

Verificarea îndeplinirii ipotezelor modelului de regresie liniarăa) Ipoteza 1 - Este modelul liniar în parametri?

Da, deoarece poate fi scris ca o funcție liniară:

$$Pret = 12.01829 + 0.99994 * Suprafata + 0.95659 * TotalEtaje - 0.5172 * Etaj$$

b) Ipoteza 2 - Nr de observații > nr variabile independente

```
> nobs(model_rm) > (model_rm$rank - 1)
[1] TRUE
```

c) Ipoteza 3 - Variabilitatea în x este pozitivă

- $var(Suprafata) = 219.9197$
- $var(TotalEtaje) = 9.85714$
- $var(Etaj) = 7.497803$

Toate valorile > 0 => ipoteza nulă este acceptată

d) Ipoteza 4 - Media reziduurilor este 0

$$mean(residuals) = 6.43988 * e^{-16} \sim 0$$

Medie aproape de 0 => ipoteză acceptată

e) Ipoteza 5 - Testare multicolinearitate

Suprafata	TotalEtaje	Etaj
1.005996	1.422534	1.427903

Table 9: Tabelul valorilor VIF

Nu există valori pt VIF > 10 => ipoteză acceptată

f) Ipoteza 6 - Reziduurile nu sunt corelate cu variabilele independente

$p - value > 0.1$ pentru toate cele 3 variabile => nu sunt corelate

g) Ipoteza 7 - Reziduurile sunt homoscedastice

- Grafic

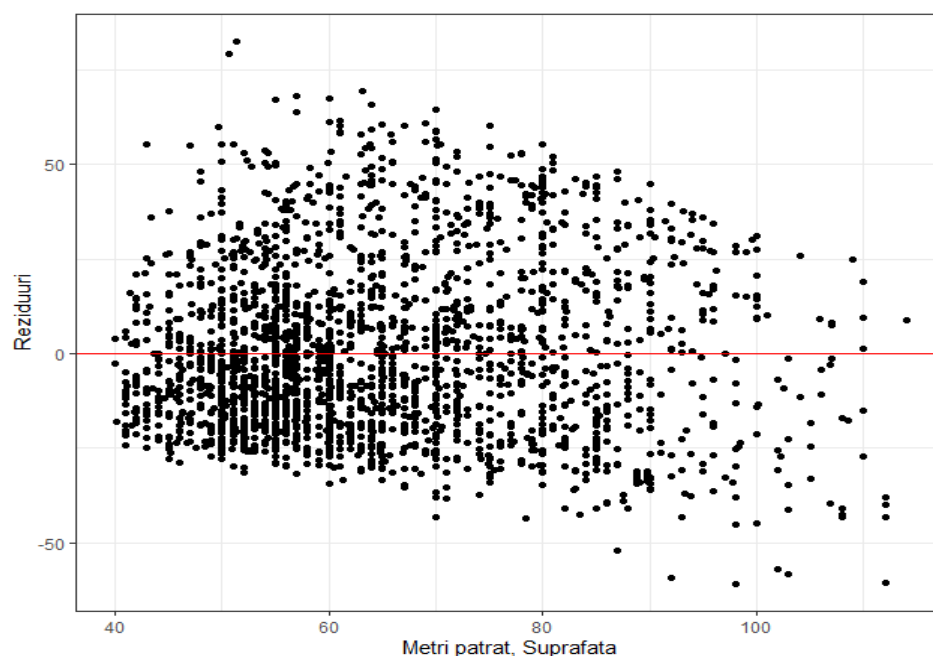


Figure 8: Graficul homoscedasticității pentru regresia multiplă

Graficul indica faptul că reziduurile nu se răspândesc aleatoriu în jurul valorii 0, fiind mai multe în partea de sus, ceea ce reprezintă un indiciu că avem heteroscedasticitate. Testele Bresch-Pagan, White și testul t al coeficienților au demonstrat heteroscedasticitatea datelor, ceea ce a dus la corectarea prin logaritmare și concluderea următorului model:

$$lPret = f(Suprafata, Etaj, lTotalEtaje)$$

- TESTUL BREUSCH-PAGAN

$$p - value = 0.4013 > 0.1 \Rightarrow \text{ipoteza nulă este acceptată}$$

- TESTUL WHITE

$$p - value = 0.493376 > 0.1 \Rightarrow \text{ipoteza nulă este acceptată}$$

În urma testelor Breusch-Pagan și White \Rightarrow ipoteza nulă este acceptată \Rightarrow erorile sunt homoscedastice

h) Ipoteza 8 - Reziduurile nu sunt autocorelate

- TESTUL DURBIN-WATSON

$$p - value = 0.2429 > 0.1$$

- TESTELE BREUSCH-GODFREY

$p - value \in \{0.5127, 0.3247, 0.5021\} > 0.1$ în toate cele 3 cazuri

=> ipoteza nulă este acceptată

=> reziduurile nu sunt autocorelate

=> modelul nu prezintă autocorelare de gradul I, II sau III

În urma testelor Durbin-Watson și Breusch-Godfrey reiese că reziduurile nu sunt autocorelate, deci nu va mai fi nevoie nicio corecție.

i) Ipoteza 9 - Reziduurile sunt normal distribuite

- Graficul 'Residuals vs Fitted'

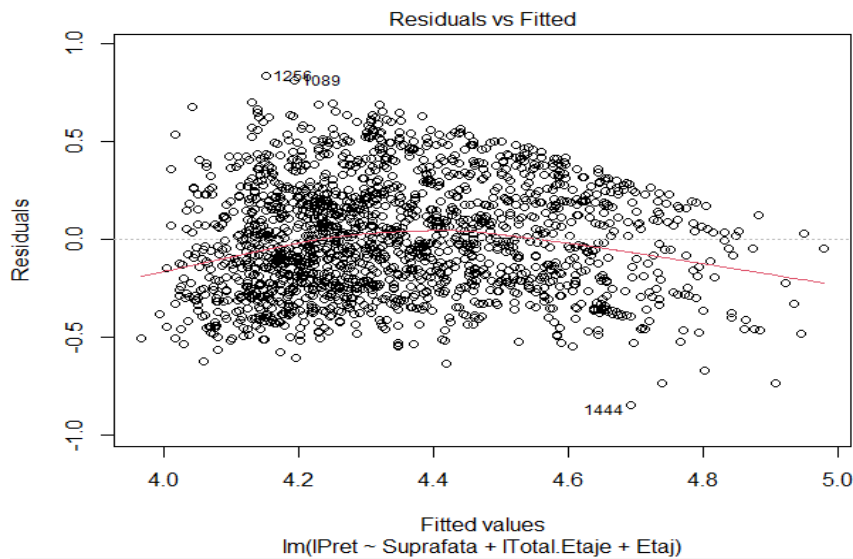


Figure 9: Graficul Residuals vs Fitted pentru regresia multiplă

Dacă linia roșie (trendul) este aproximativ dreaptă și aproape de zero, atunci se poate presupune că reziduurile sunt normal distribuite. În cazul de față reziduurile nu par a fi normal distribuite și de aceea este nevoie de mai multă investigație ca să se ofere un diagnostic final, iar testele Shapiro-Wilk, Kolmogorov-Smirnov, Jarque-Bera, Cramer-von-Mises și Anderson Darling, ale căror p-value au avut toate valori foarte mici, foarte apropiate de 0, au confirmat că reziduurile nu sunt normal distribuite. Următoarele valori sunt cele în urma corecției cu distanța Cook:

- Grafic

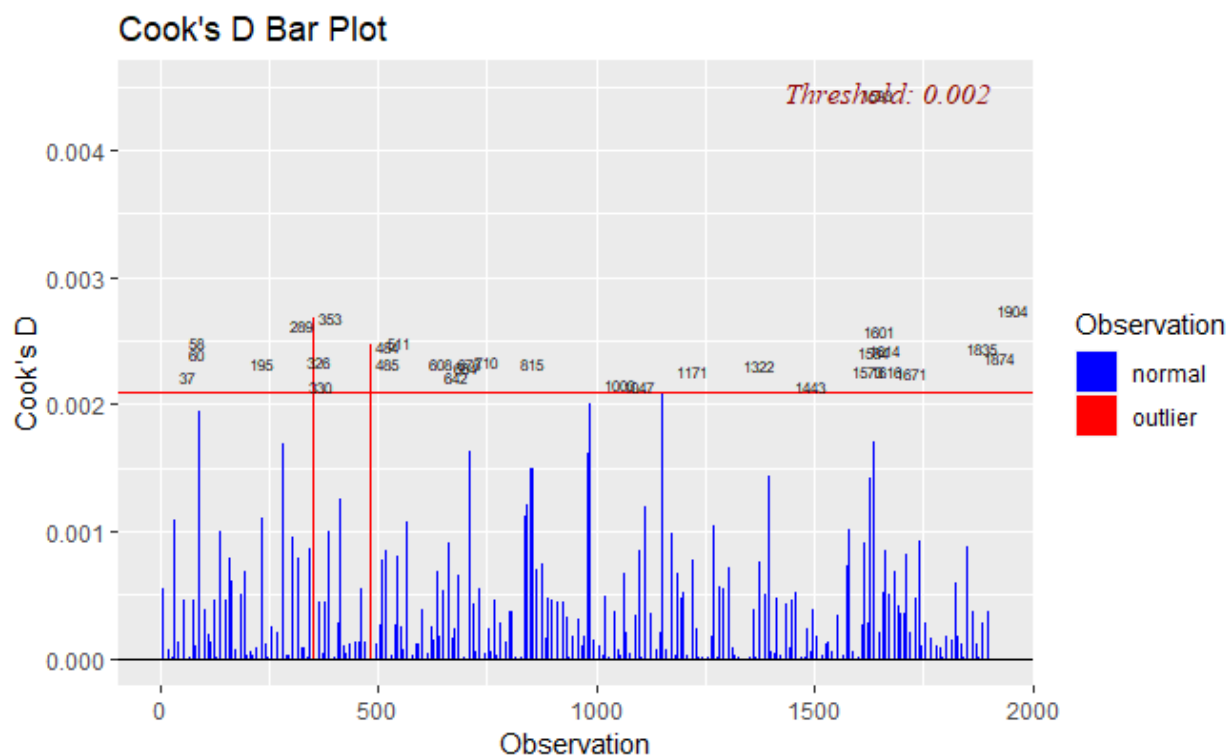


Figure 10: Graficul Cook's Bar Plot în urma corecției

Regresia cu variabilă dummy și termeni de interacțiune

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.7218935	0.0397847	93.551	< 2e-16	***
Suprafata	0.0106706	0.0005047	21.143	< 2e-16	***
lTotalEtaje	0.0406196	0.0136466	2.977	0.002952	**
Etaj	0.0061623	0.0028895	2.133	0.033081	*
Rating	-0.4638134	0.0482862	-9.606	< 2e-16	***
RatingXSupr	0.0027173	0.0007178	3.786	0.000158	***

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

Residual standard error: 0.2288 on 1900 degrees of freedom
 Multiple R-squared: 0.535, Adjusted R-squared: 0.5338
 F-statistic: 437.2 on 5 and 1900 DF, p-value: < 2.2e-16

Table 10: Tabelul regresiei cu variabilă dummy și termeni de interacțiune

Dreapta de regresie estimată este:

$$lPret = 3.721 + 0.0106 * Suprafata + 0.0406 * lTotalEtaje + 0.006 * Etaj - 0.463 \\ * Rating + 0.002 * RatingxSupr$$

De aici reiese că:

- $\alpha = 3.721$ – exprimă nivelul prețului imobilelor din București atunci când suprafața, etajul acestuia și numărul total de etaje al clădirii tind către 0.
- $\beta_1 = 0.0106$ – arată că prețul unui imobil din București se modifică în medie cu 0.0106 mii euro atunci când suprafața crește cu un m².
- $\beta_2 = 0.0406$ – arată că prețul unui imobil din București se modifică în medie cu 0.0406 mii euro atunci când numărul total de etaje crește cu un etaj.
- $\beta_3 = -0.006$ – arată că prețul unui imobil din București se modifică în medie cu -0.006 mii euro atunci când etajul se modifică la o unitate superioară.
- $\beta_4 = -0.463$ – arată diferența de preț al unui imobil din București atunci când scorul este unul bun, comparativ cu situația când nu este bun. Cum acesta este negativ, se deduce că pentru același etaj, număr total de etaje și aceeași suprafață, un scor mai mare corespunde unui preț mai mic.
- $\beta_5 = +0.002$ – arată că efectul suprafeței asupra prețului va crește, atunci când scorul este unul bun.

Verificarea semnificației parametrilor

Din tabel rezultă că suprafața, rating-ul, numărul total de etaje și termenul de interacțiune dintre suprafață și rating sunt semnificativi la un nivel de semnificație de 99%, deoarece probabilitatea lor au valori mai mici decât 0.01, deci respingem ipoteza nulă și o acceptăm pe cea alternativă => parametrii sunt semnificativi, modelul este valid. Variabila etaj este semnificativă la 95%.

Testarea validității modelului

$$F - statistic = 437.2$$

Deoarece F-statistic este semnificativ mai mare decât F-critic, este respinsă ipoteza nulă și acceptată cea alternativă modelul este valid pentru un nivel de semnificație $p - value = 2.2 * e^{-16} < 0.01$, deci la 99%.

Indicatori de bonitate

- Coeficientul de determinație

$R^2 = R - squared = 0.535 \Rightarrow$ suprafața, numărul total de etaje, etajul și scorul unui imobil explică aproximativ 53,5% din variația prețului acestuia.

- Coeficientul de nedeterminație

$K^2 = 1 - R^2 = 0.464 \Rightarrow$ acțiunea altor factori explică restul de 46.5% din variația prețului acestuia.

- Coeficientul de determinație ajustat

$Adjusted R^2 = Adjusted R - squared = 0.5338 \Rightarrow$ 53.38% din variația prețului unui imobil este explicată de suprafața, numărul total de etaje, etajul și scorul acestuia.

- Raportul de corelație

$R = 0.731 \Rightarrow$ între variabilele modelului există o legătură directă, puternică.

Modele de penalizare

- **Modelul Ridge**

	Length	Class	Mode
a0	100	-none-	numeric
beta	300	dgCMatrix	S4
df	100	-none-	numeric
dim	2	-none-	numeric
lambda	100	-none-	numeric
dev.ratio	100	-none-	numeric
nulldev	1	-none-	numeric
npasses	1	-none-	numeric
jerr	1	-none-	numeric
offset	1	-none-	logical
call	4	-none-	call
nobs	1	-none-	numeric

Table 11: Tabelul modelului de regresie Ridge

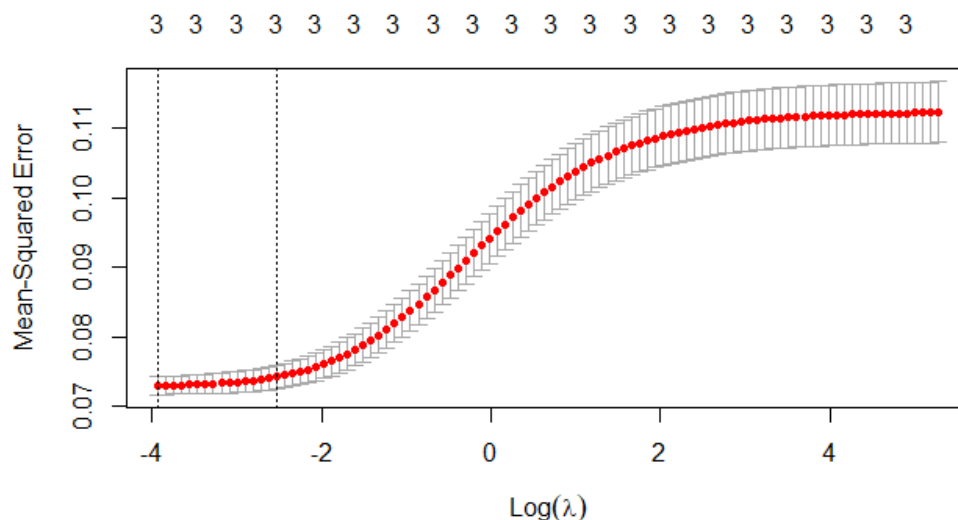


Figure 11: Plot regresia Ridge

Modelul de regresie Ridge încearcă să minimizeze $SSR + \lambda * \sum \beta^2$, unde $\lambda * \sum \beta^2$ se mai numește și *shrinkage penalty*. λ ia valoarea astfel încât să producă cea mai mică valoare pentru MSE. Pașii sunt:

- Definirea variabilei răspuns $y = \log(Pret)$
- Definirea predictorilor $x = \{Suprafata, \log(TotalEtaje), Etaj\}$
- Setarea valorii $\alpha = 0$
- Estimarea modelului Ridge
- Identificarea valorii λ pentru care avem MSE minimizat utilizând validarea încrucișată $\lambda = 0.01968$
- Testarea valorii λ și reimplementarea modelului cu valoarea optimă
- Estimarea coeficienților. Toți coeficienții au valori diferite de 0, deci sunt semnificativi modelului
- Calcularea bonității $R^2 = 35.27\%$

- **Modelul Lasso**

	Length	Class	Mode
a0	59	-none-	numeric
beta	177	dgCMatrix	S4
df	59	-none-	numeric
dim	2	-none-	numeric
lambda	59	-none-	numeric
dev.ratio	59	-none-	numeric
nulldev	1	-none-	numeric
npasses	1	-none-	numeric
jerr	1	-none-	numeric
offset	1	-none-	logical
call	4	-none-	call
nobs	1	-none-	numeric

Table 12: Tabelul modelului de regresie Lasso

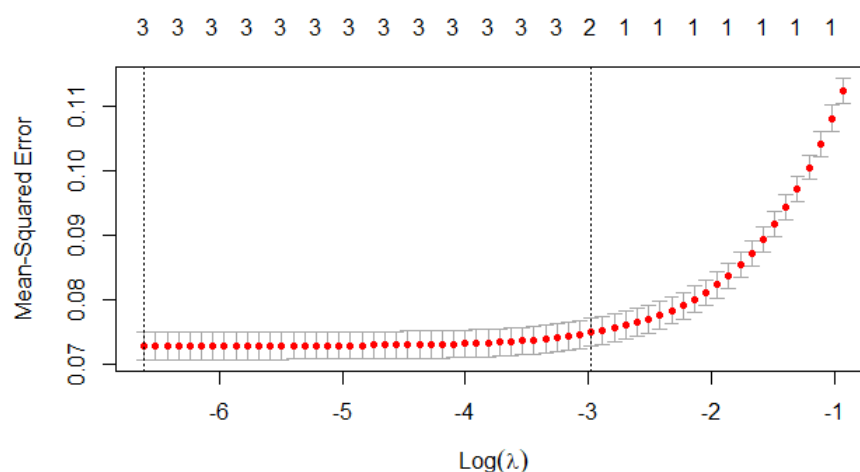


Figure 12: Plot regresia Lasso

Modelul Lasso încearcă să minimizeze $SSR + \lambda * \sum |\beta|$. Pașii sunt:

- Definirea variabilei răspuns $y = \log(Pret)$
- Definirea predictorilor $x = \{Suprafata, \log(TotalEtaje), Etaj\}$
- Setarea valorii $\alpha = 1$
- Estimarea modelului Lasso
- Identificarea valorii λ pentru care avem MSE minimizat utilizând validarea încrucișată $\lambda = 0.006$
- Testarea valorii λ și reimplementarea modelului cu valoarea optima

- Estimarea coeficienților. Toți coeficienții au valori diferite de 0, deci sunt semnificativi modelului
- Calcularea bonității $R^2 = 35.37\%$

• **Modelul Elastic Net**

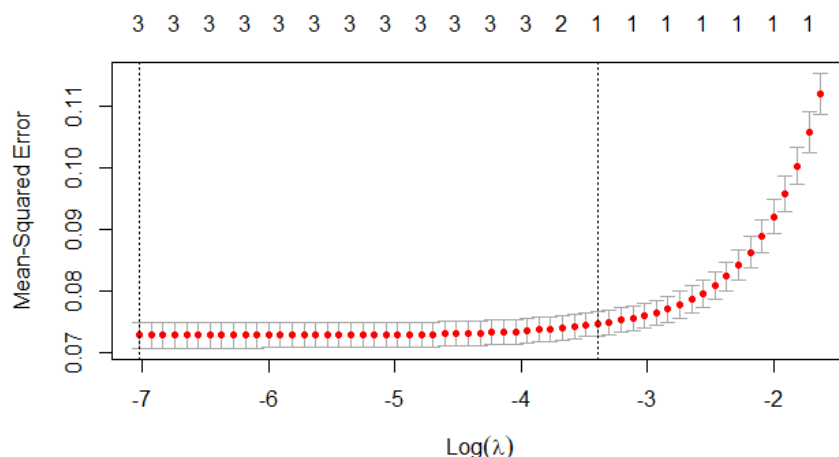


Figure 13: Plot regresia Elastic Net

Modelul Lasso încearcă să minimizeze $SSR + \lambda * \sum |\beta| + \lambda * \sum \beta^2$. Pașii sunt:

- Definirea variabilei răspuns $y = \log(Pret)$
- Definirea predictorilor $x = \{Suprafata, \log(TotalEtaje), Etaj\}$
- Setarea valorii $\alpha = 0.5$
- Estimarea modelului Lasso
- Identificarea valorii λ pentru care avem MSE minimizat utilizând validarea încrucișată $\lambda = 0.0013$
- Testarea valorii λ și reimplementarea modelului cu valoarea optimă
- Estimarea coeficienților. Toți coeficienții au valori diferite de 0, deci sunt semnificativi modelului
- Calcularea bonității $R^2 = 35.37\%$

- **Algoritmul Boruta**

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.7474932  0.0393314  95.280 < 2e-16 ***
Etaj         0.0073979  0.0039411   1.877 0.060655 .
Sector      -0.0318820  0.0037170  -8.577 < 2e-16 ***
NrCamere     0.0753213  0.0122049   6.171 8.26e-10 ***
Rating      -0.3855088  0.0483814  -7.968 2.75e-15 ***
Suprafata    0.0080368  0.0006402  12.553 < 2e-16 ***
lTotalEtaje  0.0576841  0.0134293   4.295 1.83e-05 ***
RatingXSupr  0.0025128  0.0007026   3.577 0.000357 ***
RatingXEtaj -0.0039591  0.0050009  -0.792 0.428641
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2224 on 1897 degrees of freedom
Multiple R-squared:  0.5612,    Adjusted R-squared:  0.5594
F-statistic: 303.3 on 8 and 1897 DF,  p-value: < 2.2e-16

```

Table 13: Modelul de regresie în urma algoritmului Boruta

Algoritmul Boruta a selectat următoarele variabile ca fiind importante: Etaj, Sector, NrCamere, Rating, Suprafata, log(TotalEtaje), termenul de interacțiune dintre Rating și Suprafața și termenul de interacțiune între Rating și Etaj. Bonitatea modelului nou este $R^2 = 55.94\%$.

- **Comparația celor 4 modele**

În primul rând, diferența majoră este că modelul pe baza algoritmului Boruta folosește 8 variabile independente, pe când regresiiile Ridge, Lasso și Elastic Net doar 3.

În al doilea rând, se observă o diferență în indicatorul de bonitate R^2 , care are aproximativ o valoare de 35,30% pentru Ridge, Lasso și Elastic Net și 55,95% pentru regresia prin algoritmul Boruta.

Prognoze

Am împărțit setul de date într-un set de antrenare (aprox. 80%) și unul de testare (20% dintre observații).

$RMSE = 0.2748 < 1 \Rightarrow$ predicția este bună

$MAE = 0.2342 < 1 \Rightarrow$ predicția este bună

$MSE = 0.0755 > 0 \Rightarrow$ predicția este bună

$MAPE = 0.0541 < 1 \Rightarrow$ predicția este bună

1	2	3
61.30904	73.12296	113.45393

Table 14: Tabelul prognozei regresiei multiple

Tabelul indică valorile în mii euro a prețurilor imobilelor cu:

- suprafața de 50 m² și care se află la parterul unei clădiri cu 4 etaje
- suprafața de 60 m² și care se află la etajul 2 din 6
- suprafața de 90 m², la etajul 8 al unei clădiri cu 9 etaje

	fit	lwr	upr
1	61.30904	59.95231	62.69647
2	73.12296	71.96641	74.29809
3	113.45393	109.23513	117.83567

Table 15: Tabelul prognozei pe intervale de încredere

Tabelul indică intervalele de încredere rezultate în urma prognozei regresiei multiple a valorilor în mii de euro a prețurilor imobilelor.

Pentru un interval de încredere de 95%, putem estima următoarele valori: 61.3 mii euro, 73.1 mii euro și 113.4 mii euro, pentru aceleași valori ale suprafeței, etajului și numărului total de etaje al clădirii în care se află apartamentul descrise mai sus.

Pentru primul imobil, limita inferioară a intervalului este de 59.9 mii euro, cea superioară de 62.6 mii euro, iar media de preț de 61.3 mii euro.

Pentru cel de-al doilea, limita inferioară a intervalului de încredere este 71.9 mii euro, limita superioară 74.2 mii euro și media de preț de 73.1 mii euro.

Prețul celui de-al treilea imobil a cărui medie de preț este de 113.4 mii euro se încadrează între 109.2 și 117.8 mii euro.

Limitări

Pentru modelul de regresie simplă, dar și la cea multiplă am întâmpinat probleme la ipoteza de normalitate. Am identificat punctele influente cu distanța Cook, însă fiecare corectare a dus la apariția unor noi puncte. În consecință, ipoteza nu a fost acceptată, iar modelul nu a putut fi adus la o formă normală.

Concluzii

Am folosit modele de regresie pentru a obține o perspectivă asupra modului în care anumiți factori determină prețurile imobiliarelor din București.

În urma regresiei simple a reieșit că suprafața unui imobil este un prim factor care determină prețul acestuia, dat fiind faptul că modelul preț-suprafață este cel mai semnificativ din punct de vedere statistic, explicând aproximativ 65.19% din variația prețului.

Prin modelul de regresie multiplă implementat, am constatat că pe lângă factorii adăugați și studiați, anume suprafața imobilului, etajul la care se află și totalul etajelor clădirii, există și alte elemente care au o influență destul de mare asupra prețului, acțiunea acestora reprezentând 69.88% din preț. Modelul nu este unul care să reprezinte într-o mare măsură criteriile ce sunt luate în calcul de cel care pune la vânzare imobiliarul la stabilirea prețului. Pe lângă acestea, probabil că prețul este influențat și de prețul materialelor utilizate la construcție, de numărul de camere, băi, utilități, existența unui garaj sau a unui loc de parcare, sector, etc.

Bibliografie

- CREȚU, F. R. (2019). "IMPORTANȚA TERENULUI ȘI A AMPLASAMENTULUI ÎN EVALUAREA PROPRIETĂȚILOR IMOBILIARE". *Culegere de lucrări științifice ale Conferinței Științific Internațional "Competitivitate și Inovare în economia cunoașterii"*, (pp. 93-101). Retrieved from https://irek.ase.md/xmlui/bitstream/handle/1234567890/193/Cretu-R-F_conf_09.19_teze.pdf?sequence=1&isAllowed=y
- Drd. Elena IONAȘCU, P. u. (2019). "*Transparența piețelor imobiliare: evidențe conceptuale și empirice*". Retrieved from <https://www.proquest.com/openview/087b4a7df966941e9bc5e03727102c05/1?pq-origsite=gscholar&cbl=51385>
- LEȘAN, A. (2020). "FACTORS IDENTIFICATION INFLUENCING COMPENSATION SIZE IN CASE OF PROPERTY EXPROPRIATION FOR PUBLIC UTILITY". *Journal of Social Sciences*, 65-69. Retrieved from https://ibn.idsi.md/sites/default/files/imag_file/JSS-1-2020_65-69.pdf
- ROTARU, O. (2021). "OPORTUNITĂȚI ȘI RISCURI ÎN PROCESUL DECIZIONAL DE CUMPĂRARE ÎN CADRUL PIEȚEI IMOBILIARE". *INTERNATIONAL SCIENTIFIC CONFERENCE*, (pp. 262-269). Retrieved from https://ibn.idsi.md/sites/default/files/imag_file/p-262-269.pdf
- TUREȚCHI, V. (n.d.). "*PIAȚA OBIECTELOR IMOBILIARE RURALE A ȚĂRILOR VECINE CU REPUBLICA MOLDOVA*". Retrieved from https://ibn.idsi.md/sites/default/files/imag_file/435-439.pdf

Anexa

Tabele:

Table 1: Tabelul modelului de regresie liniară unifactorială.....	12
Table 2: Testul Breusch-Pagan pentru regresia simplă	15
Table 3: Testul White pentru regresia simplă.....	16
Table 4: Testul Durbin-Watson pentru regresia simplă.....	17
Table 5: Testul Breusch-Godfrey de ordinul II pentru regresia simplă.....	17
Table 6: Testul Shapiro-Wilk pentru regresia simplă.....	19
Table 7: Tabelul prognozei regresiei simple	20
Table 8: Tabelul modelului de regresie multiplă.....	20
Table 9: Tabelul valorilor VIF.....	22
Table 10: Tabelul regresiei cu variabilă dummy și termeni de interacțiune	25
Table 11: Tabelul modelului de regresie Ridge.....	27
Table 12: Tabelul modelului de regresie Lasso.....	29
Table 13: Modelul de regresie în urma algoritmului Boruta	31
Table 14: Tabelul prognozei regresiei multiple.....	32
Table 15: Tabelul prognozei pe intervale de încredere	32

Figuri:

Figure 1: Graficul scorului imobilelor.....	11
Figure 2: Graficul etajelor imobilelor.....	11
Figure 3: Graficul numărului total de etaje al clădirilor în care se află imobilele.....	11
Figure 4: Dreapta de regresie pentru modelul unifactorial	12
Figure 5: Graficul Residuals vs Fitted pentru regresia simplă	14
Figure 6: Graficul Q-Q plot pentru regresia simplă	15
Figure 7: Graficul ACF pentru regresia simplă.....	16
Figure 8: Graficul homoscedasticității pentru regresia multiplă	23
Figure 9: Graficul Residuals vs Fitted pentru regresia multiplă.....	24
Figure 10: Graficul Cook's Bar Plot în urma corecției	25
Figure 11: Plot regresia Ridge	28
Figure 12: Plot regresia Lasso	29
Figure 13: Plot regresia Elastic Net.....	30