

# **Wycena nieruchomości w dzielnicy Xindian, Nowe Tajpej, Tajwan**

Yauheni Semianiuk

Warszawa 2023

## Spis treści

Wstęp .....	2
Rozdział I: Opis i analiza danych .....	3
Rozdział II: Model KMNK.....	9
Rozdział III: Informacja a priori .....	10
Rozdział IV: Informacja a posteriori .....	11
Rozdział V: Czynniki Bayesa. HPDI.....	13
Bibliografia .....	15
Spis rysunków.....	16

## Wstęp

Cena nieruchomości stanowi jedną z najważniejszych wskaźników statystycznych, kształtując rozwój demograficzny i ekonomiczny zarówno rynku lokalnego, jak i globalnego kraju. Niska cena mieszkania stymuluje migrację zewnętrzną i wewnętrzną, co z kolei kumuluje dodatkowe zasoby ludnościowe w danym obszarze geograficznym. Mówiąc o wpływie demograficznym, cena mieszkania jest faktorem kierującym decyzje o osiedlaniu, mianowicie dla osób w młodszym wieku produkcyjnym.<sup>1</sup> Przyczynia to do restrukturyzacji rozkładu wiekowego na terenie kraju. W końcu cena mieszkania oddziałuje na koszty prowadzenia działalności gospodarczej, wspierając tym samym ulokowanie przedsiębiorstw w obszarach o niskich kosztach pomieszczeń komercyjnych i zabudowy.

Rynek nieruchomości na Tajwanie jest ciekawy przede wszystkim ze względu na szybki wzrost cen nieruchomości. W Nowe Tajpej w latach 2002 – 2014 wartość przeciętnej nieruchomości potroiła się<sup>2</sup>. Przyczyną takiego szybkiego wzrostu było aktywne rozwinięcie gospodarcze kraju oraz spekulacje inwestycyjne. Jednak po 2014 roku rząd Tajwański wprowadził wysokie podatki, co ukroiło spekulacje na tajpejskim rynku nieruchomości.<sup>3</sup> Celem niniejszej pracy z kolei jest ocena czynników wpływających na kształtowanie ceny jednego pinga (1 ping = 3.3 metra kwadratowego) mieszkania w Nowe Tajpej w najbardziej intensywnym okresie inwestycyjnym w latach 2012 – 2013.

---

<sup>1</sup> Peng C., The long- and short-run influences of housing prices on migration, s. 257.

<sup>2</sup> <https://www.rp.pl/nieruchomosci/art4170291-tajwan-mieszkania-sa-tansze> , dostęp 04.02.2023.

<sup>3</sup> <https://biznes.interia.pl/nieruchomosci/news-wysokie-podatki-ukrocily-spekulacje-na-tajwanskim-ryнку-nier,nId,4039159> , dostęp 04.02.2023.

## Rozdział I: Opis i analiza danych

Do analizy wykorzystany będzie zbiór danych zawierający 414 transakcji zawartych w od sierpnia 2012 do grudnia 2013<sup>4</sup>. Zbiór danych zawiera 6 zmiennych objaśniających:

- year\_month – data transakcji w postaci rok-miesiąc,
- age – wiek budynku w latach,
- dist\_to\_mrt – dystans do najbliższej stacji metra,
- no\_stores – liczba sklepów ogólnospożywczych w dzielnicy mieszkalnej,
- long – długość geograficzna,
- lat – szerokość geograficzna

Natomiast nie wszystkie z przedstawionych zmiennych zostaną użyte w bieżącej pracy. Długość i szerokość geograficzna nie będą wykorzystane z powodu tego, że nie mają one bezpośredniego wpływu na kształtowanie ceny nieruchomości, a raczej zawierają w sobie osobliwości ekonomiczne, socjalne oraz tłoczone przedstawionych mieszkań. Pozostałe zmienne zostaną szczególnie przeanalizowane.

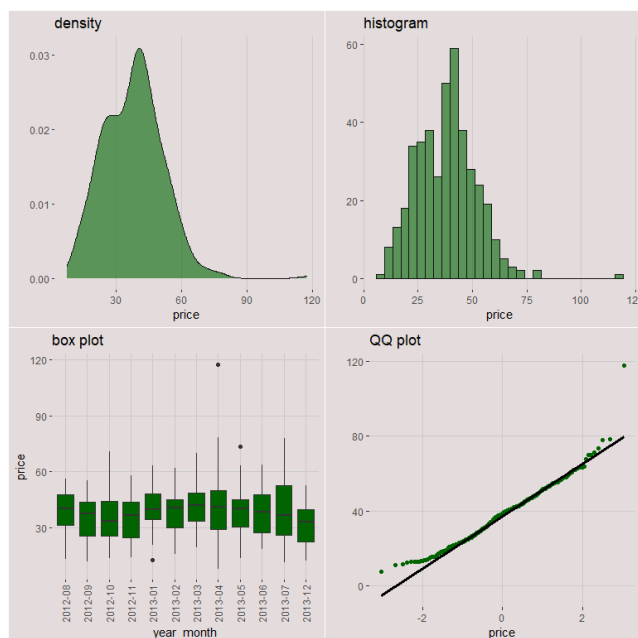
Analiza zaczyna się od zmiennej objaśnianej price. Z rysunku 1.1 wynika, że jest to zmienna o dość scentralizowanym rozkładzie z długim prawym ogonem, który odzwierciedla występowanie bardzo drogich mieszkań (wartości odstających). Ponieważ kolejnym krokiem analizy jest zbudowanie modelu liniowego oszacowanego za pomocą metody najmniejszych kwadratów, może to mieć znaczący wpływ na wartości współczynników. Występowanie wartości odstających widać także na wykresie QQ, porównującym kwantyle rozkładu zmiennej price z kwantylami rozkładu normalnego. Z ostatniego wynika także że lewe ogony są bardzo „grube”. Jednak dla celów bieżącej pracy ograniczymy się tym, że zostaną usunięte obserwacje, dla których cena 1 pingu mieszkania wynosiła więcej niż 100.

Na rysunku 1.1 pokazana również zależność ceny 1 pingu od daty transakcji. Jednak z wykresu pudełkowego wynika, że średnia wartość ceny prawie się nie zmienia dla poszczególnych miesięcy. Może to oznaczać że szczyt wartości cen mieszkań był osiągnięty mianowicie w tym okresie. W kolejnych etapach analizy, zmienna year\_month nie będzie używana.

---

<sup>4</sup> <https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set> , dostęp 04.02.2023.

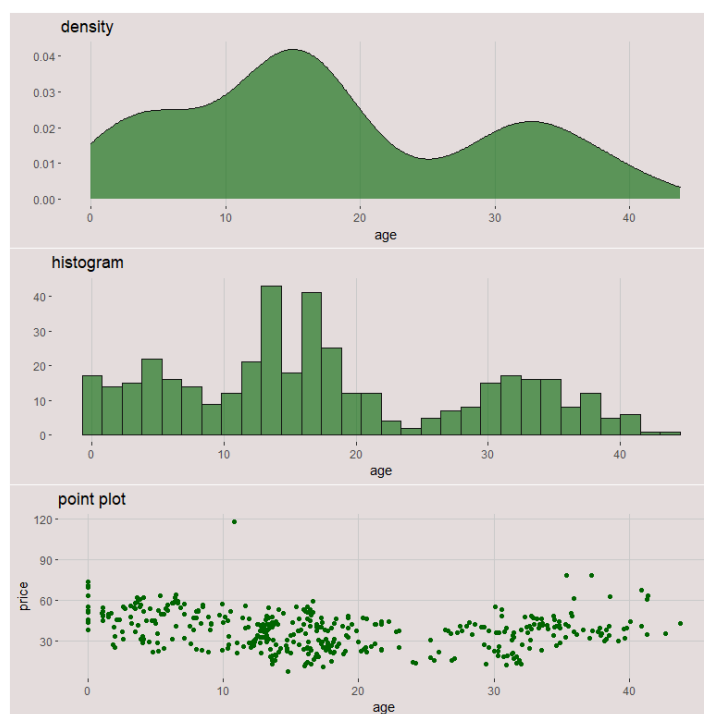
Rysunek 1. Zmienna objaśniana price



Źródło: opracowanie własne

Następnie przeanalizowana zmienna age. Po pierwsze, jest to zmienna o złożonym rozkładzie dwumodalnym (dominanty przy wieku wynoszącym 15 i 32 lata). Fakt ten może odzwierciedlać 2 boomy wędrownicze i budownicze. Po drugie, zależność pomiędzy zmienną objaśnianą i zmienną age nie wygląda jako zależność liniowa. Przypomina ona wygładzoną zależność kwadratową. Może to świadczyć o większej cenie nieruchomości w centrum miasta (przecież została centrum był zagospodarowany wcześniej). Należałoby oszacować kilka wielomianów zmiennej age w modelu KMNK.

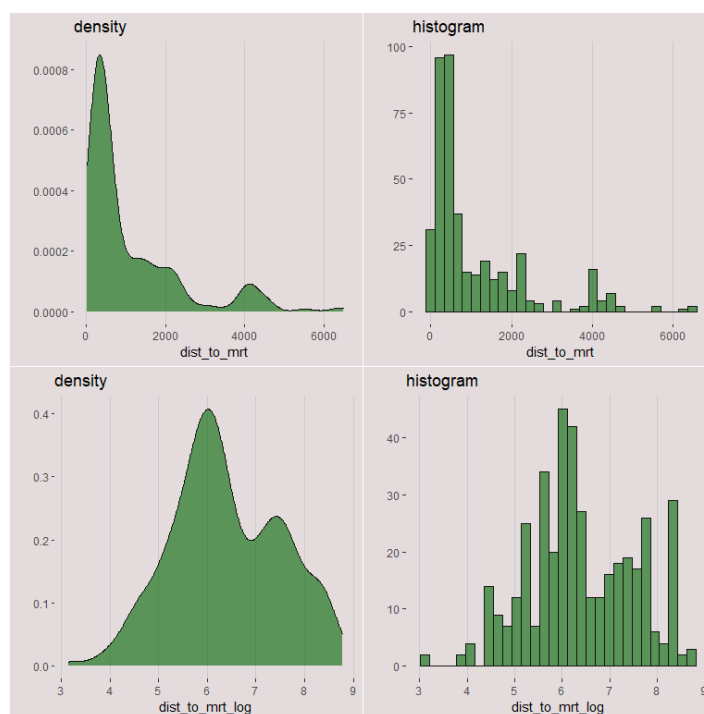
Rysunek 2. Rozkład i wykres punktowy zmiennej age



Źródło: opracowanie własne

Następną z kolei jest zmienna `dist_to_mrt`. Szczególna uwaga będzie dotyczyła postaci zmiennej, a mianowicie czy potrzebne jest zastosowanie transformacji logarytmicznej. Rysunek 3 wskazuje mocną prawoskosność rozkładu oryginalnej zmiennej `dist_to_mrt`. Z tego powodu i powstało pytanie, czy należy zastosować odpowiednią transformację. Jednak rozkład zmiennej po logarytmizacji nie przekonywał do podejmowania takiej decyzji.

Rysunek 3. Rozkład zmiennej dist\_to\_mrt



Źródło: opracowanie własne

Z tego powodu kluczowym było przedstawienie relacji zachodzącej pomiędzy zmiennej objaśnianej a każdej z wersji zmiennej dist\_to\_mrt (rysunek 4). Okazało się, że relacja przyjmuje charakter bardziej liniowy w przypadku zmiennej dist\_to\_mrt\_log, dlatego zostaje ona wybrana jako wersja końcowa.

Rysunek 4. Zależność zmiennej price od zmiennej dist\_to\_mrt

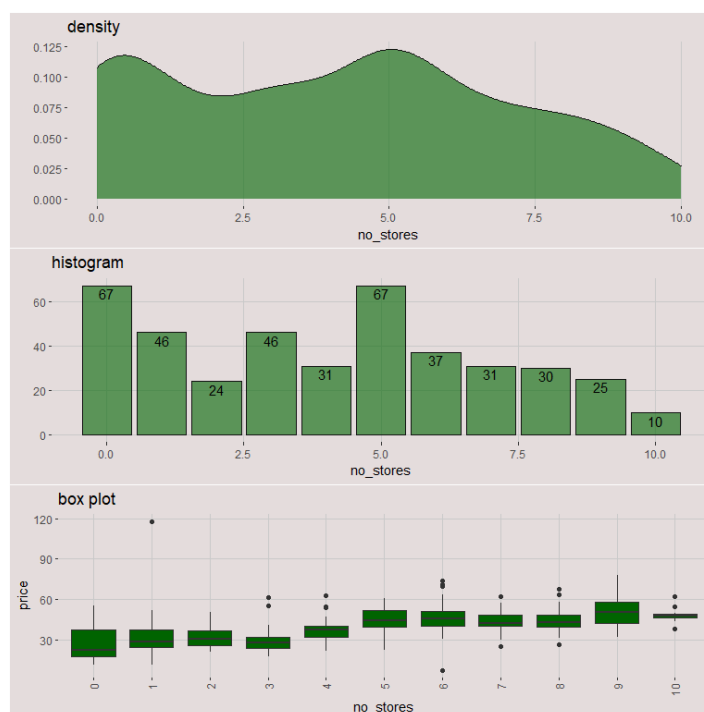


Źródło: opracowanie własne

Na koniec analizie uległa zmienna `no_stores`. Jej rozkład, jak i w przypadku zmiennej `age`, jest rozkładem dwumodalnym. Jednak jest on bardziej wypłaszczony i przypomina rozkład jednostajny. Wykres pudełkowy wskazuje na zachodzenie trendu liniowego, co zostawia możliwości do traktowania tej zmiennej jako istotnie wpływającej na wartości ceny 1 pinga. Należy także zauważyć, że można traktować zmienną `no_stores` jako zmienną kategoryjalną o 9 poziomach.



Rysunek 5. Rozkład zmiennej dist\_to\_mrt



Źródło: opracowanie własne

Po uwzględnieniu informacji zdobytych podczas wstępnej analizy danych oceniona została korelacja zachodząca pomiędzy zmiennymi. W wyniku okazało się, że mamy dwie dość mocne zmienne objaśniające, gdyż zmienna age jest dość słabo skorelowana ze zmienną price. Co więcej, raczej nie zauważalne jest zjawisko współliniowości. W ten sposób, możliwe będzie otrzymanie stabilnych oszacowań KMNK.

Rysunek 6. Korelacja zmiennych

variable	price	age	no_stores	dist_to_mrt_log
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
price	1	-0.21	0.61	-0.76
age	-0.21	1	0.05	0.07
no_stores	0.61	0.05	1	-0.69
dist_to_mrt_log	-0.76	0.07	-0.69	1

Źródło: opracowanie własne

## Rozdział II: Model KMNK

Kolejna część pracy dotyczyła zbudowania modelu liniowego oraz przeprowadzenie testów statystycznych. Na podstawie danych z rozdziału I oszacowany został model:

$$price_i = 82.47 - 0.21 \times age_i + 0.92 \times no\_stores_i - 7 \times dist\_to\_mrt\_log_i + \varepsilon_i$$

Rysunek 7. Model KMNK

```
Call:
lm(formula = price ~ age + no_stores + dist_to_mrt_log, data = df_sel)

Residuals:
    Min       1Q   Median       3Q      Max
-35.473  -4.990  -0.861   3.930  34.968

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   82.46920    3.72310   22.151 < 0.0000000000000002 ***
age          -0.20724    0.03527   -5.876  0.00000000872 ***
no_stores      0.91754    0.18794    4.882  0.00000150748 ***
dist_to_mrt_log -7.00223    0.49461  -14.157 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.075 on 409 degrees of freedom
Multiple R-squared:  0.6197,    Adjusted R-squared:  0.6169
F-statistic: 222.1 on 3 and 409 DF,  p-value: < 0.00000000000000022
```

Źródło: opracowanie własne

Wyniki testów Durбина-Watsona (wartość  $p = 0.088$ ) oraz Ljunga-Boxa (wartość  $p = 0.396$ ) wskazywały na niezachodzenie zjawiska autokorelacji zmiennej objaśnianej. Testy heteroskedastyczności Goldfelda-Quandt ( $p=0.217$ ) oraz Breuscha-Pagana ( $=0.058$ ) na poziomie istotności  $\alpha = 0.05$  świadczyły o tym, że wariancja reszt jest stała. Normalność rozkładu reszt nie jest wymagana dla próby z liczbą obserwacji większej niż 100, tabela korelacji oraz czynnik inflacji wariancji VIF mniejszy od 10 dla każdej ze zmiennych (rysunek 8) nie sugeruje współliniowości. Można więc przyjąć, że przy zadanym poziomie istotności specyfikacja modelu KMNK jest poprawna.

Rysunek 8. Czynnik inflacji wariancji

age	no_stores	dist_to_mrt_log
1.021587	1.935748	1.939782

Źródło: opracowanie własne

### Rozdział III: Informacja a priori

Za wartość wyrazu wolnego a priori ustalona wartość przeciętna 1 pinga mieszkania uzyskana z danych historycznych. Natomiast główne źródła wiedzy a priori stanowiły prace Wydziału Finansów Uniwersytetu Narodowego Kaohsiunga, raport CTBC Financial Holding oraz pracy Xu Zhanga, Qingshan Qiana i inn. W przypadku wpływu zmiennej *age* i *no\_stores*, przyjęte uśrednione wyniki z powyższych badań z małym odchyleniem standardowym. Powodem przyjęcia małego odchylenia jest to że, wyniki pochodzą z trzech zasadniczo różnych badań naśladujące różne cele, aczkolwiek wyniki się zgadzają. Inaczej stoi ze zmiennej *dist\_to\_mrt\_log*: istnieje dużo badań na temat wpływu lokalizacji na cenę mieszkań w Nowym Tajpej, aczkolwiek żadne z nich nie wykazuje główne cechy charakterystyczne badanych lokalizacji. Co więcej, liczne raporty badające czynniki wpływające na boom cen mieszkań w Nowym Tajpej wśród najbardziej istotnych przyczyn wskazują intensywnie rozwijającą się sieć transportu publicznego. Niemniej jednak, nie znaleziono żadnego słusznego badania na temat wpływu bliskości metra. Dlatego wartość a priori została ustawiona na wysoką o dużym odchyleniu standardowym. Z racji braku jednoznacznych informacji na temat korelacji pomiędzy czynnikami (w szczególności ich siły), kowariancje pomiędzy parametrami zostaną ustawione na poziom zerowy. Liczba stopni swobody a priori zostanie ustawiona na 100. Ponieważ zaufane do wiedzy a priori jest dość duże, założona wartość oczekiwana a priori odchylenia standardowego reszt (*s*) wynosi 0.15, zatem jego precyzji — 44.44. Zbiorczo, przyjętą wiedzę a priori prezentuje rysunek 9.

Rysunek 9. Wartości parametrów a priori

```
> Beta.prior
[1] 37.790 -0.050 1.106 -8.000
> sm2.prior
[1] 44.44
> U.prior
      [,1] [,2] [,3] [,4]
[1,] 25 0.0000 0.000000 0
[2,] 0 0.0001 0.000000 0
[3,] 0 0.0000 0.000625 0
[4,] 0 0.0000 0.000000 9
> v.prior
[1] 100
```

Źródło: opracowanie własne

## Rozdział IV: Informacja a posteriori

W wyniku estymacji otrzymano wartości oczekiwane parametrów a posteriori. Rysunek 10 pokazuje jak wyglądały współczynniki otrzymane po estymacji KNMK z wiedzą a priori oraz wiedzą po połączeniu obu źródeł. Widać, że założenie a priori o znacznie mniejszej stałej nie przesunęło jej zbyt dużo. Przyczyną tego jest dość mała wariancja a priori pozostałych zmiennych, co nie pozwalała na wymaganą elastyczność parametrów. Co więcej, należy pamiętać że wartość stałej a priori była wyznaczona z surowych danych, natomiast średnia z modelu wynosząca 82.5 jest wartością średnią „oczyszczoną” od wpływu pozostałych zmiennych (oznacza średnią wartość 1 pinga gdy pozostałe zmienne są równi zero).

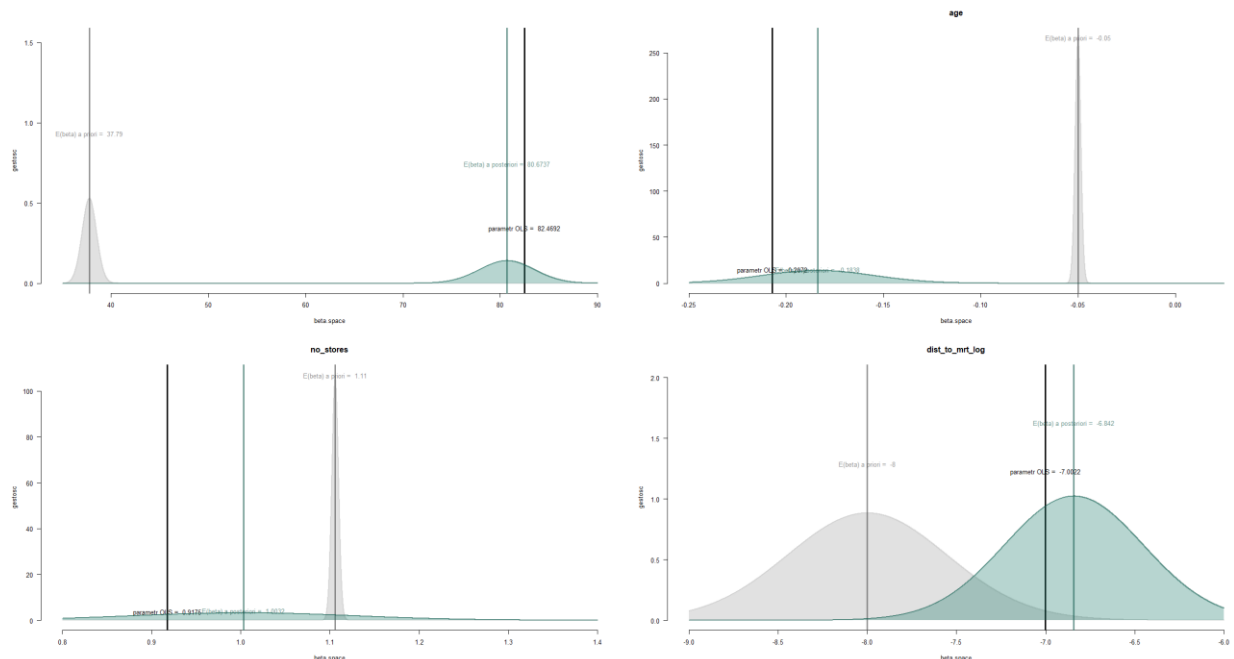
Rysunek 10. Wartości parametrów a posteriori, a priori oraz oszacowane z modelu

variable	posterior	model	prior
<chr>	<dbl>	<dbl>	<dbl>
intersept	80.7	82.5	37.8
age	-0.184	-0.207	-0.05
no_stores	1.00	0.918	1.11
dist_to_mrt_log	-6.84	-7.00	-8

Źródło: opracowanie własne

Z kolei zestaw wykresów 4.2 prezentuje rozkłady brzegowe oszacowań parametrów a posteriori, gdyż pierwszy wykres dotyczy wyrazu wolnego.

Rysunek 11. Rozkłady brzegowe parametrów



Źródło: opracowanie własne

Warto wyjaśnić niektóre fakty wynikające z wykresów rozkładów brzegowych:

- Wartości oczekiwane a posteriori dla wszystkich zmiennych znajdują się pomiędzy oszacowaniem wartości oczekiwanej a priori a oszacowaniem KMNK, co jest potwierdzeniem małej korelacji między parametrami.
- Pomimo przyjętych małej wariancji a priori oraz dużej liczby stopni swobody a priori, wartości oczekiwane a posteriori (nie licząc zmiennej `no_stores`) są bliżej oszacowań KMNK niż oszacowań a priori.
- Wartości oczekiwane parametrów a priori, a posteriori oraz parametry KMNK są spójne co do kierunku oddziaływania.

## Rozdział V: Czynniki Bayesa. HPDI

W formule obliczania czynników Bayesa używano logarytmu zmiennych oraz w przypadku funkcji gamma użyta była funkcja  $\min(\bar{v}/2, 171)$ . Jest to niezbędne aby rozwiązać problemy wynikające z dokładności numerycznej gdyż ustawiona duża liczba stopni swobody (100) oraz liczba obserwacji (414) są duże, co powoduje że  $\Gamma\left(\frac{514}{2}\right) = \infty$ .

Wyniki obliczeń są pokazane na rysunku 5.1

Rysunek 12. Czynniki Bayesa dla 3 modeli

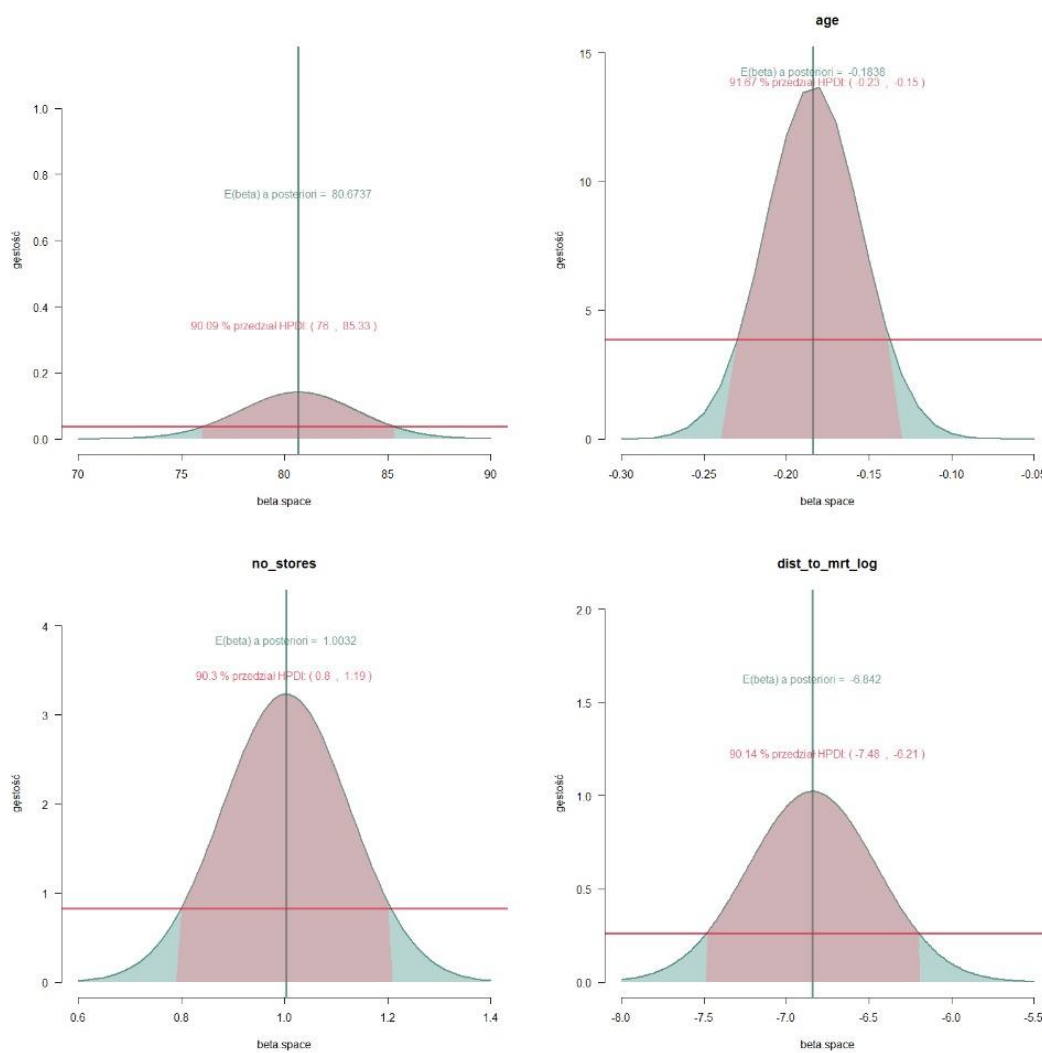
variable	BF
<chr>	<dbl>
age	9.07e-299
no_stores	2.24e-300
dist_to_mrt_log	1.22e-253

Źródło: opracowanie własne

Czynniki Bayesa są znacznie mniejsze niż 1 (de facto, są równe zero), co oznacza, że preferowane są modele  $M_2$ , czyli modele w których pozbawiono jednej ze zmiennych. Należy jednak zwrócić uwagę na fakt, że wyliczone czynniki Bayesa dotyczą modeli zagnieżdżonych, w których usunięto tylko jedną zmienną objaśniającą. Powyższe wyniki nie świadczą o tym, że wszystkie zmienne są nieistotne. W tym przypadku należałoby sprawdzić przypadki kiedy usunięte byłoby 2 zmienne naraz.

Wykres HPDI dla stałej przy 90% poziomie ufności wskazują, że stała powinna przyjmować wartości dodatnie z przedziału (76; 85.33). Wykresy HPDI dla zmiennych age, no\_stores oraz dist\_to\_mrt\_log pokazują, że przy 90% poziomie ufności (tzn. w 90% przypadkach) nie przyjmują te zmienne wartości zerowe, co oznacza, że są istotne w kształtowaniu cen mieszkań.

Rysunek 13. Wykres HDPI dla zmiennych objaśniających



Źródło: opracowanie własne

Wnioski pochodzące z analizy przedziałów HPDI są odmienne od tych, prezentowanych przez czynniki Bayesa. Niemniej jednak, wnioski po analizie przedziałów o największej gęstości a posteriori są bardziej zgodne z innymi pracami w tym temacie, oraz z własnymi oszacowaniami KMNK. Dlatego ostatecznie przyjęto że zmienne age, no\_stores oraz dist\_to\_mrt\_log mają istotny wpływ na cenę 1 pinga mieszkania w Nowym Tajpej, Tajwan.

## Bibliografia

1. Miller N., Peng L., *Exploring metropolitan housing price volatility*, Journal of Real Estate , Financial Economics, 2006, str. 5-18.
2. Peng C., *The long- and short-run influences of housing prices on migration*, National Taipei University, Tajpej 2019, str. 253-262.
3. Lin C, *A Quick Overview of Taiwan Housing Market*, CTBC Financial Holding, Tajpej 2019.
4. Qian Q., *How do mass rapid transit and convenience stores affect housing prices*, ABE International Business College, Iloilo 2022.
5. Wang W., Tsai I., *House age and housing prices: a viewpoint of the optimal time for land redevelopment*, International Journal of Strategic Property Management, Wyd. 26 Nu. 3, 2022, str. 172–187.
6. Yan Y., Song Y., Zhang, W., *Method for housing price forecasting based on TEI@I methodology*, Systems Engineering—Theory&Practice, Wyd. 27 Nu. 7, 2007 str. 1–9.
7. Zhang X. i inn, *Do Urban Rail Transit Facilities Affect Housing Prices? Evidence from China*, Southeast University, Nanjing 2016.
8. <https://biznes.interia.pl/nieruchomosci/news-wysokie-podatki-ukrocily-spekulacje-na-tajwanskim-rynku-nier,nId,4039159>, (dostęp 04.02.2023).
9. <https://www.rp.pl/nieruchomosci/art4170291-tajwan-mieszkania-sa-tansze>, (dostęp 04.02.2023).
10. <https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set>, (dostęp 04.02.2023).



## Spis rysunków

Rysunek 1. Zmienna objaśniana price .....	4
Rysunek 2. Rozkład i wykres punktowy zmiennej age .....	5
Rysunek 3. Rozkład zmiennej dist_to_mrt .....	6
Rysunek 4. Zależność zmiennej price od zmiennej dist_to_mrt .....	7
Rysunek 5. Rozkład zmiennej dist_to_mrt .....	8
Rysunek 6. Korelacja zmiennych .....	8
Rysunek 7. Model KMNK .....	9
Rysunek 8. Czynniki inflacji wariancji .....	9
Rysunek 9. Wartości parametrów a priori .....	10
Rysunek 11. Rozkłady brzegowe parametrów .....	11
Rysunek 12. Czynniki Bayesa dla 3 modeli .....	13
Rysunek 13. Wykres HDPI dla zmiennych objaśniających .....	14