

Uniwersytet Warszawski  
Wydział Nauk Ekonomicznych

**Marta Pleń**

Nr albumu: 442060

**Analiza czynników determinujących ceny mieszkań  
na rynku wtórnym w Polsce.**

Praca przygotowana

w ramach ćwiczeń z Ekonometrii.

Koordinator przedmiotu: dr Kateryna Zabarina

Warszawa, styczeń 2024 r.

# SPIS TREŚCI

|  |    |
|--|----|
| <b>1. Wstęp</b>                                    | 2  |
| <b>2. Przegląd literatury</b>                      | 3  |
| <b>3. Hipotezy</b>                                 | 6  |
| <b>4. Opis zbioru danych</b>                       | 6  |
| <b>5. Analiza statystyczna zmiennych</b>           | 9  |
| 5.1 Analiza zmiennej zależnej                      | 9  |
| 5.2 Analiza zmiennych niezależnych ciągłych        | 11 |
| 5.3 Analiza zmiennych zależnych zero – jedynkowych | 22 |
| <b>6. Estymacja modelu</b>                         | 23 |
| 6.1 Model pierwotny                                | 23 |
| 6.2 Testy diagnostyczne                            | 27 |
| 6.3 Model ostateczny                               | 29 |
| <b>7. Interpretacja wyników</b>                    | 31 |
| <b>8. Weryfikacja hipotez</b>                      | 35 |
| <b>9. Problemy z danymi</b>                        | 37 |
| <b>10. Wnioski</b>                                 | 40 |
| <b>Bibliografia</b>                                | 42 |

## 1. Wstęp

Ceny mieszkań są tematem wieloaspektowym, dającym się badać zarówno z perspektywy makro, jak i mikroekonomicznej. Podczas gdy analizy makroekonomiczne często podkreślają krajowe trendy i skutki polityki, perspektywa mikroekonomiczna skupia się na indywidualnych determinantach kształtujących wartość nieruchomości. W tym raporcie zawęzę swoją uwagę do poziomu mikroekonomicznego, mając na celu zbadanie skomplikowanych relacji pomiędzy różnymi czynnikami a cenami mieszkań na rynku wtórnym.

W kontekście rosnącej w ostatnich latach luki podażowej mieszkań w Polsce, szczególnie napędzanej rządowymi programami i czynnikami zewnętrznymi np. pandemią COVID-19 czy wojną w Ukrainie, dynamika rynku nieruchomości zajmuje istotne miejsce w dyskursie krajowym.<sup>1</sup> Gwałtowny wzrost popytu spowodował wyraźną tendencję wzrostową cen mieszkań, co skłoniło konsumentów do zmierzenia się z wyzwaniem dostosowania swoich preferencji mieszkaniowych do realiów gospodarczych. W miarę jak w Polsce zachodzą zmiany sytuacji na rynku nieruchomości oraz warunków gospodarczych, zrozumienie czynników wpływających na ceny mieszkań staje się coraz większą koniecznością zarówno dla właścicieli domów, inwestorów, jak i badaczy.

Wielu aspirujących właścicieli nieruchomości znajduje się w sytuacji, w której marzenie o posiadaniu idealnego mieszkania często krzyżuje się z ograniczeniami związanymi z przystępną ceną. W odpowiedzi na rosnące ceny konsumenci zmuszeni są ponownie skalibrować swoje preferencje mieszkaniowe, decydując się na bardziej pragmatyczne wybory, które odpowiadają ich ograniczeniom budżetowym.

Celem niniejszego raportu jest próba rozkwikłania zawilej sieci czynników wpływających na ceny mieszkań na rynku wtórnym w największych miastach w Polsce. Badanie ma na celu wyjaśnienie, w jaki sposób charakterystyka nieruchomości, udogodnienia i aspekty lokalizacyjne wpływają na różnice w ofertach. Wykorzystując klasyczny model regresji liniowej, postaram się określić konkretne czynniki, które mogą mieć istotny wpływ na ceny mieszkań.

---

<sup>1</sup> Raport PwC, 2021

## 2. Przegląd literatury

Dynamika rynków nieruchomości mieszkalnych od dawna jest przedmiotem żywego zainteresowania i analizy. Literatura dotycząca cen mieszkań jest obszerna i różnorodna, co świadczy o złożoności tego zjawiska. Często metodą wykorzystywaną w tego typu badaniach są hedoniczne modele cenowe, które oparte są głównie na teorii zachowań konsumentów Lancastera. Stwierdził on, że to nie samo dobro kreuje jego użyteczność, ale indywidualne cechy, które je charakteryzują.<sup>2</sup> Ponadto „mieszkanie” jest wielowymiarowym i heterogenicznym dobrem<sup>3</sup>, a jego analizę można podzielić na kilka części: charakterystykę samego lokalu oraz charakterystykę lokalizacji, a w tym okolicy/sąsiedztwa i dostępności. Przy czym ze względu na nieruchomość mieszkań, ta druga kwestia jest szczególnie istotna.<sup>4</sup>

W swoim przeglądzie modeli hedonicznych dotyczących rynku mieszkaniowego, Anthony Owusu-Ansah omawia zastosowanie modeli parametrycznych m.in. MNK i WMNK, gdzie krzywa regresji hedonicznej pokazuje związek między zmienną zależną (w tym przypadku ceną domu) a zmienną niezależną lub objaśniającą (cechą mieszkania, np. liczbą sypialni). Analizuje on także inne modele nieparametryczne i półparametryczne. Szeroko wykorzystywana w hedonicznej analizie cen jest logarytmiczna postać zmiennej objaśnianej, co podkreślano w licznych badaniach. Szczególnie dlatego, iż ułatwia ona interpretację współczynników regresji oraz pomaga zminimalizować problem heteroskedastyczności reszt.<sup>5</sup>

Dirk Wittowsky, Josje Hoekveld, Janina Welsch i Michael Steier podjęli się szerokiej analizy cen mieszkań w Dortmundzie, biorąc pod uwagę ich cechy, dostępność do różnych punktów i dóbr publicznych oraz wpływ sąsiedztwa. W tym celu wykorzystali zwykłą metodę najmniejszych kwadratów oraz modele przestrzenne (*spatial lag models*), które okazały się bardziej efektywne, gdyż uwzględniały oddziaływanie cen mieszkań sąsiednich, a co za tym idzie zapobiegały przeszacowaniu wpływu pozostałych zmiennych.<sup>6</sup>

Jeśli chodzi o charakterystykę mieszkań, to ich stan wykazał stosunkowo wysokie współczynniki na tle pozostałych zmiennych niezależnych. Istotna była także liczba pokoi, choć relacja ta była ujemna: im więcej pokoi, tym niższa cena – co może być dość

---

<sup>2</sup> Lancaster, 1966

<sup>3</sup> Bourne, 1981

<sup>4</sup> Wittowsky et al., 2020

<sup>5</sup> Anthony Owusu-Ansah, 2011

<sup>6</sup> Wittowsky et al., 2020

nieoczekiwanym rezultatem. Wielkość powierzchni mieszkalnej była dodatnio skorelowana z ceną mieszkania, ale okazała się istotna jedynie w przypadku mieszkań własnościowych.

Z badań wynika również, że ważne są usługi dostępne w niewielkiej odległości od mieszkania. Jednak tylko dwa udogodnienia pozytywnie korelują z cenami mieszkań: restauracje i parki. Zgodnie z przewidywaniami autorów, najważniejsze czynniki są bezpośrednio związane ze stanem mieszkań, niezależnie od ich typu.<sup>7</sup>

Analiza rynku mieszkaniowego w Turynie, zaprezentowana przez Luca D'Acci, pokazuje, jak w zależności od obszaru miasta, wartość domu wzrasta/maleje, nawet wśród terenów zlokalizowanych dość blisko siebie. Wyniki badania wskazują m.in., że wartość nieruchomości spada o 0,23% na każdy 1% wzrostu odległości od centrum miasta. Koszty pieniężne zakupu mieszkania, czas i transport oraz korzyści jakościowe wynikające z charakterystyki danego miejsca odgrywają główną rolę w procesach decyzyjnych gospodarstw domowych przy wyborze mieszkań.<sup>8</sup>

Do podobnych wniosków dochodzą również Emilia Tomczyk i Marta Widłak w swojej pracy na podstawie danych na temat transakcji zawieranych na warszawskim rynku wtórnym, w której poświęcają uwagę konstrukcji różnych modeli hedonicznego indeksu cen. Z modelu oszacowanego metodą imputacji<sup>9</sup> wynika, że lokalizacja mieszkania w dobrej dzielnicy Warszawy podnosi cenę metra kwadratowego o około 29% w porównaniu z mieszkaniem o przeciętnej lokalizacji.

Autorki wnioskuje także, że mieszkania bardzo duże są droższe od mieszkań małych i średnich, co pokazuje, że powierzchnia ma istotny wpływ na cenę nieruchomości. Ponadto mieszkania o wysokim standardzie wyposażenia są droższe od mieszkań o przeciętnym standardzie średnio o 9%, natomiast niski standard zaniża cenę całkowitą o około 6% w stosunku do standardu przeciętnego.<sup>10</sup>

Alastair Adair , Stanley McGreal , Austin Smyth , James Cooper i Tim Ryley wskazują jednak, że dostępność do „punktów centralnych” takich jak Central Business District (CBD) ma niewielkie znaczenie w wyjaśnianiu zmian cen domów w skali całego miasta. Jedynie na poziomie subrynkowym, szczególnie na obszarach o niższych dochodach, dostępność może

---

<sup>7</sup> Wittowsky et al., 2020

<sup>8</sup> Luca D'Acci, 2018

<sup>9</sup> Imputacja – sztuczne wstawienie pewnych wartości do tabeli danych. Na ogół imputacja jest wykonywana w celu usunięcia tzw. braków danych.

<sup>10</sup> Tomczyk, Widłak 2010

mieć istotny wpływ. W artykule skupiono się na czynnikach wpływających na strukturę cen nieruchomości mieszkalnych na obszarze miejskim Belfastu, badając cechy nieruchomości, czynniki społeczno-ekonomiczne i wpływ dostępności. Analiza podkreśla znaczenie badań na poziomie subryнку i wyciąga wnioski dotyczące złożoności relacji w obrębie obszaru miejskiego.<sup>11</sup>

Badanie Katarzyny Małeckiej dla wtórnego rynku mieszkaniowego w Łodzi potwierdziło związek cen mieszkań z ich lokalizacją oraz powierzchnią. Dzielnicą, w której mieszkania uzyskały cenę powyżej średniej ceny w Łodzi było m.in. Śródmieście. Ponadto autorka stwierdza w swoim artykule, że im bardziej rozwinięty dany rynek nieruchomości, tym więcej atrybutów ma wpływ na kształtowanie się cen na tym rynku.<sup>12</sup>

Jak pokazuje dotychczasowa literatura, dostęp do instytucji publicznych, np. szkół czy przedszkoli ma znaczący wpływ na ceny nieruchomości. Haizhen Wen, Yue Xiao i Ling Zhang wykorzystali dane na temat mieszkań i placówek oświatowych w Hangzhou w Chinach do opracowania hedonicznych i przestrzennych modeli cen w celu ilościowej oceny wpływu placówek oświatowych na ceny mieszkań.

Placówki edukacyjne mają pozytywny wpływ na cenę mieszkań, a wyniki badania ukazują znaczenie placówek edukacyjnych na rynku mieszkaniowym w Chinach. Cena wzrasta średnio o 2,737% lub 0,904%, gdy dom znajduje się w odległości mniejszej niż 1 km od szkoły średniej lub uczelni. Jednakże różne rodzaje placówek edukacyjnych mają inny poziom oddziaływania na cenę mieszkań, a mieszkańcy są skłonni płacić więcej za dostęp do wysokiej jakości placówek edukacyjnych. Bliskość okolicznych przedszkoli, liceów i uniwersytetów zwiększa ceny domów. Autorzy wskazują jednak, że tradycyjny model hedoniczny przeszacowuje dodatni efekt placówek oświatowych, a wyeliminować ten problem pomaga zastosowanie przestrzennych modeli ekonometrycznych.<sup>13</sup>

W artykule autorstwa Joseph T.L. Ooi, Thao T.T. Le i Nai-Jia Lee zbadano wpływ jakości wykonania i konstrukcji nowych mieszkań na cenę ich sprzedaży i wzrost wartości. Do pomiaru jakości budownictwa w Singapurze wykorzystano System Oceny Jakości Budownictwa (CONQUAS). Znalezione istotne dowody na to, że cena sprzedaży i stopa wzrostu są w istotny sposób powiązane z jakością konstrukcji nowych domów. Co szczególnie ciekawe, premia

---

<sup>11</sup> Adair et al., 2000

<sup>12</sup> Małeczka, 2009

<sup>13</sup> Wen et al., 2017

„jakościowa” występuje zarówno na rynku pierwotnym, jak i wtórnym, a wpływ jakości na rynku odsprzedaży jest prawie dwukrotnie większy niż na rynku przedsprzedaży. Atorzy stwierdzają, że nabywcy, którzy zapłacili wysoką cenę za domy dobrej jakości, mogą przynajmniej odzyskać premię za jakość konstrukcji, gdy odsprzedają domy na rynku wtórnym.<sup>14</sup>

### 3. Hipotezy

W celu przeprowadzenia analizy czynników mogących mieć wpływ na ceny mieszkań w polskich miastach, zdecydowałam się na postawienie przedstawionych poniżej hipotez. Bazują one na dotychczasowej literaturze oraz własnych przypuszczeniach.

**H1:** Powierzchnia mieszkania jest istotna i dodatnio skorelowana z ceną – mieszkania duże są średnio droższe od mniejszych.

**H2:** Liczba pokoi jest istotna i ujemnie skorelowana z ceną – mieszkania z dużą liczbą pokoi są średnio tańsze od tych z mniejszą liczbą.

**H3:** Stan/kondycja mieszkania ma istotny, dodatni wpływ na jego cenę.

**H4:** Ceny mieszkań różnią się w zależności od dostępności do istotnych dla mieszkańców punktów, w szczególności placówek edukacyjnych i restauracji.

**H5:** Własność mieszkania ma istotny wpływ na jego cenę.

**H6:** Mieszkania wykonane z lepszego jakościowo materiału (cegła) są średnio droższe od tych zbudowanych z innych materiałów.

**H7:** Mieszkania na najniższym piętrze są najdroższe, a mieszkania na najwyższym piętrze są najtańsze.

**H8:** Obecność udogodnień takich jak miejsce parkingowe, winda, balkon, ochrona czy schowek ma istotny wpływ na cenę i jest z nią dodatnio skorelowana.

**H9:** Wiek mieszkania jest ujemnie skorelowany z ceną – im starsze mieszkanie, tym średnio tańsze.

**H10:** Mieszkania w stolicy są średnio droższe niż mieszkania w innych miastach.

---

<sup>14</sup> T.L. Ooi et al., 2014

## 4. Opis zbioru danych

Wykorzystana w pracy baza danych pochodzi z platformy Kaggle.com i została utworzona przez Krzysztofa Jamroza na podstawie internetowych ofert sprzedaży mieszkań z 15 największych miast w Polsce (Warszawa, Łódź, Kraków, Wrocław, Poznań, Gdańsk, Szczecin, Bydgoszcz, Lublin, Katowice, Białystok, Częstochowa). Dodatkowo uwzględnia ona także dane pochodzące z Open Street Map, które dotyczą odległości danej nieruchomości od tzw. *points of interest*. Zebrane obserwacje pochodzą z okresu od sierpnia do grudnia 2023 roku.

Na potrzeby utworzenia modelu ekonometrycznego baza poddana została oczyszczeniu z brakujących lub niepotrzebnych wartości oraz przekształceniu niektórych zmiennych. Ostateczna próba użyta do estymacji modelu objęła **2094 obserwacji**. Baza danych zawiera zmienne typu ciągłego jak i zmienne zero-jedynkowe (Tabela 1.).

| NAZWA ZMIENNEJ                | OPIS ZMIENNEJ   | TYP ZMIENNEJ                       |
|-------------------------------|---|------------------------------------|
| id                            | Numer identyfikacyjny danego mieszkania.  | -                                  |
| price                         | Cena w złotych.   | Ciągła                             |
| squareMeters                  | Powierzchnia w metrach kwadratowych.  | Ciągła                             |
| rooms: rooms1, rooms2, rooms3 | Liczba pokoi podzielona na trzy poziomy.<br>rooms1: 1–2 pokoje<br>rooms2: 3–4 pokoje<br>rooms3: 5–6 pokoi | Zero – jedynkowa (dummy variables) |
| centreDistance                | Odległość od centrum miasta w metrach.  | Ciągła                             |
| schoolDistance                | Odległość od najbliższej szkoły w metrach.  | Ciągła                             |
| clinicDistance                | Odległość od najbliższej przychodni w metrach.  | Ciągła                             |
| kindergartenDistance          | Odległość od najbliższego przedszkola w metrach.  | Ciągła                             |
| restaurantDistance            | Odległość od najbliższej restauracji w metrach.   | Ciągła                             |
| collegeDistance               | Odległość od najbliższej szkoły wyższej w metrach.  | Ciągła                             |
| pharmacyDistance              | Odległość od najbliższej apteki w metrach.  | Ciągła                             |



| NAZWA ZMIENNEJ            | OPIS ZMIENNEJ  | TYP ZMIENNEJ     |
|---------------------------|--|------------------|
| ownership                 | Własność mieszkania:<br>1 - Własnościowe<br>0 - Spółdzielcze | Zero - jedynkowa |
| condition                 | Stan mieszkania:<br>1 – wysoki<br>0 - niski                  | Zero - jedynkowa |
| buildingMaterial          | Materiał budowlany:<br>1 – cegła<br>0 – inne                 | Zero - jedynkowa |
| hasParkingSpace           | Miejsce parkingowe:<br>1 – posiada<br>0 – nie posiada        | Zero - jedynkowa |
| hasBalcony                | Balkon:<br>1 – posiada<br>0 – nie posiada                    | Zero - jedynkowa |
| hasElevator               | Winda:<br>1 – posiada<br>0 – nie posiada                     | Zero - jedynkowa |
| hasSecurity               | Ochrona:<br>1 – posiada<br>0 – nie posiada                   | Zero - jedynkowa |
| hasStorageRoom            | Schowek:<br>1 – posiada<br>0 – nie posiada                   | Zero - jedynkowa |
| first_floor               | Najniższe piętro:<br>1 – tak<br>0 – inne piętro              | Zero - jedynkowa |
| top_floor                 | Najwyższe piętro:<br>1 – tak<br>0 – inne piętro              | Zero - jedynkowa |
| age                       | Wiek mieszkania w latach.                                    | Ciągła           |
| Building type: block      | Mieszkanie w bloku:<br>1 – tak<br>0 - nie                    | Zero - jedynkowa |
| Building type: apartmentB | Mieszkanie<br>w apartamentowcu:<br>1 – tak<br>0 - nie        | Zero - jedynkowa |
| Building type: tenement   | Mieszkanie w kamienicy:<br>1 – tak<br>0 - nie                | Zero - jedynkowa |
| stolica                   | Mieszkanie w Warszawie:<br>1 – tak<br>0 - nie                | Zero - jedynkowa |

**Tabela 1.** Wykaz zmiennych w bazie danych. Opracowanie własne.

## 5. Analiza statystyczna zmiennych

### 5.1 Zmienna zależna

Zmienną zależną w badaniu jest zmienna *price*, czyli cena mieszkania podana w złotych. Maksymalna cena mieszkania w bazie wyniosła 3,2 mln zł, a minimalna 187 tysięcy złotych. Średnia cena mieszkań uplasowała się na poziomie 832104 zł. Ponadto 25% nieruchomości kosztowało mniej niż 536500 złotych, a 75% mniej niż 999000 zł (Tabela 2.).

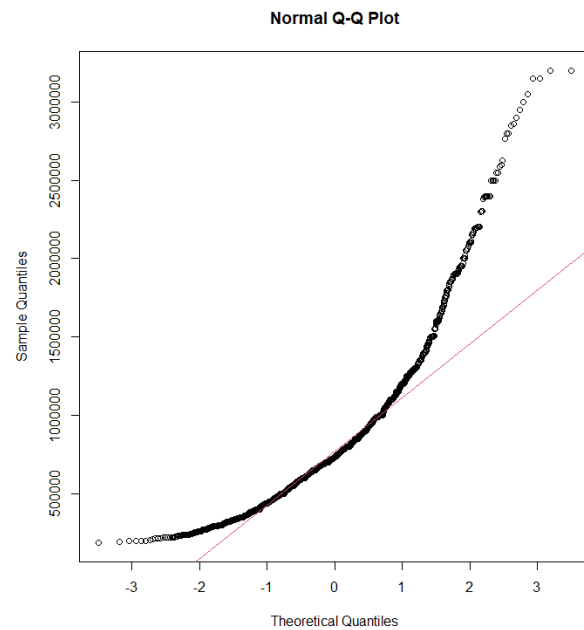
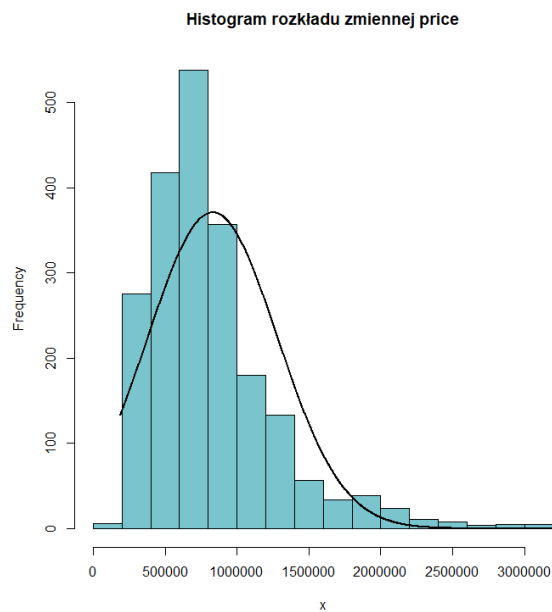
| Minimum   | 1st Qu.   | Median    | Mean      | 3rd Qu.   | Maximum    |
|-----------|-----------|-----------|-----------|-----------|------------|
| 187000.00 | 536500.00 | 730000.00 | 832104.00 | 999000.00 | 3200000.00 |

**Tabela 2.** Podstawowe statystyki dla zmiennej *price*. Opracowanie własne.

Rozkład zmiennej *price* znacząco odbiega od rozkładu normalnego, co potwierdza histogram z linią rozkładu normalnego oraz wykres Normal Q-Q Plot. Można wyraźnie zauważyć asymetrię prawostronną rozkładu. Wynik przeprowadzonego testu Jarque – Bera również pozwala na odrzucenie hipotezy zerowej o normalności rozkładu zmiennej *price* ( $p\text{-value} < 5\%$ ).

Na podstawie literatury, która wskazuje na użycie w podobnych modelach ceny mieszkań w postaci zlogarytmowanej oraz własnej analizy zmiennej *price*, zdecydowałam się na uwzględnienie w swoim badaniu tejże zmiennej w logarytmie:  $\log(\text{price})$ . Poprawiło to znacząco histogram rozkładu – jest on dużo bardziej zbliżony do linii rozkładu normalnego, a także wykres Normal Q-Q plot – kwantyle z próbki mniej odbiegają od linii teoretycznego rozkładu.

## Rozkład zmiennej *price*



### Jarque - Bera Normality Test

Test Results:

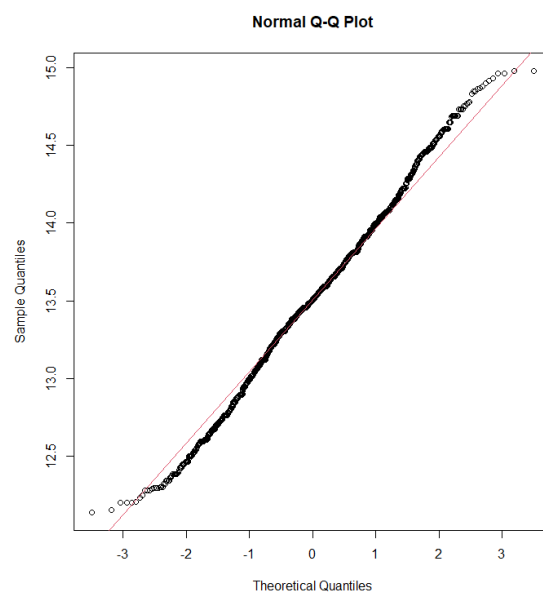
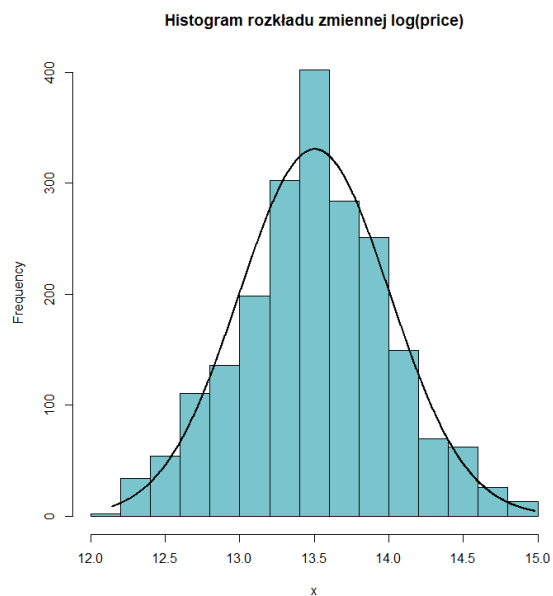
STATISTIC:

X-squared: 2440.9719

P VALUE:

Asymptotic p Value: < 0.00000000000000022

## Rozkład zmiennej *log(price)*



## 5.2 Zmienne niezależne ciągłe

Pierwszą analizowaną ciągłą zmienną objaśniającą jest powierzchnia mieszkania w metrach kwadratowych, czyli zmienna *squareMeters*. Najmniejsze mieszkanie w bazie miało zaledwie 25 m<sup>2</sup>, natomiast największe charakteryzowało się metrażem aż 150 m<sup>2</sup>. Średnia powierzchnia lokalu wyniosła 61,44 m<sup>2</sup>, przy czym jednocześnie nawet 75% mieszkań było mniejszych niż 73 m<sup>2</sup> (Tabela 3.)

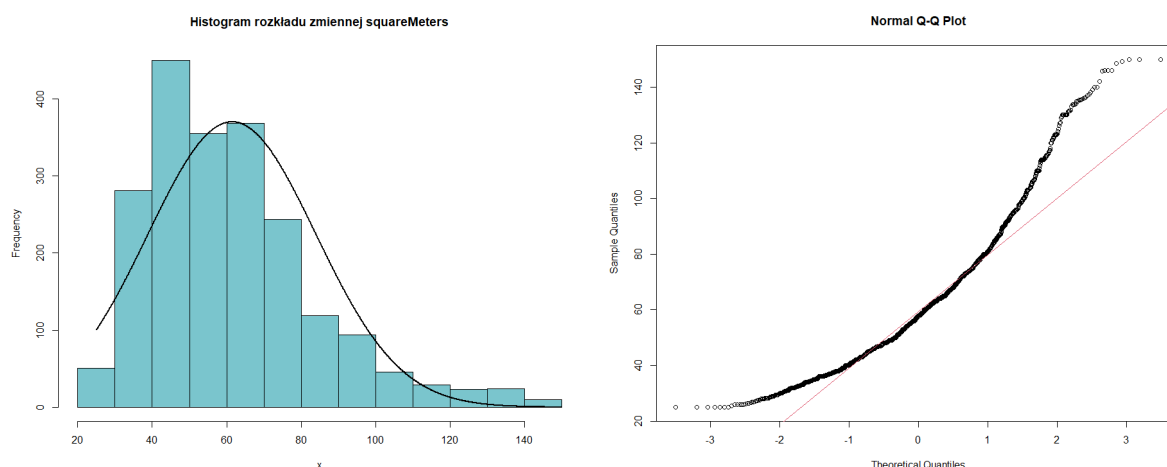
| Minimum | 1st Qu. | Median | Mean  | 3rd Qu. | Maximum |
|---------|---------|--------|-------|---------|---------|
| 25.00   | 45.51   | 57.22  | 61.44 | 73.00   | 150.00  |

**Tabela 3.** Podstawowe statystyki dla zmiennej *squareMeters*. Opracowanie własne.

Powierzchnia mieszkań ma, jak wskazuje histogram oraz wykres Normal Q-Q plot, rozkład asymetryczny prawostronny. Również potwierdza to wynik testu Jarque – Bera, którego p-value było bliskie zera i przy założonym poziomie istotności 5% pozwala na odrzucenie hipotezy o normalności rozkładu zmiennej *squareMeters*.

Analogicznie do przypadku zmiennej zależnej, a także na podstawie wykresów rozproszon zmiennej *log(price)* od zmiennej *squareMeters* i zmiennej *log(price)* od *log(squareMeters)*, zdecydowałam się na uwzględnienie w modelu powierzchni mieszkania w postaci zlogarytmowanej. Dzięki temu rozkład zmiennej uległ znacznej poprawie, tak samo jak wykres rozproszenia, co może mieć pozytywny wpływ na poprawność formy funkcyjnej modelu.

### Rozkład zmiennej *squareMeters*:



## Jarque - Bera Normality Test

Test Results:

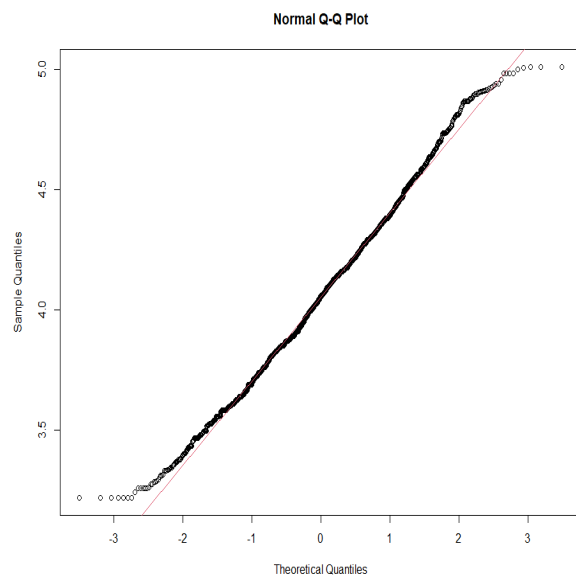
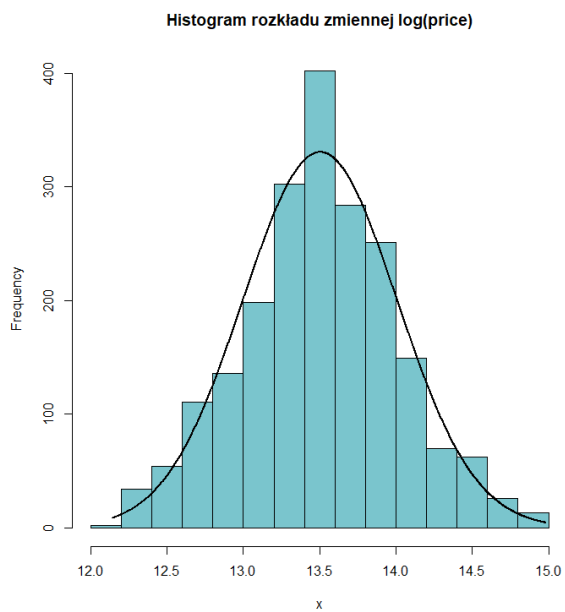
STATISTIC:

X-squared: 666.6859

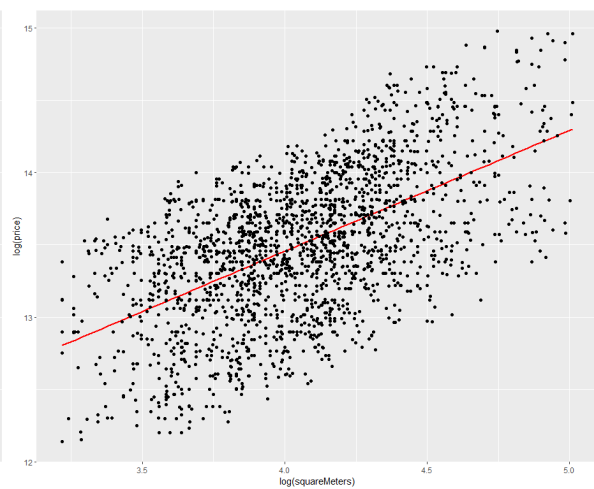
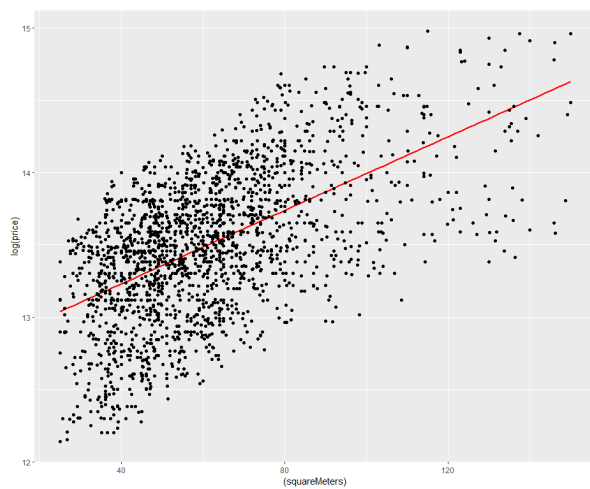
P VALUE:

Asymptotic p Value: < 0.00000000000000022

## Rozkład zmiennej $\log(\text{squareMeters})$ :



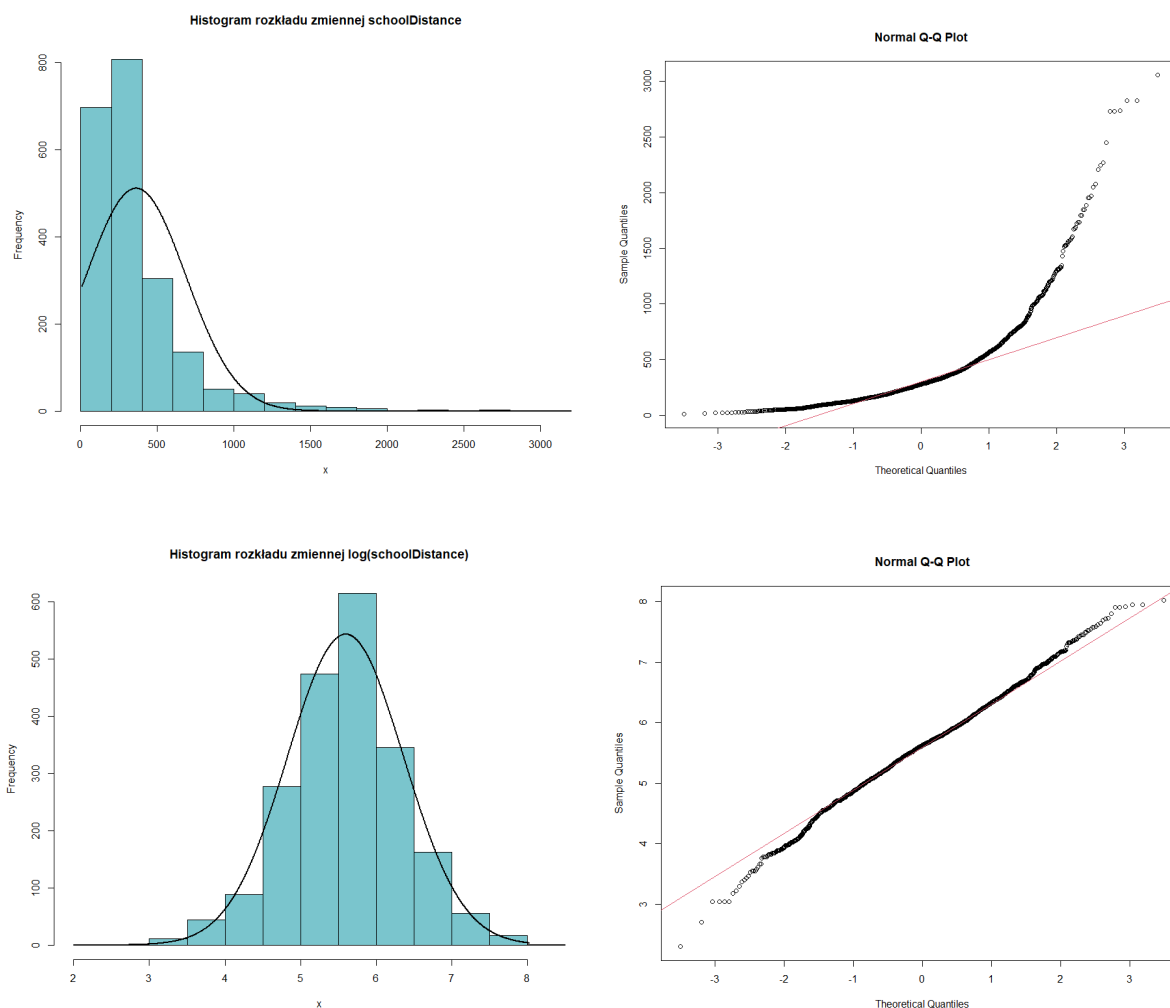
## Wykresy rozproszenia:



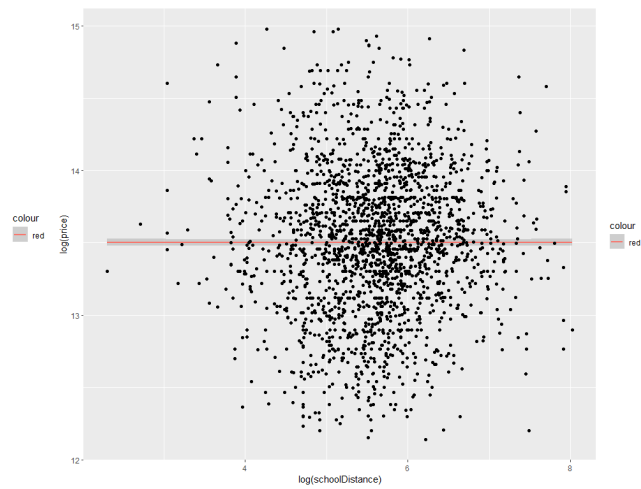
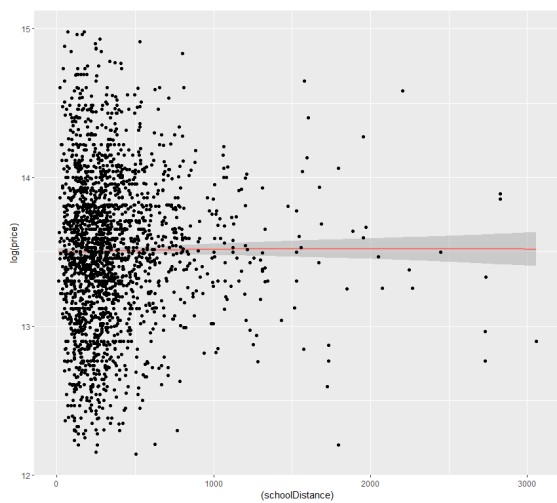
Uwzględnione w analizie zmienne opisujące odległość mieszkania od *points of interest*, czyli istotnych dla mieszkańców miasta punktów (takich jak np. centrum, placówki edukacyjne, instytucje publiczne czy restauracje) również mają charakter ciągły i są wyrażone w metrach.

Każda z tych zmiennych poddana została analizie rozkładu, na podstawie histogramów rozkładu, wykresów Normal Q-Q plot oraz wykresów rozproszenia ze zmienną  $\log(\text{price})$ , a także testom Jarque-Bera na normalność rozkładu. Wszystkie te zmienne zdecydowałam się uwzględnić w modelu w postaci zlogarytmowanej, co w większości przypadków znacząco zbliżyło ich rozkład do rozkładu normalnego i poprawiło wykres rozproszenia. Wyjątkami okazały się zmienne *centreDistance* i *postOfficeDistance*, gdzie zlogarytmowanie nie zrobiło znaczącej różnicy, jednak ułatwiło późniejszą interpretację wyników estymacji.

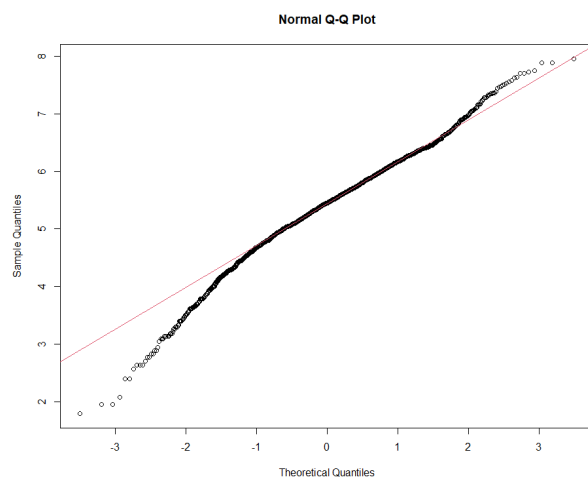
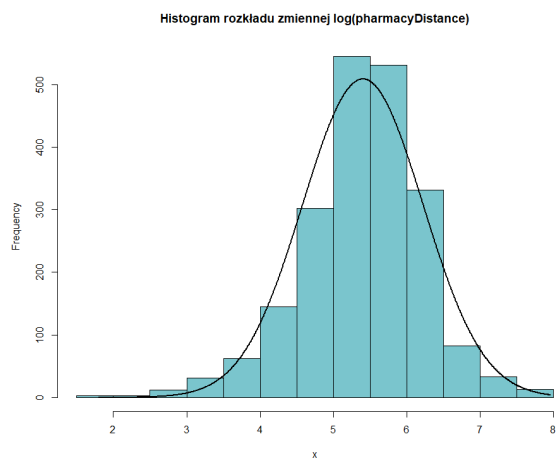
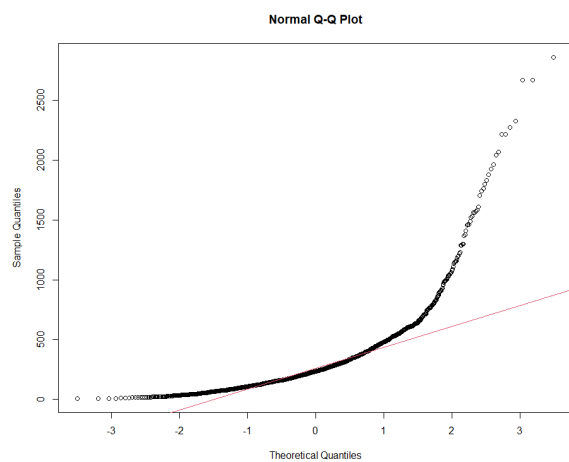
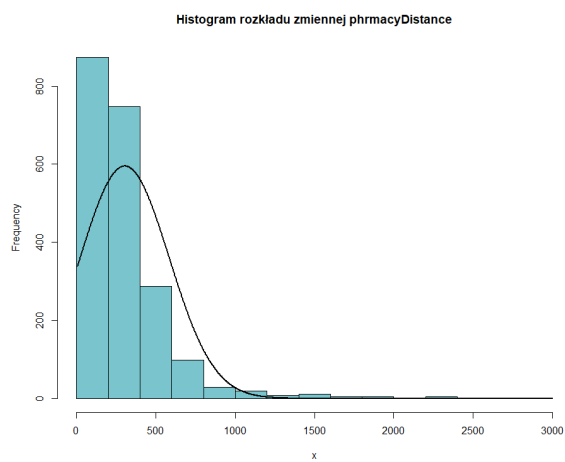
### Rozkład zmiennej *schoolDistance* i $\log(\text{schoolDistance})$ :



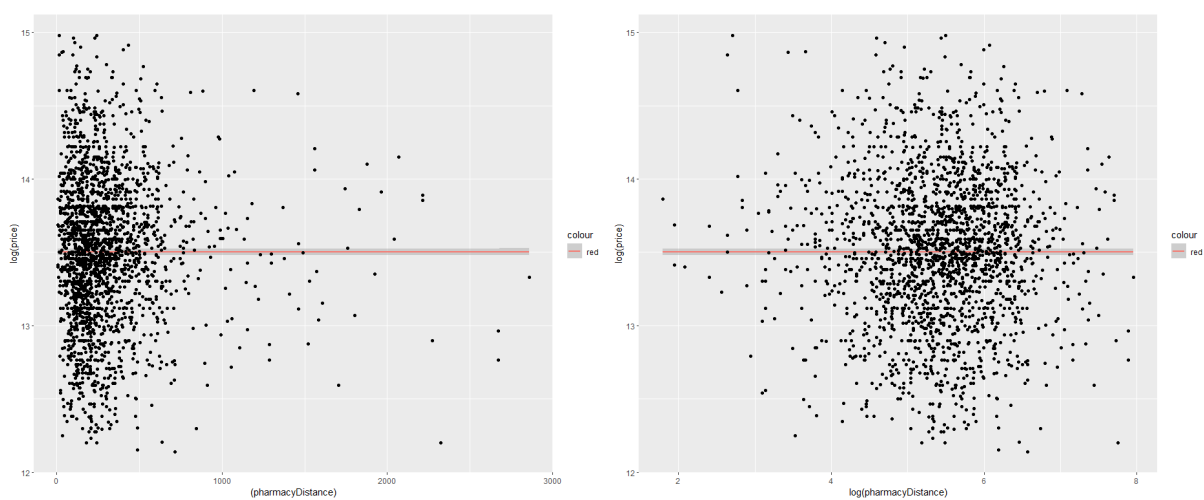
## Wykresy rozproszenia:



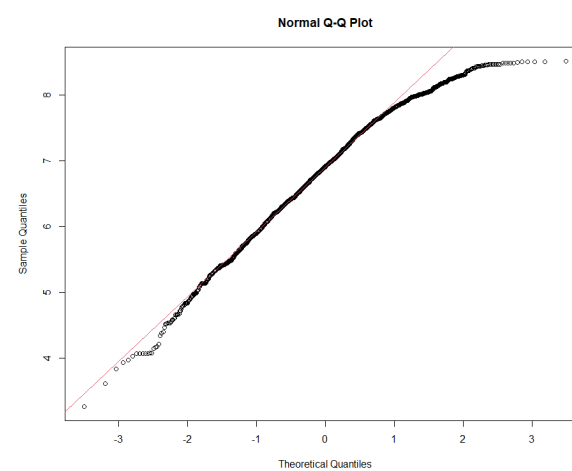
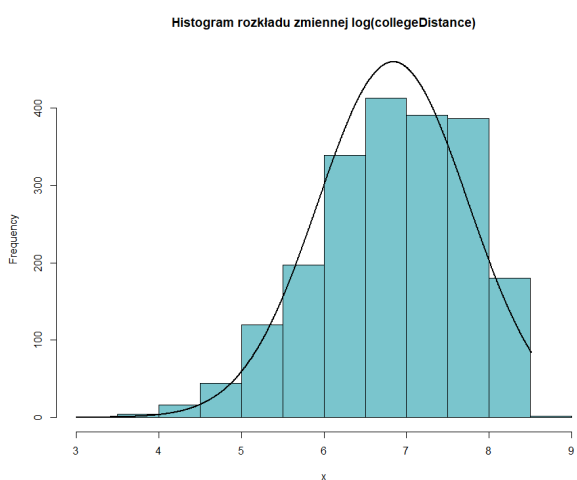
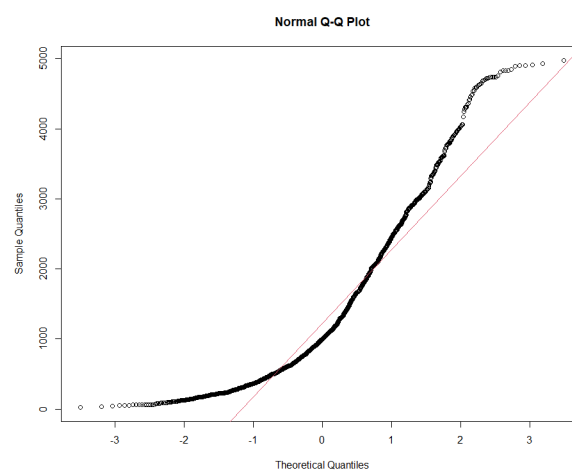
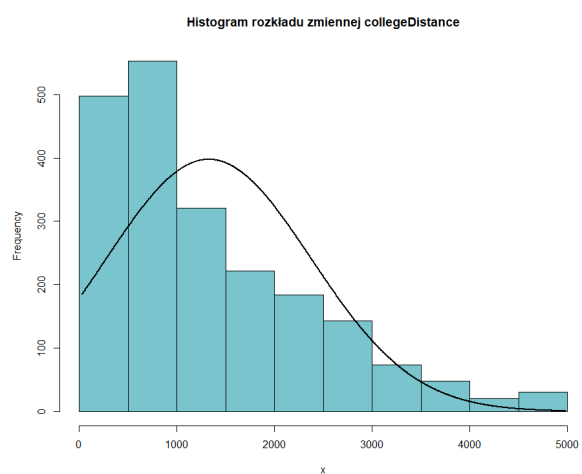
## Rozkład zmiennej *pharmacyDistance* i $\log(\text{pharmacyDistance})$ :



## Wykresy rozproszenia:

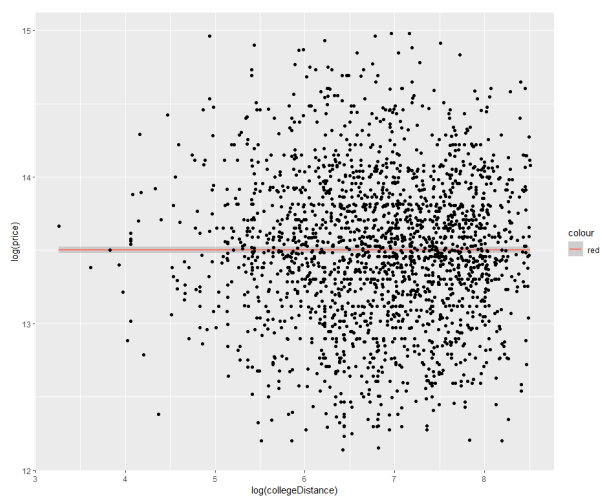
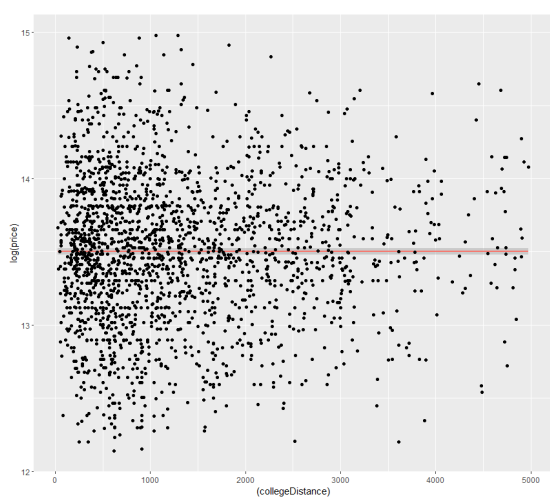


## Rozkład zmiennej *collegeDistance* i zmiennej $\log(\text{collegeDistance})$ :

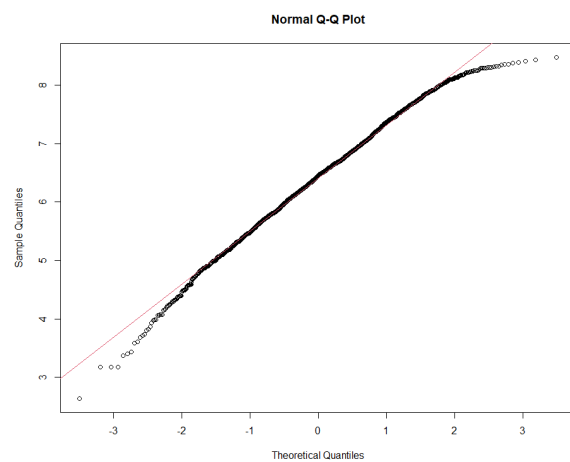
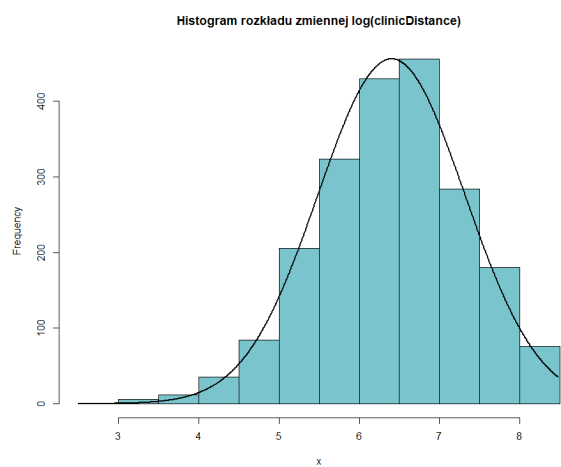
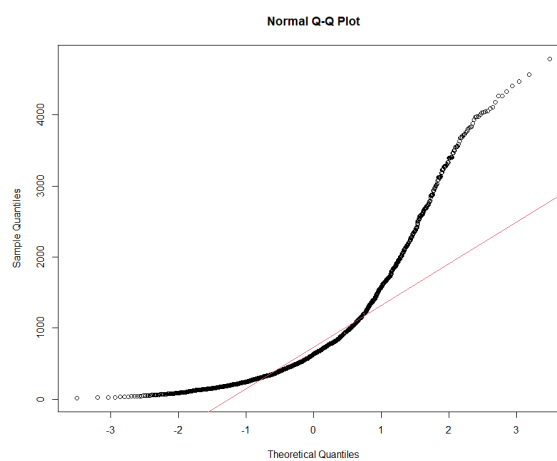
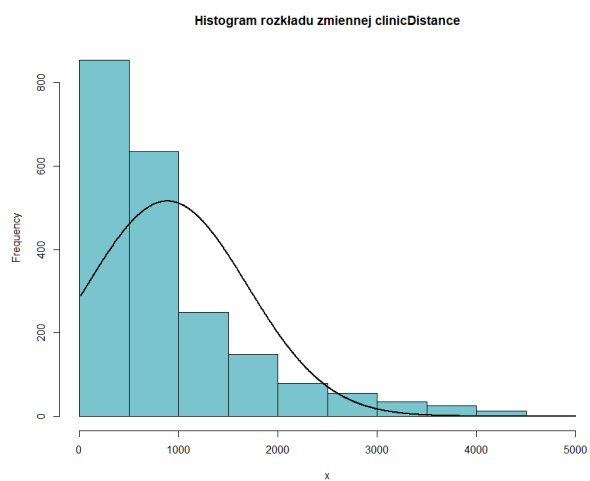




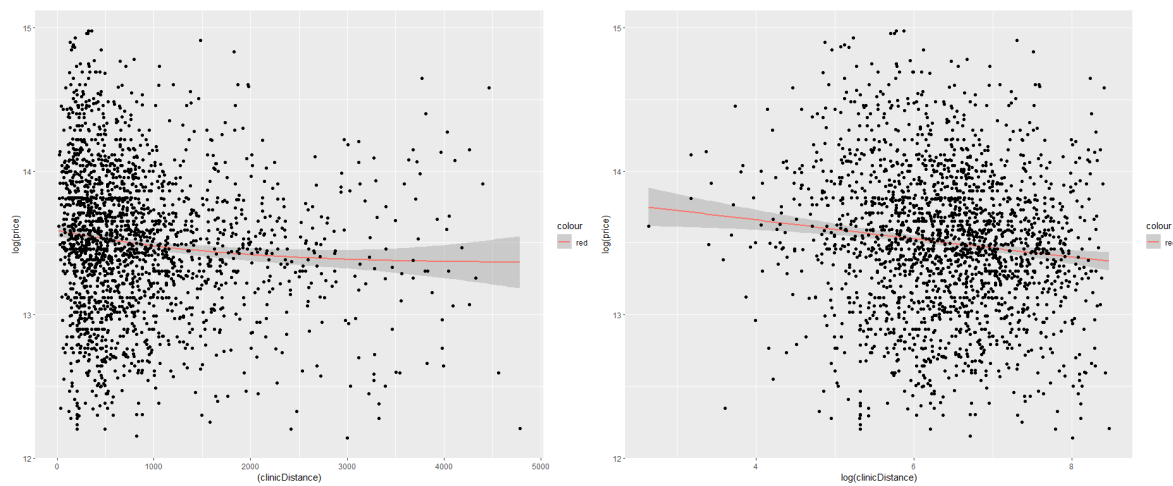
## Wykresy rozproszenia:



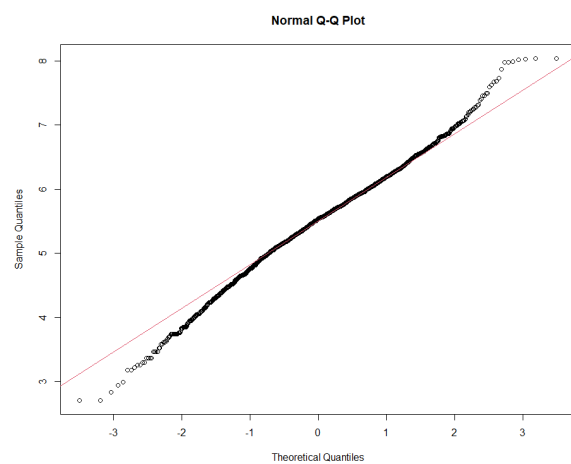
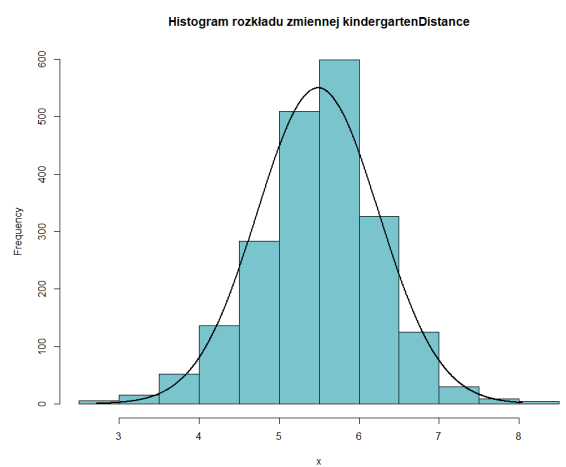
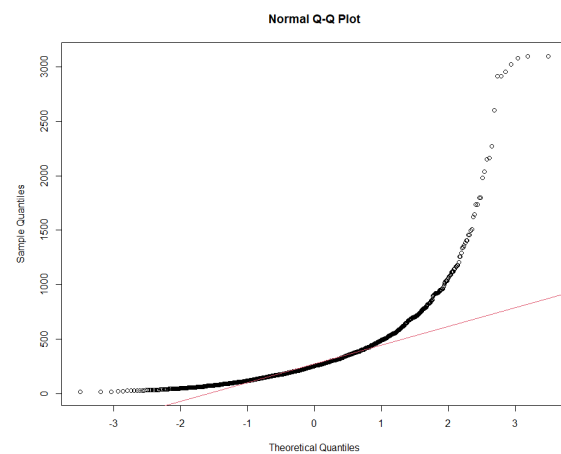
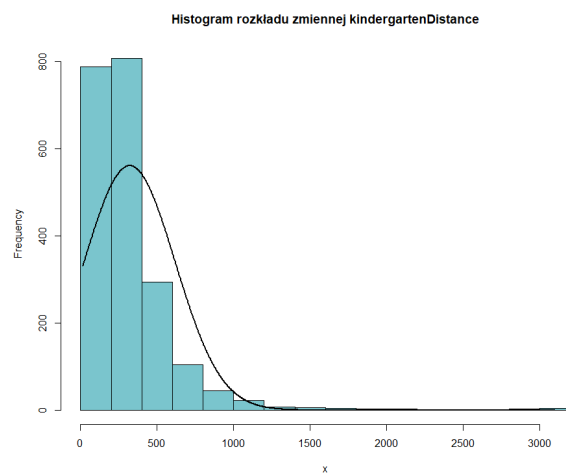
## Rozkład zmiennej *clinicDistance* i zmiennej *log(clinicDistance)*:



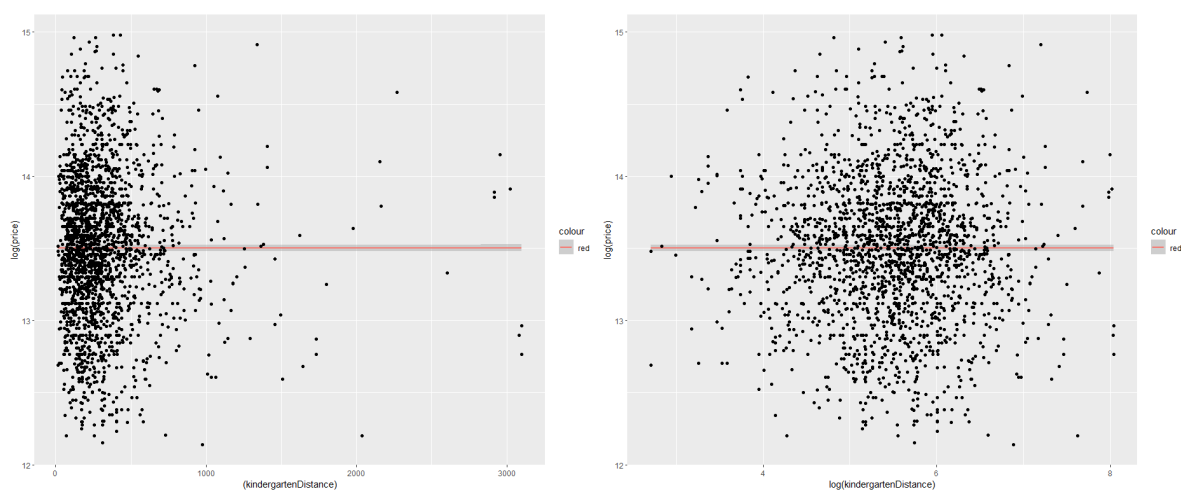
## Wykresy rozproszenia:



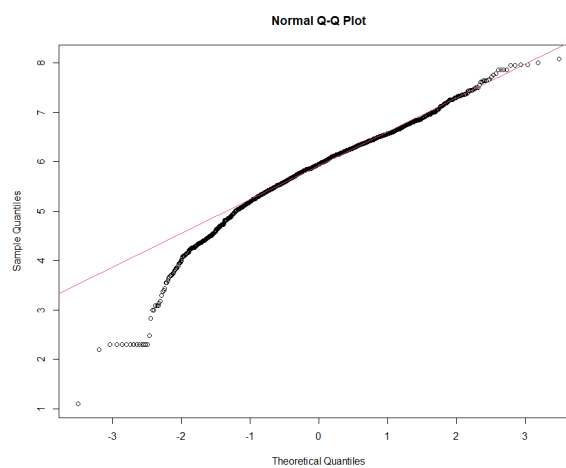
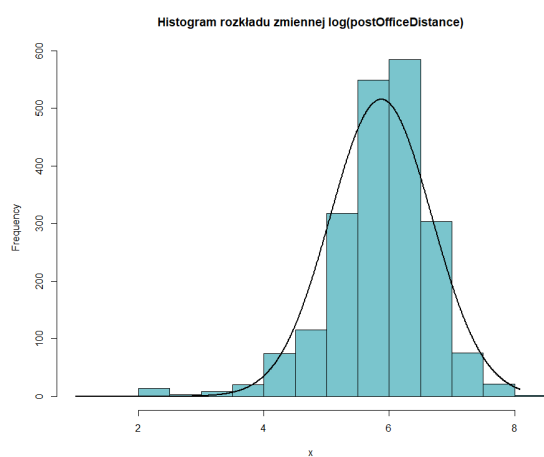
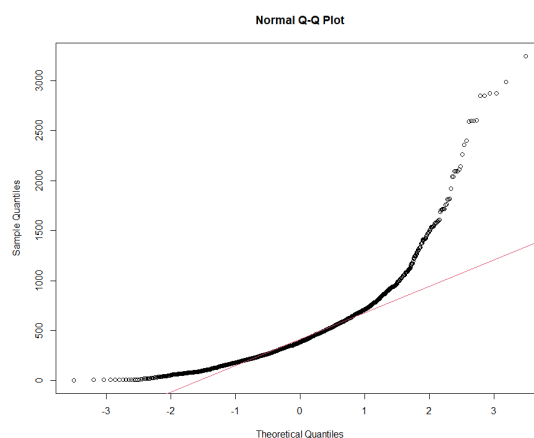
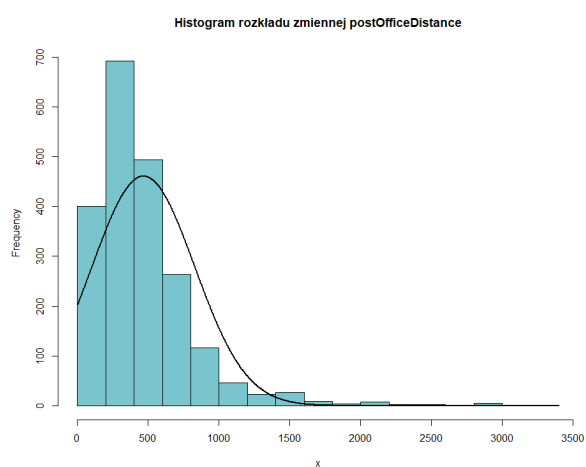
## Rozkład zmiennej *kindergartenDistance* i zmiennej *log(kindergartenDistance)*:



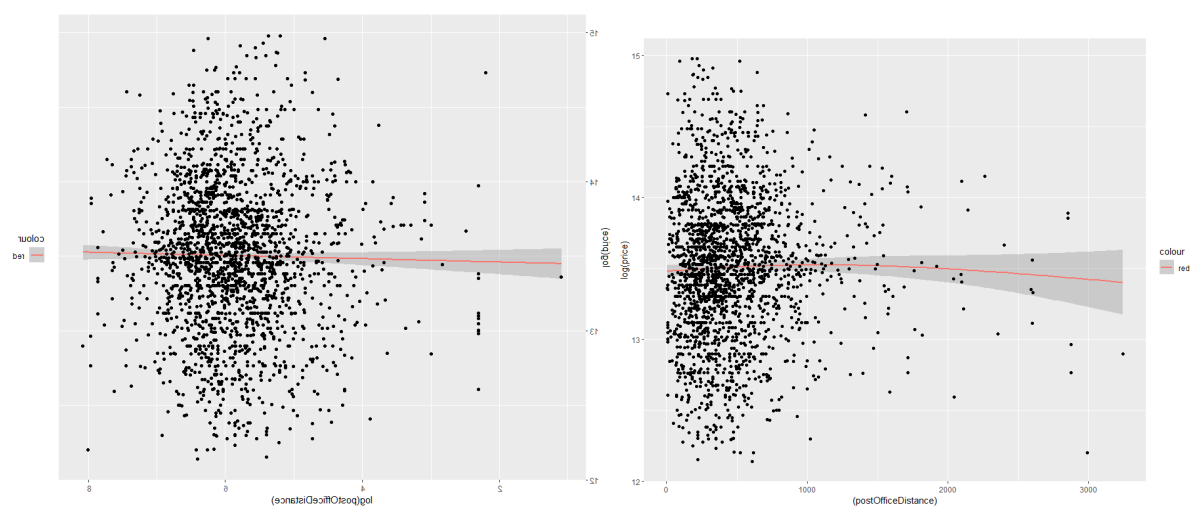
## Wykresy rozproszenia:



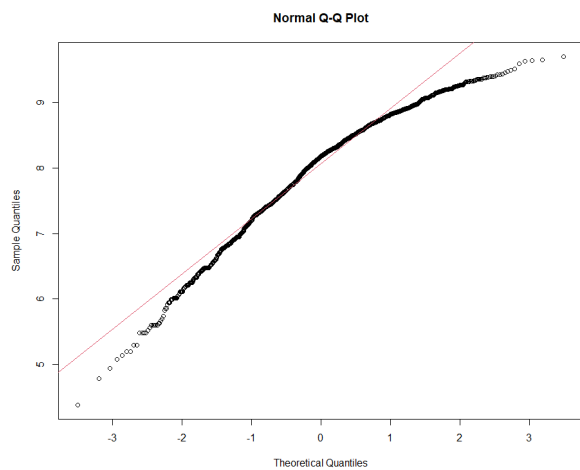
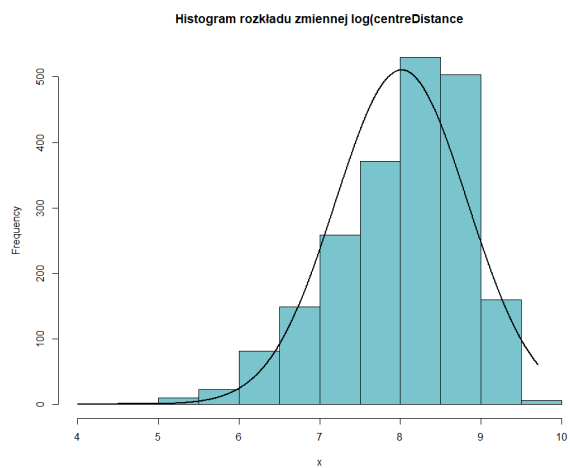
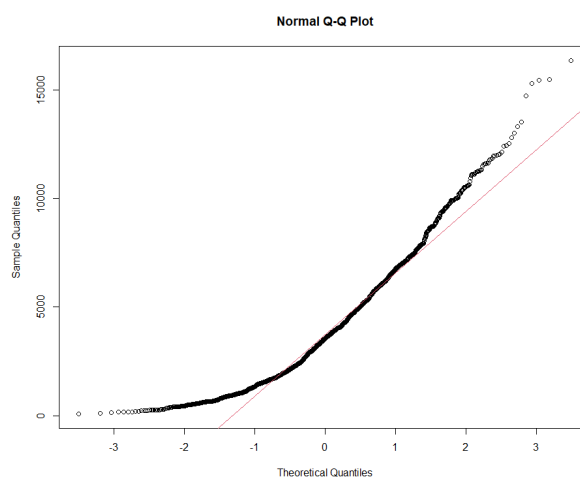
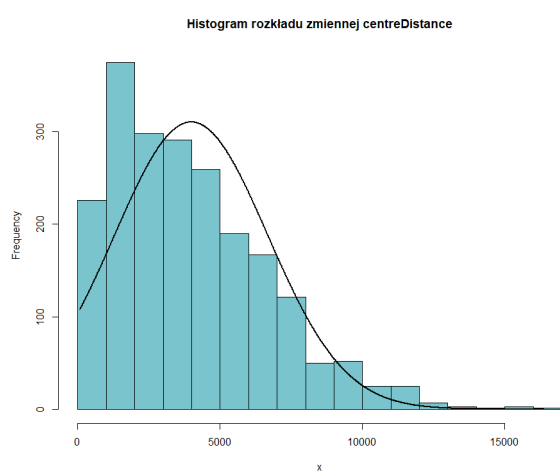
## Rozkład zmiennej $\text{postOfficeDistance}$ i zmiennej $\log(\text{postOfficeDistance})$ :



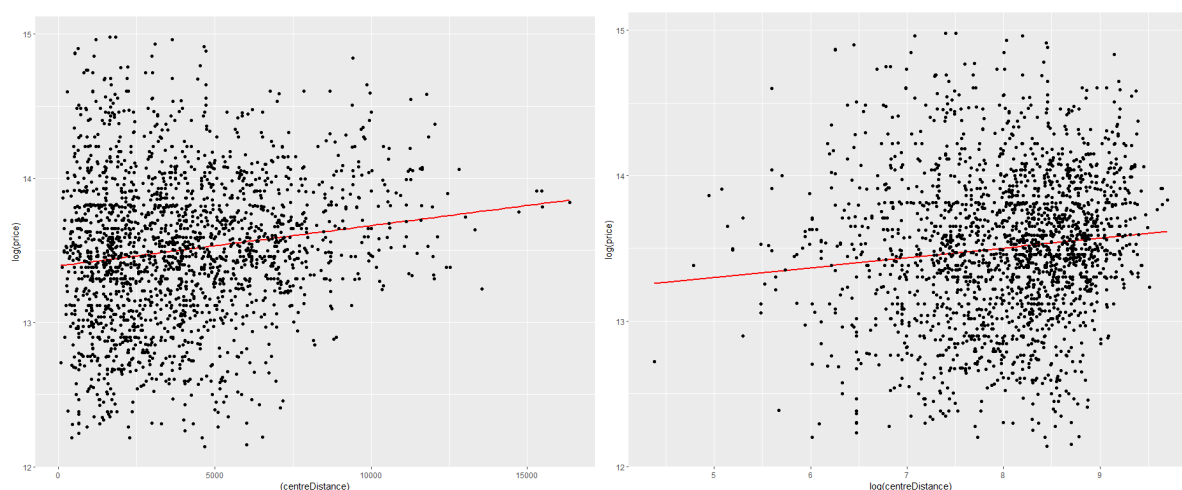
## Wykresy rozproszenia:



## Rozkład zmiennej $\text{centreDistance}$ i $\log(\text{centreDistance})$ :



### Wykresy rozproszenia:



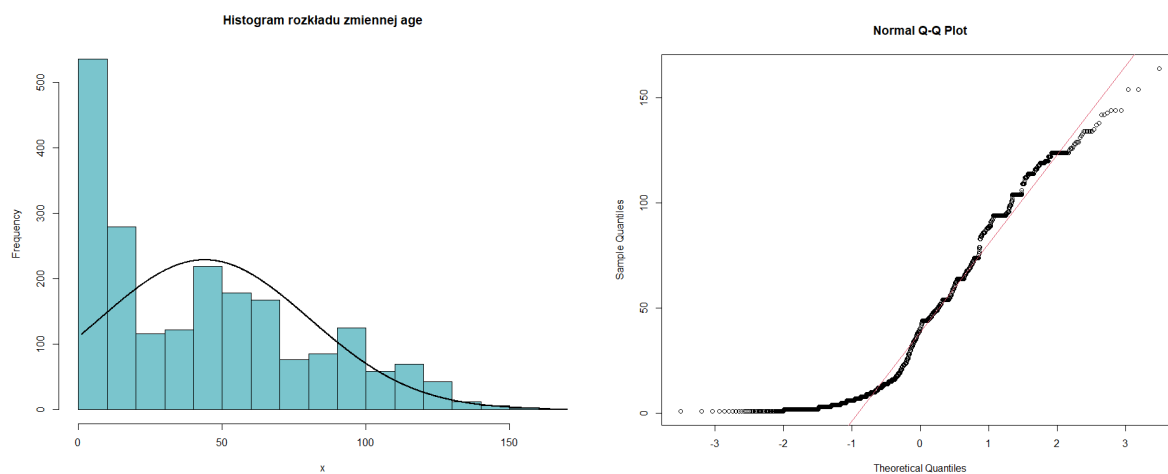
Wiek mieszkania również został tutaj potraktowany jako zmienna ciągła – *age*. Oferty uwzględnione w bazie danych obejmowały wyłącznie rynek wtórny, zatem minimalny wiek mieszkania wyniósł 1 rok. Najstarsze mieszkanie liczyło natomiast aż 164 lata i jak można się spodziewać, była to kamienica. Przeciętny lokal miał 43,86 lat. 25% mieszkań wybudowane zostało mniej niż dekadę temu, a trzy czwarte mieszkań mniej niż 67 lat temu.

| Minimum | 1st Qu. | Median | Mean  | 3rd Qu. | Maximum |
|---------|---------|--------|-------|---------|---------|
| 1.00    | 10.00   | 40.00  | 43.86 | 67.00   | 164.00  |

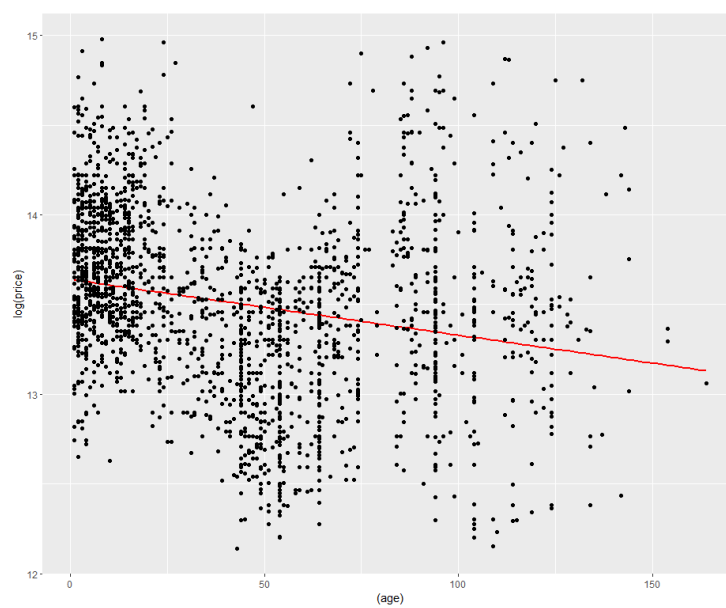
**Tabela 4.** Podstawowe statystyki dla zmiennej *age*. Opracowanie własne.

Na rynku występuje zatem znaczna przewaga mieszkań młodych, szczególnie 10-20 letnich, co bardzo dobrze widać na histogramie rozkładu zmiennej *age*. Wynik testu Jarque – Bera (p-value <5%) wskazuje na konieczność odrzucenia hipotezy o normalności rozkładu zmiennej *age*. Niestety podjęte próby „polepszenia” sytuacji, poprzez przekształcanie zmiennej również okazały się nieskuteczne.

## Rozkład zmiennej *age*:



## Wykres rozproszenia:



### 5.3 Zmienne niezależne zero – jedynkowe

W modelu zawarte są dwie charakterystyki mieszkań, które opisane są za pomocą tzw. *dummy variables*: typ budownictwa oraz liczba pokoi. Aby wykluczyć problem liniowej zależności poszczególnych kategorii tych zmiennych konieczne było ustalenie ich poziomów bazowych. Zazwyczaj w tym celu wybierane są grupy najmniej lub najbardziej liczne, ja zdecydowałam się na wybranie tych drugich. Najobszerniejszą kategorią, jeśli chodzi o zmienną *building type* była grupa mieszkań zlokalizowanych w zwykłych blokach mieszkalnych – 908 obserwacji. Jako poziom bazowy liczby pokoi przyjąłam kategorię, zawierającą mieszkania 3 – 4 pokojowe - liczącą 1074 obserwacje.

Wśród badanych lokali mieszkaniowych większość jest własnościowa i w dobrym stanie. Ponadto przeważająca część z nich nie znajduje się ani na parterze ani na najwyższym piętrze. Lekko powyżej połowa mieszkań może pochwalić się obecnością takich udogodnień jak winda czy balkon, a dużo z nich (ale mniej niż połowa) ma także miejsce parkingowe. Tylko nieliczne domy objęte są opieką firm ochroniarskich, co raczej nie jest zaskakujące, szczególnie na starszych osiedlach. Warto zwrócić uwagę na fakt, iż ponad ¼ badanych ofert pochodzi z Warszawy (Tabela 5.)

| ZMIENNA                         | LICZEBNOŚĆ DLA WARTOŚCI: 1 | LICZEBNOŚĆ DLA WARTOŚCI: 0 |
|---------------------------------|----------------------------|----------------------------|
| ownership                       | 1802                       | 292                        |
| condition                       | 1367                       | 727                        |
| hasParkingSpace                 | 868                        | 1226                       |
| hasBalcony                      | 1245                       | 849                        |
| hasElevator                     | 1129                       | 965                        |
| hasSecurity                     | 95                         | 1999                       |
| hasStorageRoom                  | 1038                       | 1056                       |
| first_floor                     | 397                        | 1697                       |
| top_floor                       | 435                        | 1659                       |
| stolica                         | 631                        | 1463                       |
| LICZEBNOŚCI DLA DUMMY VARIABLES |                            |                            |
| building type                   | block:                     | 908                        |
|                                 | apartmentB:                | 697                        |
|                                 | tenement:                  | 489                        |
| rooms                           | rooms1:                    | 933                        |
|                                 | rooms2:                    | 1074                       |
|                                 | rooms3:                    | 87                         |

**Tabela 5.** Liczebności dla zmiennych 0-1. Opracowanie własne.

## 6. Estymacja modelu

### 6.1 Model pierwotny

Analizę zależności zmiennej *price* od zmiennych objaśniających rozpocząłam od doprowadzenia modelu regresji liniowej do poprawnej formy funkcyjnej oraz zbadania, czy nie występuje współliniowość między czynnikami. W przeprowadzonych testach przyjęty poziom ufności to 5%.

Model pierwszy (model 1 w Tabeli 6.) zawierał prawie wszystkie zmienne w wybranych wcześniej formach, jednak jeszcze bez interakcji i uwzględnienia oddzielnej kategorii dla mieszkań ze stolicy. Wskaźnik  $R^2$  wyniósł 64,3%, co oznaczało, że model ten wyjaśniał 64,3% zmienności cen mieszkań w logarytmie. Wynik testu RESET, z p-value znacznie większym niż 5%, pozwolił na nie odrzucenie hipotezy o liniowości formy funkcyjnej. Spełnione zostało zatem pierwsze założenie Klasycznego Modelu Regresji Liniowej.

```
RESET test
data: model1
RESET = 0.14829, df1 = 2, df2 = 2067, p-value = 0.8622
```

Następnie w celu sprawdzenia, czy konieczne jest uwzględnienie zmiennej *stolica* w regresji, przeprowadziłam test Chowa. Jego ideą jest porównanie, czy parametry modelu zmieniają się między dwoma lub więcej podgrupami danych. Wynik testu był jednoznaczny: p-value na poziomie bliskim zera wskazywało na konieczność odrzucenia hipotezy zerowej o stabilności parametrów między próbą mieszkań ze stolicy, a próbą spoza Warszawy. Zdecydowałam się więc na włączenie do modelu zmiennej *stolica* (model 2 w Tabeli 6.). Model drugi przeszedł również test RESET.

```
RESET test
data: model2
RESET = 1.0663, df1 = 2, df2 = 2066, p-value = 0.3445
```

Kolejnym istotnym aspektem przy tworzeniu modelu regresji liniowej jest współliniowość zmiennych objaśniających. To zjawisko, w którym dwie lub więcej zmiennych niezależnych w modelu są silnie skorelowane, co może prowadzić do problemów interpretacyjnych. Współliniowość może sprawić, że ocena oddziaływania poszczególnych zmiennych staje się trudna, a estymatory parametrów stają się nieprecyzyjne. Aby skontrolować poziom współliniowości zmiennych użyłam wskaźnika Variance Inflation Factor (VIF).



Wartość VIF wyższa niż 5 wskazywałaby bowiem na wysoką współliniowość, a powyżej 10 na konieczność usunięcia skorelowanych zmiennych.

```
> vif(model2)
log(squareMeters)      rooms1      rooms3      log(centreDistance)      log(schoolDistance)      log(clinicDistance)
log(postOfficeDistance) log(kindergartenDistance) log(restaurantDistance) log(collegeDistance) log(pharmacyDistance)      ownership
buildingMaterial      condition      hasParkingSpace      hasBalcony      hasElevator      hasStorageRoom
1.964664      1.338527      1.312598      1.160863      1.810418      1.379976
hasSecurity      first_floor      top_floor      age      apartmentB      tenement
1.040404      1.100392      1.098870      5.902201      2.525149      4.040095
stolica
1.487971
```

Jedynie VIF dla zmiennej *age* okazał się być na poziomie wyższym niż 5, co świadczyło o dość wysokiej korelacji z innymi zmiennymi w modelu. Jednak żadna zmienna nie zakwalifikowała się do usunięcia.

W modelu trzecim (model 3 w Tabeli 6.) zdecydowałam się na uwzględnienie kilku interakcji, które mogły mieć wpływ na badaną zmienną zależną oraz estymacje współczynników:

- ***Log(centreDistance)\*log(squareMeters)*** – mieszkania budowane w centrum miast mogą mieć raczej mniejszą powierzchnię, niż te na obrzeżach, gdzie zaludnienie jest rzadsze, dlatego relacja „cena – bliskość centrum” może być zaburzona przez czynnik jakim jest metraż;
- ***Log(centreDistance)\*hasParkingSpace*** – miejsce parkingowe może być szczególnie ważne dla osób dojeżdżających do np. pracy w centrum samochodem, a zatem efekt zmiennej *centreDistance* w połączeniu ze zmienną *has ParkingSpace* może być wzmocniony;
- ***Log(schoolDistance)\*hasParkingSpace*** – osoby zawożące dzieci do szkoły samochodem, czyli mieszkające dalej od szkół, mogą bardziej cenić sobie posiadanie miejsca parkingowego i być w stanie jeszcze więcej zapłacić za mieszkanie;
- ***Ownership\*tenement*** – nabywcy mieszkań w kamienicach często decydują się na ich poważne renowacje i remonty, a właściciele mieszkań własnościowych mają większą autonomię w podejmowaniu decyzji odnośnie modernizacji swoich lokali niż właściciele mieszkań spółdzielczych;

- ***Log(squareMeters)\*condition*** – kondycja i stan mieszkania może być bardzo istotny dla nabywców dużych lokali, gdyż większe mieszkania zazwyczaj generują wyższe koszty utrzymania, a w związku z tym, jeśli stan techniczny i kondycja mieszkania są dobre, właściciel może uniknąć kosztownych napraw i modernizacji w przyszłości;
- ***Log(centreDistance)\*stolica*** – szczególnie w dużych metropoliach dostępność gruntów budowlanych w centrum jest ograniczona, a mieszkania tam są uważane za najbardziej prestiżowe, zatem efekt odległości od centrum może być najbardziej widoczny właśnie w Warszawie.

Model trzeci charakteryzował się zdecydowanie najwyższym skorygowanym  $R^2$  z dotychczasowych modeli oraz wyjaśniał on 73,8% zmienności  $y$  (Tabela 6.). W następnej części badania przystąpiłam więc do przeprowadzenia kolejnych testów diagnostycznych w celu sprawdzenia, czy regresja ta spełnia założenia Klasycznego Modelu Regresji Liniowej.

| Dependent variable:                   |                            |                            |                            |
|---------------------------------------|----------------------------|----------------------------|----------------------------|
|                                       | (1)                        | log(price)<br>(2)          | (3)                        |
| log(squareMeters)                     | 0.791***<br>(0.032)        | 0.775***<br>(0.028)        | 0.431***<br>(0.163)        |
| rooms1                                | -0.010<br>(0.020)          | -0.014<br>(0.018)          | -0.008<br>(0.017)          |
| rooms3                                | 0.025<br>(0.037)           | 0.019<br>(0.033)           | 0.006<br>(0.032)           |
| log(centreDistance)                   | 0.157***<br>(0.011)        | 0.040***<br>(0.011)        | -0.107<br>(0.083)          |
| log(schoolDistance)                   | -0.025**<br>(0.011)        | -0.011<br>(0.009)          | -0.022*<br>(0.012)         |
| log(clinicDistance)                   | -0.062***<br>(0.009)       | -0.022***<br>(0.008)       | -0.017**<br>(0.008)        |
| log(postOfficeDistance)               | 0.029***<br>(0.010)        | 0.015*<br>(0.009)          | 0.010<br>(0.008)           |
| log(kindergartenDistance)             | -0.013<br>(0.010)          | 0.001<br>(0.009)           | -0.001<br>(0.008)          |
| log(restaurantDistance)               | -0.100***<br>(0.009)       | -0.060***<br>(0.008)       | -0.065***<br>(0.007)       |
| log(collegeDistance)                  | -0.056***<br>(0.010)       | -0.049***<br>(0.009)       | -0.048***<br>(0.008)       |
| log(pharmacyDistance)                 | -0.011<br>(0.010)          | -0.011<br>(0.009)          | -0.005<br>(0.009)          |
| ownership                             | -0.013<br>(0.022)          | 0.020<br>(0.020)           | 0.029<br>(0.019)           |
| buildingMaterial                      | 0.121***<br>(0.023)        | 0.103***<br>(0.020)        | 0.094***<br>(0.020)        |
| condition                             | 0.158***<br>(0.016)        | 0.168***<br>(0.014)        | 0.427***<br>(0.144)        |
| hasParkingSpace                       | 0.026*<br>(0.015)          | 0.023*<br>(0.014)          | 0.315**<br>(0.131)         |
| hasBalcony                            | 0.015<br>(0.015)           | 0.019<br>(0.013)           | 0.022*<br>(0.013)          |
| hasElevator                           | 0.227***<br>(0.018)        | 0.153***<br>(0.016)        | 0.136***<br>(0.017)        |
| hasStorageRoom                        | -0.003<br>(0.016)          | 0.006<br>(0.014)           | -0.003<br>(0.013)          |
| hasSecurity                           | 0.065**<br>(0.032)         | 0.007<br>(0.029)           | 0.009<br>(0.028)           |
| first_floor                           | 0.011<br>(0.018)           | 0.025<br>(0.016)           | 0.024<br>(0.015)           |
| top_floor                             | -0.031*<br>(0.017)         | -0.028*<br>(0.015)         | -0.033*<br>(0.020)         |
| age                                   | -0.002***<br>(0.0004)      | -0.002***<br>(0.0004)      | -0.001***<br>(0.0004)      |
| apartmentB                            | 0.139***<br>(0.022)        | 0.143***<br>(0.020)        | 0.147***<br>(0.019)        |
| tenement                              | 0.196***<br>(0.031)        | 0.129***<br>(0.028)        | 0.245***<br>(0.089)        |
| stolica                               |                            | 0.364***<br>(0.016)        | 1.971***<br>(0.149)        |
| hasElevator:top_floor                 |                            |                            | 0.021<br>(0.029)           |
| log(squareMeters):log(centreDistance) |                            |                            | 0.053***<br>(0.020)        |
| log(centreDistance):hasParkingSpace   |                            |                            | -0.064***<br>(0.016)       |
| log(schoolDistance):hasParkingSpace   |                            |                            | 0.040**<br>(0.016)         |
| ownership:tenement                    |                            |                            | -0.125<br>(0.087)          |
| log(squareMeters):condition           |                            |                            | -0.063*<br>(0.035)         |
| log(centreDistance):stolica           |                            |                            | -0.194***<br>(0.018)       |
| Constant                              | 10.098***<br>(0.170)       | 10.419***<br>(0.152)       | 11.368***<br>(0.670)       |
| Observations                          | 2,094                      | 2,094                      | 2,094                      |
| R2                                    | 0.644                      | 0.717                      | 0.738                      |
| Adjusted R2                           | 0.640                      | 0.714                      | 0.734                      |
| Residual Std. Error                   | 0.304 (df = 2069)          | 0.271 (df = 2068)          | 0.261 (df = 2061)          |
| F Statistic                           | 155.723*** (df = 24; 2069) | 209.552*** (df = 25; 2068) | 181.711*** (df = 32; 2061) |
| Note: *p<0.1; **p<0.05; ***p<0.01     |                            |                            |                            |

Tabela 6. Kolejno wyniki modelu 1, 2 oraz 3.

## 6.2 Testy diagnostyczne

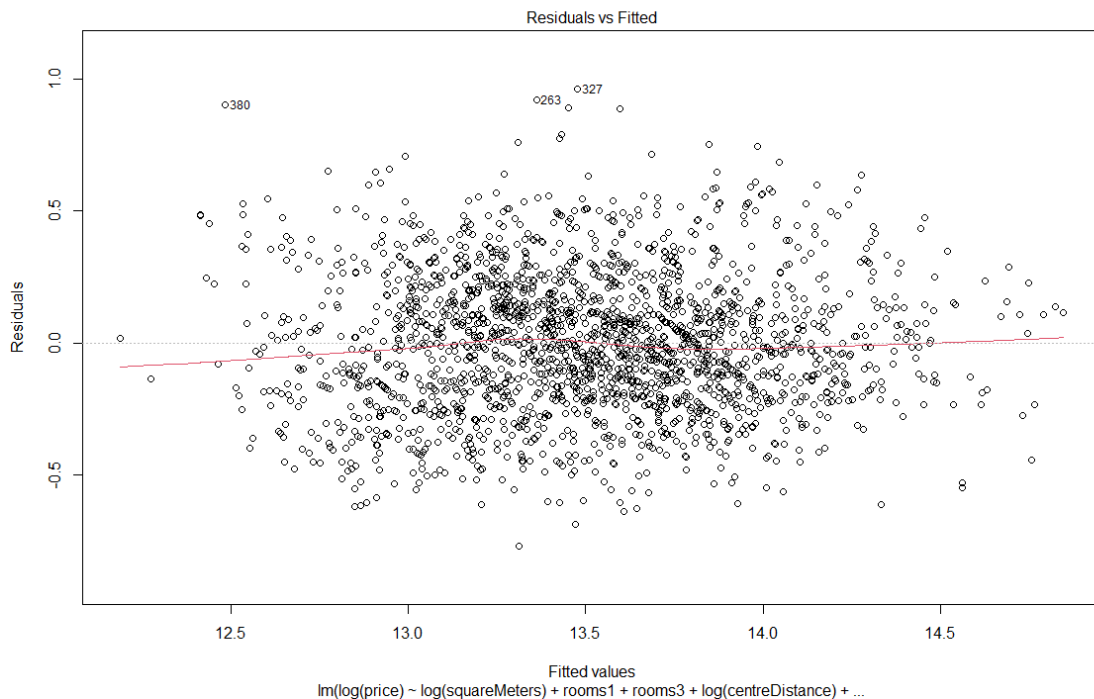
### a) Liniowość formy funkcyjnej

Model trzeci poddany został testowi RESET, którego wynik jednoznacznie pozwolił na przyjęcie  $H_0$  o poprawności formy funkcyjnej. P – value wyniosło bowiem 0.8681 przy założonym poziomie istotności 5%.

```
RESET test  
data: model3  
RESET = 0.14146, df1 = 2, df2 = 2059, p-value = 0.8681
```

### b) Homoskedastyczność reszt

Stałość wariancji reszt w modelu jest kluczowa, gdyż w przypadku niespełnienia tego warunku błędy standardowe dla współczynników regresji są nieprawidłowe, a co za tym idzie odpowiadające im p-value również. Nie możemy wtedy interpretować parametrów modelu w poprawny sposób, gdyż są one obciążone. Założenie o homoskedastyczności reszt dla modelu 3. sprawdziłam przy pomocy wykresu Residuals vs. Fitted oraz posługując się testem Breuscha – Pagana.



Wykres Residuals vs. Fitted pozwala nam skontrolować, czy w naszym modelu istnieje zależność między resztami a wartościami dopasowanymi. Niestety w tym przypadku wskazywał on na heteroskedastyczność reszt – punkty miały tendencję do skupiania się głównie w środkowej części wykresu, a czerwona linia opadała w dół przy początku osi OX. Potwierdził to także wynik testu Breuscha – Pagana. Wartość p-value bliska zera przesądziła o odrzuceniu hipotezy zerowej o homoskedastyczności reszt.

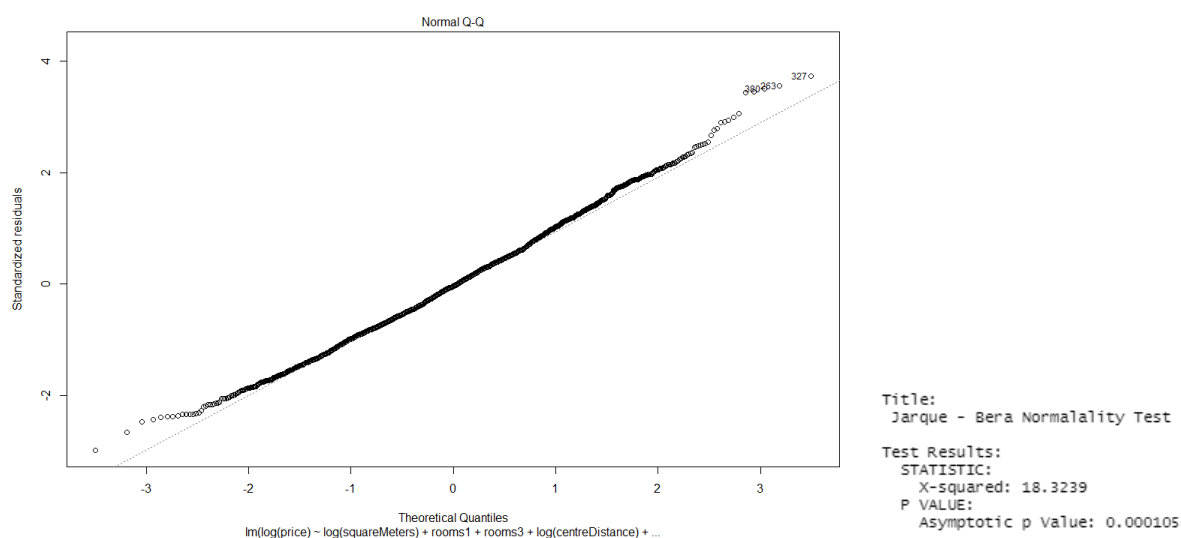
```
Breusch-Pagan test

data: model3
BP = 373.87, df = 32, p-value < 0.00000000000000022
```

### c) Normalność reszt

Testowanie założenia o normalności reszt jest szczególnie ważne w przypadku badania małej próby obserwacji. W takim przypadku jeśli model nie spełnia tego warunku, to nie możemy użyć standardowego wnioskowania, gdyż statystyki nie mają rozkładu t oraz F.

Do sprawdzenia tego założenia dla modelu trzeciego posłużyłam się wykresem Normal Q-Q Plot oraz testem Jarque – Bera na normalność rozkładu.



Na wykresie Normal Q-Q Plot na osi OX rozmieszczone są kwantyle teoretyczne dla rozkładu normalnego, a na osi OY kwantyle empiryczne reszt. Dla omawianego modelu większość punktów ułożona była na linii przerywanej, poza obserwacjami na końcach osi, które zdecydowanie zaburzyły normalność rozkładu. Świadczył również

o tym wynik testu na normalność Jarque – Bera:  $p\text{-value} = 0.000105$  oznaczało konieczność odrzucenia  $H_0$  o normalności rozkładu.

Nie był to jednak powód do zmartwienia, gdyż baza danych wzięta do badania obejmowała 2094 obserwacje, co można uznać za wielką próbę. W tym wypadku zawsze możemy korzystać ze standardowego wnioskowania, ze względu na Prawo Wielkich Liczb i Centralne Twierdzenie Graniczne.

### 6.3 Model ostateczny

Aby rozwiązać występujący w modelu problem heteroskedastyczności reszt i móc w poprawny sposób interpretować parametry zastosowałam macierz odporną wariancji – kowariancji reszt MacKinnona i White’a. Wybrałam tę metodę, gdyż heteroskedastyczność prawdopodobnie wynikała z oddziaływania kilku zmiennych (wyklucza użycie WMNK), a zidentyfikowanie ich byłoby w tym przypadku bardzo pracochłonne (wyklucza użycie USMNK). Zastosowanie macierzy wpłynęło na poziomy istotności zmiennych (Tabela 7.)

Duża część zmiennych użytych w pierwotnym modelu okazała się być statystycznie nieistotna na poziomie ufności 5%. Za pomocą testu Walda zdecydowałam się na sprawdzenie ich łącznej nieistotności.

```
wald test:
-----

Chi-squared test:
X2 = 964.1, df = 11, P(> X2) = 0.0
```

Wynik testu pozwolił na odrzucenie  $H_0$  o łącznej nieistotności zmiennych. Zatem aby wyeliminować zmienne nieistotne statystycznie z modelu musiałam zastosować metodę usuwania ich po kolei. Patrząc na ich estymacje w macierzy odpornej dla modelu 3. kolejno odejmowałam po jednej zmiennej, zaczynając od tych z najwyższym p-value.

Model końcowy zawarł w sobie 15 zmiennych i 4 interakcje (Tabela 8.). Spełniał on założenie o poprawności funkcyjnej, co sprawdzone zostało testem RESET. Brak normalności reszt, tak jak dla poprzedniego modelu, nie jest w tym przypadku problemem. Heteroskedastyczność reszt zminimalizowana została poprzez zastosowanie macierzy odpornej wariancji-kowariancji MacKinnonna i White’a. Tym sposobem uzyskałam ostateczny model, na podstawie którego przystąpię do interpretacji współczynników i weryfikacji postawionych na początku badania hipotez.

| Dependent variable:                   |                             |                       |
|---------------------------------------|-----------------------------|-----------------------|
|                                       | log(price)<br>OLS           | coefficient<br>test   |
|                                       | (1)                         | (2)                   |
| log(squareMeters)                     | 0.431***<br>(0.163)         | 0.431**<br>(0.193)    |
| rooms1                                | -0.008<br>(0.017)           | -0.008<br>(0.017)     |
| rooms3                                | 0.006<br>(0.032)            | 0.006<br>(0.036)      |
| log(centreDistance)                   | -0.107<br>(0.083)           | -0.107<br>(0.093)     |
| log(schoolDistance)                   | -0.022*<br>(0.012)          | -0.022*<br>(0.012)    |
| log(clinicDistance)                   | -0.017**<br>(0.008)         | -0.017**<br>(0.008)   |
| log(postofficeDistance)               | 0.010<br>(0.008)            | 0.010<br>(0.008)      |
| log(kindergartenDistance)             | -0.001<br>(0.008)           | -0.001<br>(0.008)     |
| log(restaurantDistance)               | -0.065***<br>(0.007)        | -0.065***<br>(0.008)  |
| log(collegeDistance)                  | -0.048***<br>(0.008)        | -0.048***<br>(0.008)  |
| log(pharmacyDistance)                 | -0.005<br>(0.009)           | -0.005<br>(0.008)     |
| ownership                             | 0.029<br>(0.019)            | 0.029*<br>(0.017)     |
| buildingMaterial                      | 0.094***<br>(0.020)         | 0.094***<br>(0.019)   |
| condition                             | 0.427***<br>(0.144)         | 0.427**<br>(0.170)    |
| hasParkingSpace                       | 0.315**<br>(0.131)          | 0.315**<br>(0.145)    |
| hasBalcony                            | 0.022*<br>(0.013)           | 0.022*<br>(0.013)     |
| hasElevator                           | 0.136***<br>(0.017)         | 0.136***<br>(0.017)   |
| hasStorageRoom                        | -0.003<br>(0.013)           | -0.003<br>(0.014)     |
| hasSecurity                           | 0.009<br>(0.028)            | 0.009<br>(0.025)      |
| first_floor                           | 0.024<br>(0.015)            | 0.024<br>(0.016)      |
| top_floor                             | -0.033*<br>(0.020)          | -0.033<br>(0.022)     |
| age                                   | -0.001***<br>(0.0004)       | -0.001***<br>(0.0004) |
| apartments                            | 0.147***<br>(0.019)         | 0.147***<br>(0.017)   |
| tenement                              | 0.245***<br>(0.089)         | 0.245***<br>(0.088)   |
| stolica                               | 1.971***<br>(0.149)         | 1.971***<br>(0.133)   |
| hasElevator:top_floor                 | 0.021<br>(0.029)            | 0.021<br>(0.029)      |
| log(squareMeters):log(centreDistance) | 0.053***<br>(0.020)         | 0.053**<br>(0.023)    |
| log(centreDistance):hasParkingSpace   | -0.064***<br>(0.016)        | -0.064***<br>(0.017)  |
| log(schoolDistance):hasParkingSpace   | 0.040**<br>(0.016)          | 0.040**<br>(0.016)    |
| ownership:tenement                    | -0.125<br>(0.087)           | -0.125<br>(0.087)     |
| log(squareMeters):condition           | -0.063*<br>(0.035)          | -0.063<br>(0.042)     |
| log(centreDistance):stolica           | -0.194***<br>(0.018)        | -0.194***<br>(0.016)  |
| Constant                              | 11.368***<br>(0.670)        | 11.368***<br>(0.788)  |
| Observations                          | 2,094                       |                       |
| R2                                    | 0.738                       |                       |
| Adjusted R2                           | 0.734                       |                       |
| Residual Std. Error                   | 0.261 (df = 2061)           |                       |
| F Statistic                           | 181.711*** (df = 32; 2061)  |                       |
| Note:                                 | *p<0.1; **p<0.05; ***p<0.01 |                       |

**Tabela 7.** Wyniki odpowiednio dla modelu 3. i macierzy odpornej.

| Dependent variable:                   |                             |                       |
|---------------------------------------|-----------------------------|-----------------------|
|                                       | log(price)<br>OLS           | coefficient<br>test   |
|                                       | (1)                         | (2)                   |
| log(squareMeters)                     | 0.430***<br>(0.160)         | 0.430**<br>(0.184)    |
| log(centreDistance)                   | -0.093<br>(0.082)           | -0.093<br>(0.091)     |
| log(schoolDistance)                   | -0.021*<br>(0.011)          | -0.021*<br>(0.012)    |
| log(clinicDistance)                   | -0.018**<br>(0.008)         | -0.018**<br>(0.008)   |
| log(restaurantDistance)               | -0.064***<br>(0.007)        | -0.064***<br>(0.007)  |
| log(collegeDistance)                  | -0.049***<br>(0.008)        | -0.049***<br>(0.008)  |
| buildingMaterial                      | 0.101***<br>(0.018)         | 0.101***<br>(0.017)   |
| condition                             | 0.171***<br>(0.014)         | 0.171***<br>(0.014)   |
| hasParkingSpace                       | 0.289**<br>(0.131)          | 0.289**<br>(0.143)    |
| hasElevator                           | 0.140***<br>(0.015)         | 0.140***<br>(0.015)   |
| top_floor                             | -0.031**<br>(0.014)         | -0.031**<br>(0.014)   |
| age                                   | -0.002***<br>(0.0004)       | -0.002***<br>(0.0004) |
| apartmentB                            | 0.150***<br>(0.019)         | 0.150***<br>(0.017)   |
| tenement                              | 0.132***<br>(0.027)         | 0.132***<br>(0.028)   |
| stolica                               | 2.018***<br>(0.147)         | 2.018***<br>(0.132)   |
| log(squareMeters):log(centreDistance) | 0.049**<br>(0.020)          | 0.049**<br>(0.022)    |
| log(centreDistance):hasParkingSpace   | -0.062***<br>(0.016)        | -0.062***<br>(0.017)  |
| log(schoolDistance):hasParkingSpace   | 0.042**<br>(0.016)          | 0.042***<br>(0.016)   |
| log(centreDistance):stolica           | -0.199***<br>(0.018)        | -0.199***<br>(0.016)  |
| Constant                              | 11.414***<br>(0.652)        | 11.414***<br>(0.748)  |
| Observations                          | 2,094                       |                       |
| R2                                    | 0.736                       |                       |
| Adjusted R2                           | 0.734                       |                       |
| Residual Std. Error                   | 0.261 (df = 2074)           |                       |
| F Statistic                           | 304.809*** (df = 19; 2074)  |                       |
| Note:                                 | *p<0.1; **p<0.05; ***p<0.01 |                       |

**Tabela 8.** Wyniki ostatecznego modelu przed i po użyciu macierzy odpornej.



#### Test RESET dla ostatecznego modelu:

```
RESET test  
data: model16  
RESET = 0.70578, df1 = 2, df2 = 2072, p-value = 0.4938
```

#### Test Breuscha – Pagana dla ostatecznego modelu:

```
Breusch-Pagan test  
data: model16  
BP = 338.48, df = 19, p-value < 0.00000000000000022
```

#### Test Jarque – Bera dla ostatecznego modelu:

```
Title:  
Jarque - Bera Normality Test  
  
Test Results:  
STATISTIC:  
X-squared: 17.7344  
P VALUE:  
Asymptotic p Value: 0.0001409
```

## **7. Interpretacja wyników**

Model wyjaśnia 73.6% zmienności zmiennej objaśnianej. Poniżej przedstawiam interpretację oszacowań współczynników  $\beta$  dla wszystkich zmiennych i interakcji. Stała została celowo pominięta, gdyż jej interpretacja nie ma w tym przypadku sensu.

#### Zmienna log(squareMeters):

Współczynnik wskazuje, że wzrost powierzchni mieszkania o 1% przekłada się na wzrost jego ceny o średnio 0.43% (przy innych warunkach niezmiennych). Zmienna jest istotna na poziomie ufności 5%.

#### Zmienna log(centreDistance):

Współczynnik przy zmiennej wskazuje na ujemną zależność ceny od odległości od centrum miasta, jednak nie jest on statystycznie istotny w tym modelu.

#### Zmienna log(schoolDistance):

Zmienna nie jest istotna statystycznie na poziomie ufności 5%, jednak współczynnik wskazuje na ujemną zależność ceny mieszkania od odległości od najbliższej szkoły. Wzrost dystansu o 1% przekłada się na spadek ceny o średnio 0.021%.

#### Zmienna log(clinicDistance):

Wzrost odległości od najbliższej przychodni o 1% powoduje spadek ceny mieszkania o 0.018%. Zmienna jest istotna statystycznie na poziomie ufności 5%.

#### Zmienna log(restaurantDistance):

Wzrost odległości od najbliższej restauracji o 1% przekłada się na spadek ceny mieszkania o średnio 0.064%. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna log(collegeDistance):

Wzrost odległości od najbliższej szkoły wyższej o 1% powoduje spadek ceny mieszkania o 0.049%. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna buildingMaterial:

Współczynnik wskazuje, że mieszkanie zbudowane z cegły jest średnio o 10.1% droższe od mieszkań wybudowanych z innych materiałów (przy pozostałych warunkach takich samych). Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna condition:

Mieszkanie o wyższym standardzie jest średnio o 17.1% droższe od mieszkań w gorszym stanie. Zmienna jest istotna statystycznie o na poziomie ufności 1%.

#### Zmienna hasParkingSpace:

Mieszkanie posiadające miejsce parkingowe jest średnio o 28.9% droższe od mieszkań nie mających go w swojej ofercie. Zmienna jest istotna statystycznie na poziomie ufności 5%.

#### Zmienna hasElevator:

Mieszkanie posiadające w budynku windę jest średnio o 14% droższe od lokali bez windy. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna top\_floor:

Mieszkanie na ostatnim piętrze jest średnio o 3.1% tańsze od mieszkań na innych piętrach. Zmienna jest istotna statystycznie na poziomie ufności 5%.

#### Zmienna age:

Mieszkanie „starsze” o rok jest średnio o 0.2% tańsze od mieszkania młodszego. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna apartmentB:

Mieszkania w apartamentowcach są średnio o 15% droższe od mieszkań w blokach mieszkalnych. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna tenement:

Mieszkania w kamienicach są średnio o 13.2% droższe od mieszkań w blokach mieszkalnych. Zmienna jest istotna statystycznie na poziomie ufności 1%.

#### Zmienna stolica:

Zmienna jest istotna statystycznie na poziomie ufności 1%. Współczynnik przy tej zmiennej wyniósł 2.018, co wskazuje, że mieszkania w Warszawie są średnio 6.52 razy droższe od mieszkań w innych miastach.

#### Interakcja $\log(\text{squareMeters}) * \log(\text{centreDistance})$ :

Jednoczesny wzrost powierzchni i odległości od centrum miasta o 1% przekłada się na wzrost ceny średnio o 0.049% - wraz ze wzrostem odległości od centrum mieszkania większe będą doświadczać mniejszego spadku ceny niż mieszkania mniejsze. Interakcja jest istotna statystycznie na poziomie 5%.

#### Interakcja $\log(\text{centreDistance}) * \text{hasParkingSpace}$ :

W przypadku mieszkania z miejscem parkingowym, wzrost odległości od centrum o 1% przekłada się na spadek ceny o średnio 0.155% - dla mieszkań z miejscem parkingowym spadek ceny związany wyłącznie ze wzrostem odległości od centrum jest silniejszy niż dla mieszkań bez miejsca parkingowego. Zmienna jest statystycznie istotna na poziomie 1%.

#### Interakcja $\log(\text{schoolDistance}) * \text{hasParkingSpace}$ :

W przypadku mieszkania z miejscem parkingowym, wzrost odległości od najbliższej szkoły o 1% przekłada się na wzrost ceny o średnio 0.21% - dla mieszkań z miejscem parkingowym spadek ceny związany wyłącznie ze wzrostem odległości od szkoły nie występuje, a wręcz przeciwnie wraz ze wzrostem odległości od szkoły cena rośnie. Zmienna jest statystycznie istotna na poziomie 5%.

#### Interakcja $\log(\text{centreDistance}) * \text{stolica}$ :

Jeśli mieszkanie znajduje się w Warszawie, wzrost odległości od centrum miasta o 1% wiąże się ze spadkiem jego ceny o średnio 0.292% - spadek cen związany wyłącznie ze wzrostem odległości od centrum jest silniejszy dla mieszkań w stolicy niż w innych miejscowościach.

## **8. Weryfikacja hipotez**

**H1:** Tak jak wskazywała literatura, powierzchnia mieszkania okazała się istotna statystycznie na poziomie ufności 5%, a jej wzrost wiąże się ze wzrostem cen mieszkań, co potwierdza hipotezę 1.

**H2:** Liczba pokoi okazała się statystycznie istotna w modelu, zatem nie można na tej podstawie stwierdzić czy mieszkania z dużą liczbą pokoi są średnio tańsze od tych z mniejszą ich liczbą. Hipoteza 2. odrzucona.

**H3:** Stan mieszkania okazał się zmienną statystycznie istotną w modelu i wskazywał na pozytywną relację wysokiego standardu mieszkania z jego ceną, co wskazuje na słuszność hipotezy 3.

**H4:** Statystycznie istotne na poziomie przynajmniej 5% okazały się zmienne dotyczące jedynie dystansu do najbliższej restauracji, szkoły wyższej i przychodni. Wzrost odległości od tych

*points of interest* przekładał się na spadek cen mieszkań. Odległość od centrum może mieć istotny wpływ na cenę mieszkań w stolicy lub w przypadku mieszkań posiadających miejsce parkingowe. Ciężko jest jednoznacznie określić efekt tej zmiennej. Jeśli chodzi o odległość od najbliższej szkoły, to na poziomie ufności przynajmniej 5% była ona istotna statystycznie tylko dla mieszkań z miejscem parkingowym. Uważam, że hipotezę 4. można uznać za potwierdzoną, gdyż faktycznie bliskość od ważnych miejsc ma wpływ na cenę w większości przypadków.

**H5:** Własność mieszkania okazała się zmienną nieistotną statystycznie w modelu, co wskazuje na konieczność odrzucenia hipotezy 5.

**H6:** Zmienna *buildingMaterial* ma w modelu istotny statystycznie wpływ na cenę. Mieszkania wykonane z lepszego jakościowo materiału (cegła) są średnio droższe od tych zbudowanych z innych materiałów o średnio około 10%, co potwierdza hipotezę 6.

**H7:** Zmienna *first\_floor* okazała się zmienną nieistotną statystycznie w modelu, więc fakt, że mieszkanie znajduje się na pierwszym piętrze nie ma wpływu na cenę. Możemy natomiast jednoznacznie stwierdzić, że mieszkania na najwyższym piętrze są najtańsze. Hipoteza 7. nie została więc potwierdzona całkowicie.

**H8:** Z hipotezą 8. również nie można zgodzić się w 100%, gdyż jedynymi udogodnieniami, które okazały się istotne statystycznie i miały dodatni wpływ na cenę były obecność miejsca parkingowego lub windy.

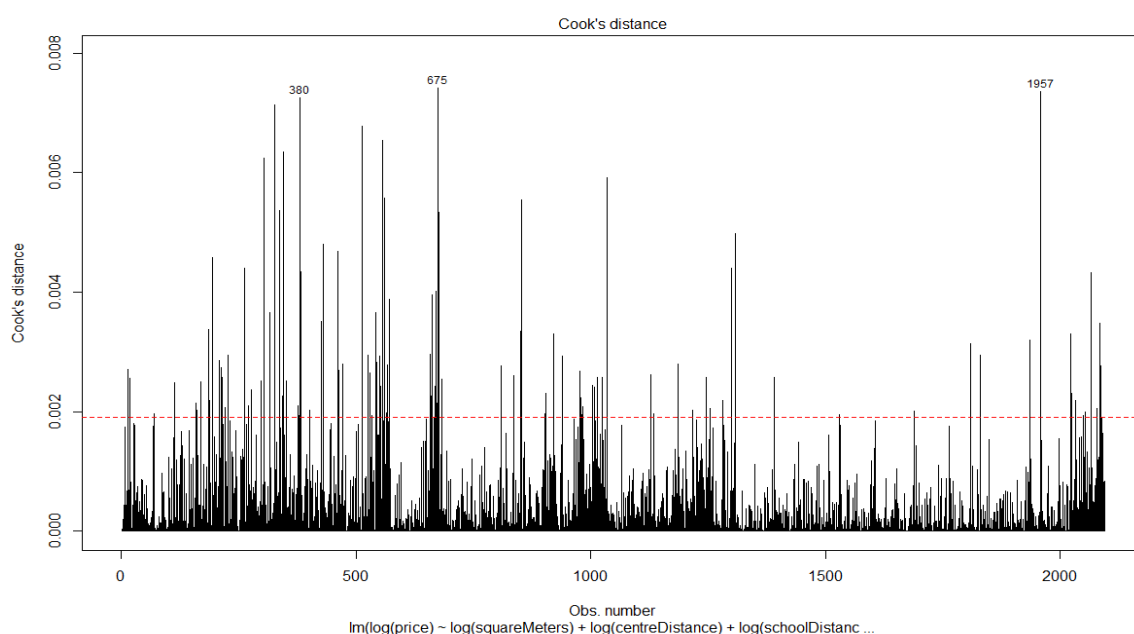
**H9:** Hipoteza 9. została jednoznacznie potwierdzona przez wyniki modelu: wiek mieszkania jest ujemnie skorelowany z ceną – im starsze mieszkanie, tym średnio tańsze.

**H10:** Wysoki i istotny statystycznie współczynnik przy zmiennej *stolica* potwierdza hipotezę 10., że mieszkania w stolicy są średnio droższe niż mieszkania w innych miastach. Jednakże efekt mógł zostać przeszacowany i należałoby uwzględnić więcej interakcji ze zmienną *stolica* lub zrobić oddzielny model tylko dla mieszkań w Warszawie, gdyż mogą charakteryzować się one zgoła innymi cechami od nieruchomości w mniejszych miastach.

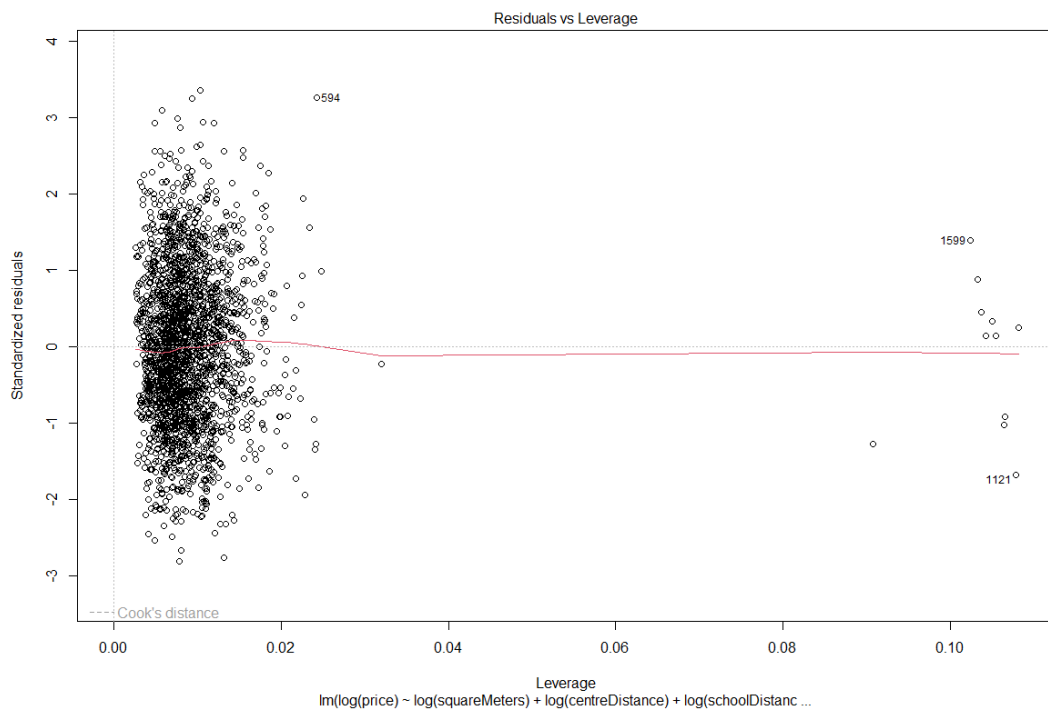
## 9. Problemy z obserwacjami

W modelowaniu ekonometrycznym ważna jest identyfikacja obserwacji nietypowych czy błędnych. Określenie nietypowe odnosi się do obserwacji, które mają unikalne cechy na tle zbioru danych. Natomiast obserwacje błędne to te, których powstania nie da się wyjaśnić w ramach teoretycznego modelu, a zatem najczęściej powstają one podczas niepoprawnego przepisywania czy gromadzenia danych. Zarówno jedno jak i drugie mogą negatywnie wpływać na estymację modelu. Oddziaływanie zmiennych nietypowych może być różne, w zależności od ich położenia w stosunku do linii regresji

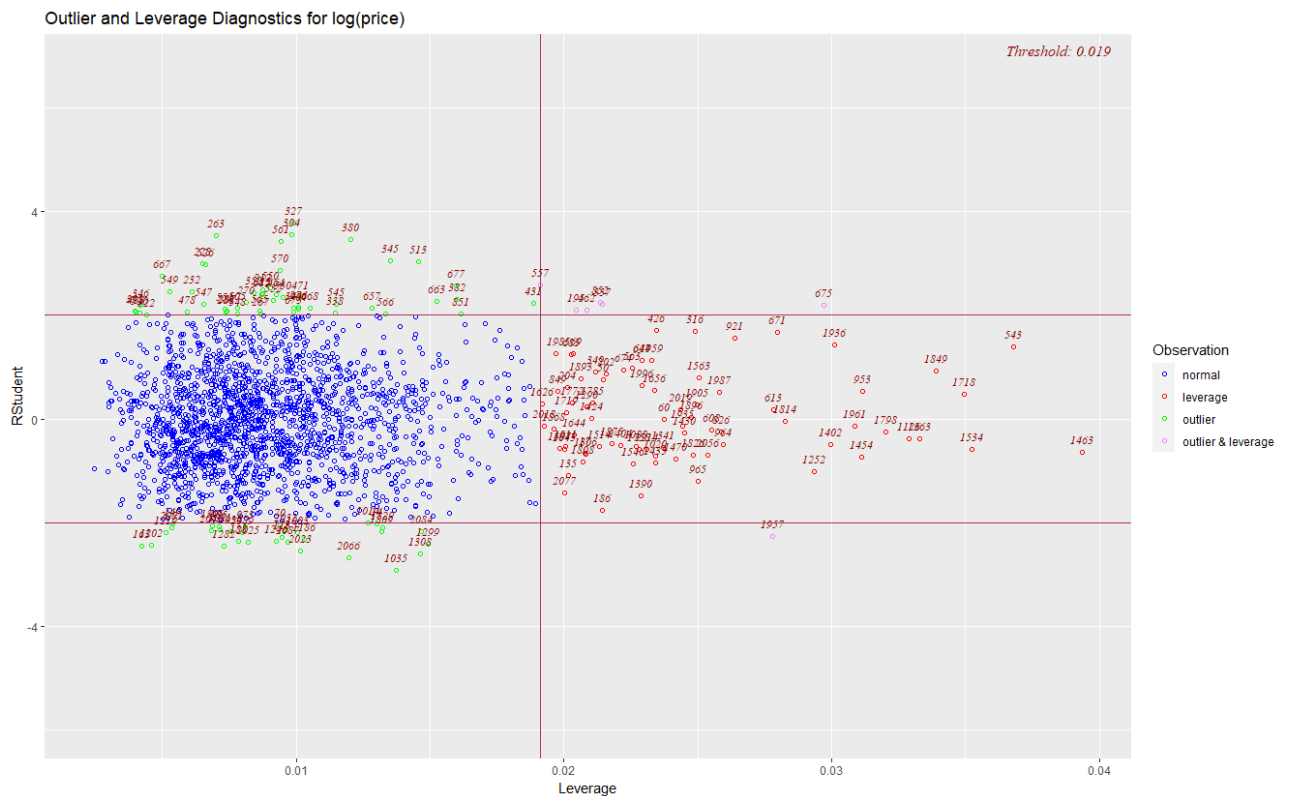
Aby zidentyfikować takie zmienne w swoim modelu, przeanalizowałam reszty i dźwignie przy pomocy wykresów oraz posługując się odpowiednim wskaźnikiem – odległością Cook'a.



Obserwacjami, które charakteryzowały się największymi odległościami Cook'a okazały się być obserwacje o numerach 675, 1957 oraz 380. To sugeruje, że mogły mieć one znaczący wpływ na regresję.



Obserwacjami o szczególnie wysokiej wartości dźwigni były m.in. obserwacje 1121, 1463, 1500 czy 543. Znajdowały się one daleko od chmury pozostałych punktów, ale jednocześnie relatywnie blisko linii regresji. Można więc je było zaklasyfikować jako *good leverage points*.



Znaczna część obserwacji w modelu to obserwacje normalne/typowe. Niektóre z przypadków można by natomiast uznać za *vertical outliers*, czyli te punkty, które nie wpływają na nachylenie krzywej regresji, tylko jej wysokość. Takimi obserwacjami były np. 327, 1035 czy właśnie wspomniana wcześniej obserwacja 380. Charakteryzowały się one małą dźwignią, ale dużą resztą.

Biorąc pod uwagę jednocześnie trzy warunki pod uwagę:  $dźwignia > 2k/n$ ,  $|std.errors| > 2$  oraz odległość Cook'a  $> 4/n$ , zidentyfikowałam 6 obserwacji nietypowych, które można uznać za *bad leverage points* (Tabela 9.). Znajdują się one daleko od chmury punktów i znacząco wpłynęły na estymację parametrów, gdyż mają jednocześnie dużą dźwignię i duże reszty.

Obserwacje 195, 337, 462 i 852 to bardzo stare mieszkania bez dodatkowych udogodnień takich jak winda czy miejsce parkingowe, ale jednocześnie o dużym metrażu i bardzo wysokiej, zastanawiającej cenie. Jednak są to kamienice, które mogą mieć znaczącą wartość historyczną czy znajdować się w uważanej za prestiżową lokalizacji. Konieczne mogłoby tutaj być wprowadzenie do modelu jeszcze jednej interakcji: *tenement\*age*, gdyż wzrost lat w przypadku kamienic nie musi wcale przekładać się na spadek ich ceny, a wręcz przeciwnie.

Rekordowo, jak na inne mieszkania o podobnych warunkach, wysoką ceną powyżej 2 milionów charakteryzuje się obserwacja nr 675. Prawdopodobnie bardzo duży wpływ ma tutaj fakt, że jest to mieszkanie w apartamentowcu, być może w prestiżowej lokalizacji. Wartościowym byłoby spojrzenie na ceny nieruchomości w sąsiedztwie tego mieszkania.

Ostatnia nietypowa obserwacja, numer 1957 to mieszkanie w bloku, prawdopodobnie wybudowane ponad 40 lat temu z materiałów niższej jakości, bez dodatkowych udogodnień i o średniej powierzchni. Ma ono jednak relatywnie wysoką cenę, ale może znajdować się w drogiej dzielnicy Warszawy czy w pobliżu metra, co nie zostało uwzględnione w modelu.



| Nr                 | 195     | 337     | 462     | 675     | 852    | 1957   |
|--------------------|---------|---------|---------|---------|--------|--------|
| price              | 1049000 | 1950000 | 1950000 | 2190000 | 720000 | 725000 |
| centreDistance     | 140     | 590     | 710     | 270     | 180    | 630    |
| squareMeters       | 72.13   | 150.00  | 150.00  | 95.89   | 46.59  | 67.60  |
| schoolDistance     | 245     | 291     | 228     | 136     | 312    | 243    |
| clinicDistance     | 1090    | 528     | 295     | 672     | 330    | 131    |
| restaurantDistance | 60      | 22      | 106     | 70      | 12     | 59     |
| collegeDistance    | 307     | 424     | 695     | 531     | 301    | 229    |
| buildingMaterial   | 1       | 1       | 1       | 1       | 1      | 0      |
| Condition          | 1       | 0       | 0       | 1       | 0      | 1      |
| hasParkingSpace    | 0       | 0       | 1       | 1       | 0      | 0      |
| hasElevator        | 0       | 0       | 0       | 1       | 0      | 0      |
| Top_floor          | 0       | 0       | 0       | 0       | 1      | 0      |
| Age                | 86      | 95      | 95      | 1       | 126    | 44     |
| apartmentB         | 0       | 0       | 0       | 1       | 0      | 0      |
| Tenement           | 1       | 1       | 1       | 0       | 1      | 0      |
| stolica            | 0       | 0       | 0       | 0       | 0      | 1      |

**Tabela 9.** Obserwacje nietypowe.

## 10. Wnioski

W niniejszym raporcie skoncentrowałam się na analizie mikroekonomicznej rynku nieruchomości wtórnych w największych miastach Polski. Celem badania było zrozumienie wpływu różnych czynników, takich jak charakterystyka nieruchomości, lokalizacja czy udogodnienia, na ceny mieszkań.

Literatura przedstawiała różnorodne metody analizy cen mieszkań, z naciskiem na modele hedoniczne. Zastosowanie tych modeli pozwalało na jednoczesne uwzględnienie wielu zmiennych dotyczących konkretnych cech mieszkań. Istotne były także badania analizujące wpływ instytucji publicznych, placówek edukacyjnych, jakości budownictwa czy różnic cenowych w zależności od obszarów miast. Weryfikacja hipotez postawionych na podstawie literatury oraz własnych przypuszczeń pozwoliła mi lepiej zrozumieć specyfikę polskiego rynku nieruchomości.

Model regresji liniowej wyjaśnił 73.6% zmienności cen mieszkań. Z interpretacji współczynników dla poszczególnych zmiennych można wyciągnąć znaczące wnioski. Potwierdziła się m.in. istotna statystycznie dodatnia zależność ceny mieszkania od jego

powierzchni oraz standardu i wysokiej jakości materiałów budowlanych. Hipotezy o wpływie liczby pokoi czy własności mieszkania na jego cenę zostały odrzucone, gdyż te zmienne okazały się statystycznie nieistotne. Wzrost wieku, odległości od restauracji, szkoły wyższej i przychodni przekładał się na spadek cen mieszkań. Jeśli chodzi o udogodnienia, to jedynie posiadanie miejsca parkingowego lub windy okazało się istotne statystycznie i miało dodatni wpływ na cenę mieszkania. Istotny statystycznie współczynnik przy zmiennej stolica potwierdza, że mieszkania w stolicy są średnio droższe niż mieszkania w innych miastach, choć efekt ten może być przeszacowany i prawdopodobnie należałoby przeanalizować przypadek Warszawy odrębnie.

Model ma pewne ograniczenia - dalsze badania mogłyby zawrzeć więcej interakcji, szczegółowe analizy dla poszczególnych miast, a nawet dzielnic, czy uwzględnić wpływ komunikacji miejskiej lub wartość sąsiadujących z mieszkaniami nieruchomości.

## Bibliografia

1. Bourne Larry Stuart. *The geography of housing*. Londyn, 1981.
2. D'Acci Luca. *Quality of urban area, distance from city centre, and housing value. Case study on real estate values in Turin*. *Cities*. Vol 91, 71-92, ISSN 0264-2751. 2019.
3. Lancaster Kelvin J. *A New Approach to Consumer Theory*. *Journal of Political Economy*, Vol. 74, No. 2, 132-157. 1966.
4. Owusu-Ansah Anthony. *A review of hedonic pricing models in housing research*. *Journal of International Real Estate and Construction Studies*. ISSN: 2153-6813. Vol. 1, Nr 1. 2011.
5. Pricewaterhouse Coopers. *Skąd ten boom? Zmiany na rynku mieszkaniowym w Polsce*. 2021.
6. Wittowsky Dirk, Hoekveld Josje, Welsch Janina, Steier Michael. *Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany*. *Urban, Planning and Transport Research*, 8:1, 44-70, DOI: 10.1080/21650020.2019.1704429. 2020.
7. Tomczyk Emilia, Widłak Marta. *Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy*. *Bank i Kredyt* 41 (1), 99–128. 2010.
8. Adair Alastair, McGreal Stanley, Smyth Austin, Cooper James, Ryley Tim *House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area*. *Housing Studies*, 15:5, 699-716, DOI: 10.1080/02673030050134565. 2000.
9. Wen Haizhen, Xiao Yue, Zhang Ling. *School district, education quality, and housing price: Evidence from a natural experiment in Hangzhou, China*. *Cities*, Vol. 66, 72-80. 2017.
10. Małecka Katarzyna. *Badanie cen na rynku nieruchomości mieszkaniowych w Łodzi*. *Świat Nieruchomości*, 3(69), 52-55. 2009.