**Temat**: Analiza danych panelowych - zmiany cen mieszkań w 16 polskich województwach w latach 2010-2023.

**Cel badania**: Zbadanie wpływu wybranych zmiennych na ceny mieszkań w latach 2010-2023.

# Dane

**Źródło**

Główny Urząd Statystyczny – Bank Danych Lokalnych.

**Zmienna objaśniana**

Cena mieszkania za jeden m2 na rynku pierwotnym.

**Zmienne objaśniające**

* przeciętne miesięczne wynagrodzenie brutto,
* mieszkania oddane do użytkowania przeznaczone na sprzedaż lub wynajem,
* gęstość zaludnienia powierzchni zabudowanej i zurbanizowanej (osoby/km2),
* Produkt Krajowy Brutto na 1 mieszkańca województwa,
* nowo utworzone miejsca pracy,
* nakłady inwestycyjne na 1 mieszkańca w cenach bieżących,
* przeciętne miesięczne wydatki na 1 mieszkańca.

# Opis danych

* Cena mieszkania za 1m2 na rynku pierwotnym - średnia cena w PLN za 1m2 lokali mieszkalnych sprzedanych w ramach transakcji rynkowych, z podziałem na województwa. Cena obejmuje wszystkie składniki nieruchomości objętej transakcją. Ceny lokali mieszkalnych sprzedanych na rynku pierwotnym zawierają podatek VAT.
* Przeciętne miesięczne wynagrodzenie brutto – średnia kwota zarobków w PLN przed potrąceniem podatków i składek na ubezpieczenia społeczne, którą otrzymuje pracownik w danym okresie, z podziałem na województwa.
* Mieszkania oddane do użytkowania przeznaczone na sprzedaż lub wynajem – ilość nieruchomości, w których zakończono prace budowlane, uzyskano pozwolenie na użytkowanie i oferowane są na rynku w celu sprzedaży lub wynajmu, zarówno w pełni wykończone, jak i w standardzie deweloperskim; z podziałem na województwa.
* Gęstość zaludnienia powierzchni zabudowanej i zurbanizowanej (osoby/km2) – liczba mieszkańców przypadających na 1 km2 obszarów, które zostały przekształcone w przestrzeń zabudowaną lub miejską, z podziałem na województwa.
* Produkt krajowy brutto na 1 mieszkańca województwa – wartość dóbr i usług wytworzonych w danym województwie w PLN, w przeliczeniu na jednego mieszkańca.
* Nowo utworzone miejsca pracy – miejsca pracy powstałe w wyniku zmian organizacyjnych, rozszerzenia lub zmiany profilu działalności oraz wszystkie miejsca pracy w jednostkach nowo powstałych, podane w tysiącach miejsc, z podziałem na województwa.
* Nakłady inwestycyjne na 1 mieszkańca w cenach bieżących – wartość wydatków przeznaczonych na inwestycje (takie jak infrastruktura, budownictwo, modernizacje) w przeliczeniu na jednego mieszkańca, uwzględniająca aktualne ceny rynkowe, w PLN, z podziałem na województwa.
* Przeciętne miesięczne wydatki na 1 mieszkańca – średnia wartość kosztów ponoszonych przez jednego mieszkańca w PLN, z podziałem na województwa. Wydatki obejmują wydatki na towary i usługi konsumpcyjne oraz pozostałe wydatki. Wydatki na towary i usługi konsumpcyjne przeznaczone są na zaspokojenie potrzeb gospodarstwa domowego. Obejmują one towary zakupione za gotówkę również przy użyciu karty płatniczej lub kredytowej, na kredyt, otrzymane bezpłatnie oraz spożycie naturalne (towary i usługi konsumpcyjne pobrane na potrzeby gospodarstwa domowego z działalności rolniczej bądź działalności gospodarczej na własny rachunek). Pozostałe wydatki obejmują m.in. kwoty przekazane innym gospodarstwom domowym i instytucjom niekomercyjnym, w tym dary; koszty zakwaterowania młodzieży i studentów uczących się poza domem; alimenty dla osób prywatnych; niektóre podatki, np.: od nieruchomości, od spadków i darowizn, od dochodów z własności, z wynajmu i sprzedaży nieruchomości; straty pieniężne.

# Koncepcja i metody analizy

**Koncepcja wykorzystanego modelu**

Model dwuczynnikowy z efektami stałymi jest narzędziem statystycznym, które umożliwia uwzględnienie zarówno efektów indywidualnych, jak i czasowych w analizie danych panelowych. W tym modelu kontrolujemy dla różnic między jednostkami (np. województwami), które mają stały wpływ na zmienną zależną, takim jak cechy specyficzne dla regionów (np. infrastruktura, gęstość zaludnienia) oraz dla czynników, które zmieniają się w czasie i mają wpływ na wszystkie jednostki w danym okresie, jak zmiany w polityce gospodarczej czy kryzysy finansowe. Dzięki zastosowaniu tego modelu, eliminujemy wpływ nieobserwowanych zmiennych, które mogą wprowadzać zakłócenia w analizie, poprawiając precyzję oszacowań. Model dwuczynnikowy jest szczególnie odpowiedni w kontekście analizy cen mieszkań, ponieważ pozwala uwzględnić różnorodność między regionami oraz wspólne czynniki makroekonomiczne, które wpływają na ceny nieruchomości w różnych latach.

**Sposób estymacji parametrów**

Sposób estymacji parametrów w modelu dwuczynnikowym z efektami stałymi polega na kontrolowaniu zarówno dla efektów indywidualnych, jak i czasowych, co pozwala na uwzględnienie różnic między jednostkami (np. województwami) oraz zmienności w czasie. Estymacja parametrów odbywa się przy użyciu estymatora wewnątrzgrupowego, który usuwa wpływ stałych efektów indywidualnych i czasowych, analizując zmiany w obrębie jednostek w czasie. W tym procesie, zmienne są transformowane przez odjęcie ich średnich, co pozwala na eliminację wpływu stałych efektów, a następnie estymacja współczynników regresji (β) odbywa się za pomocą klasycznego algorytmu najmniejszych kwadratów (MNK). Efekty indywidualne (αi) i czasowe (λt) są traktowane jako zmienne losowe, które podlegają estymacji, a całość modelu jest rozwiązywana przy użyciu odpowiednich procedur obliczeniowych, takich jak transformacja zmiennych i operacje na macierzach.

Obraz zawierający tekst, Czcionka, pismo odręczne, biały

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Opis estymatora wewnątrzgrupowego

Obraz zawierający tekst, Czcionka, typografia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Opis estymatora parametrów

**Testowanie hipotez**

Testowanie hipotez w modelu dwuczynnikowym z efektami stałymi polega na ocenie istotności poszczególnych zmiennych oraz efektów. Na początek przeprowadzamy testy **t-Studenta** dla współczynników regresji β, aby sprawdzić, czy poszczególne zmienne mają istotny wpływ na ceny mieszkań:

-> zmienna nie ma istotnego wpływu na ceny mieszkań

-> zmienna ma istotny wpływ na ceny mieszkań

W przypadku całkowitej istotności modelu wykonujemy test **F**:

i to suma kwadratów reszt w modelu bez i z efektami, a i ​ to liczba parametrów w obu modelach.

Redukcja modelu zaczyna się od ogólnego modelu z efektami stałymi indywidualnymi i czasowymi. Stopniowo usuwamy zmienne i efekty, które nie są istotne, bazując na wynikach testów t-Studenta i F, aż do uzyskania prostszego modelu, który wciąż wyjaśnia zmienność w cenach mieszkań.

**Koncepcja efektów krańcowych**

Efekty krańcowe w modelu dwuczynnikowym z efektami stałymi reprezentują wpływ zmiany jednej zmiennej objaśniającej na zmienną zależną, przy założeniu, że pozostałe zmienne pozostają niezmienne. W przypadku modelu dwuczynnikowego z efektami stałymi, efekty krańcowe są obliczane na podstawie współczynników regresji β, które reprezentują wpływ poszczególnych zmiennych na zmienną zależną, kontrolując dla efektów indywidualnych (efekty stałe związane z jednostkami) i czasowych (efekty związane z różnicami w czasie).

**Wybrane mierniki dopasowania modelu do danych**

**R² (Współczynnik determinacji)**

R² mierzy, jaka część zmienności zmiennej zależnej jest wyjaśniona przez model. Wartość R² waha się od 0 do 1, gdzie wartość 1 oznacza, że model wyjaśnia całą zmienność danych, a wartość 0 wskazuje, że model nie wyjaśnia żadnej zmienności. W przypadku modelu z efektami stałymi, R² wskazuje, jak dobrze model z efektami indywidualnymi (i czasowymi) wyjaśnia zmienność cen mieszkań w różnych regionach i latach. Wartość R² wyższa niż 0.9 świadczy o bardzo dobrym dopasowaniu.

**Skorygowany R²**

Skorygowany R²jest to użyteczna miara w porównaniu modeli z różną liczbą zmiennych, ponieważ penalizuje dodawanie zbędnych zmiennych. Wysoki skorygowany R² oznacza, że model dobrze dopasowuje się do danych, biorąc pod uwagę liczbę zmiennych i parametrów.

**Test F dla całkowitej istotności modelu**

Test F służy do sprawdzenia, czy cały model ma istotne dopasowanie do danych, tzn. czy przynajmniej jedna z zmiennych objaśniających ma istotny wpływ na zmienną zależną. Jeśli wartość p-testu F jest mniejsza niż 0.05, możemy odrzucić hipotezę zerową, że wszystkie współczynniki regresji są równe zero, co wskazuje na istotność modelu.

**Sumy kwadratów reszt (RSS)**

Suma kwadratów reszt mierzy, jak dobrze model dopasowuje się do danych. Im mniejsza suma kwadratów reszt, tym lepsze dopasowanie modelu. Niższa wartość RSS oznacza, że model lepiej przewiduje wartości zmiennej zależnej.

**Test Hausmana**

Test Hausmana służy do wyboru pomiędzy modelem z efektami stałymi a modelem z efektami losowymi. Wskazuje, czy różnice w estymacjach parametrów między tymi dwoma modelami są istotne. Jeśli test Hausmana wykazuje, że efekty indywidualne są skorelowane z zmiennymi objaśniającymi, preferowany jest model z efektami stałymi. Jeśli wynik testu Hausmana jest istotny (p-value < 0.05), preferujemy model z efektami stałymi, ponieważ sugeruje to, że efekty indywidualne są skorelowane z zmiennymi objaśniającymi.

**Akaike Information Criterion (AIC)**

AIC jest miarą, która ocenia jakość dopasowania modelu, karząc za złożoność modelu. Im mniejsza wartość AIC, tym lepszy model, który lepiej dopasowuje się do danych, jednocześnie zachowując prostotę. AIC pomaga wybrać model, który nie jest nadmiernie dopasowany do danych. Niska wartość AIC oznacza, że model wyjaśnia dużą część zmienności zmiennej zależnej przy minimalnej liczbie parametrów.

**Wstępna estymacja dla modelu podstawowego**

**Obraz zawierający tekst, zrzut ekranu, Czcionka, dokument

Zawartość wygenerowana przez AI może być niepoprawna.**

**Test F dla modelu dwuczynnikowego**

1. Model z efektami indywidualnymi vs dwuczynnikowy

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez AI może być niepoprawna.

1. Model z efektami czasowymi vs dwuczynnikowy

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez AI może być niepoprawna.

Na przyjętym poziomie istotności w oby dwu testach odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej mówiącej że istotne są zarówno efekty czasowe i indywidualne.

**Model dwuczynnikowy wyniki (efekty indywidualne vs losowe)**

1.Oszacowane parametry

|  |  |  |
| --- | --- | --- |
|  | Ef. indywidualne | Ef. losowe |
| Wynagrodzenia miesięczne brutto | 0,3956 | 0,4283 |
| Mieszkania oddane do użytku | 0,0275 | 0,0280 |
| Gęstość zaludnienia miast | -0,7755 | -0,7057 |
| Produkt krajowy brutto | -0,5635 | -0,5100 |
| Nowe miejsca pracy | -0,0110 | -0,0110 |
| Nakłady inwestycyjne | 0,0258 | 0,0163 |
| Przeciętne miesięczne wydatki | 0,1155 | 0,1259 |

2. Błąd oceny

|  |  |  |
| --- | --- | --- |
|  | Ef. indywidualne | Ef. losowe |
| Wynagrodzenia miesięczne brutto | 0,1447 | 0,1351 |
| Mieszkania oddane do użytku | 0,0124 | 0,0117 |
| Gęstość zaludnienia miast | 0,2773 | 0,2587 |
| Produkt krajowy brutto | 0,1463 | 0,1357 |
| Nowe miejsca pracy | 0,0128 | 0,0123 |
| Nakłady inwestycyjne | 0,0310 | 0,0294 |
| Przeciętne miesięczne wydatki | 0,0839 | 0,0789 |

3. Istotność statystyczna

|  |  |  |
| --- | --- | --- |
|  | Ef. indywidualne | Ef. losowe |
| Wynagrodzenia miesięczne brutto | 0,0069 | 0,0017 |
| Mieszkania oddane do użytku | 0,0273 | 0,0177 |
| Gęstość zaludnienia miast | 0,0057 | 0,0069 |
| Produkt krajowy brutto | 0,0002 | 0,0002 |
| Nowe miejsca pracy | 0,3914 | 0,3748 |
| Nakłady inwestycyjne | 0,4063 | 0,5794 |
| Przeciętne miesięczne wydatki | 0,1704 | 0,1119 |

4. Test Hausmana

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez AI może być niepoprawna.

Na przyjętym poziomie istotności nie ma podstaw do odrzucenia hipotezy zerowej mówiącej że dane popierają model z losowymi efektami indywidualnymi

**Model dwuczynnikowy z efektami losowymi vs model ze wspólnym wyrazem wolnym**

**(test Breuscha-Pagana Lagrange’a)**

Estymacja modelu z wspólnym wyrazem wolnym

Obraz zawierający tekst, zrzut ekranu, Czcionka, dokument

Zawartość wygenerowana przez AI może być niepoprawna.

Wyniki testu

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez AI może być niepoprawna.

Na przyjętym poziomie istotności odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej mówiącej że dane popierają model z losowymi efektami indywidualnymi i czasowymi.

**Interpretacja efektów krańcowych**

**Przeciętne miesięczne wynagrodzenia brutto**: Wzrost przeciętnych miesięcznych wynagrodzeń o 1% względem poziomu z 2010 roku prowadzi, do wzrostu indeksu cen mieszkań za 1 m² o około 0,43%. Oznacza to, że wyższe dochody w regionie istotnie podnoszą poziom cen mieszkań, co może świadczyć o istnieniu silnej zależności między siłą nabywczą gospodarstw domowych a rynkiem nieruchomości.

**Mieszkania oddane do użytkowania na sprzedaż lub wynajem:** Wzrost liczby mieszkań oddanych do użytku o 1% względem poziomu z 2010 roku prowadzi skutkuje przeciętnym wzrostem indeksu cen mieszkań o 0,028%. Wskazuje to na niewielką dodatnią zależność między podażą nowych lokali a poziomem cen, co może świadczyć o niedostatecznej skali podaży względem popytu lub o opóźnionym efekcie rynkowym.

**Gęstość zaludnienia powierzchni zabudowanej i zurbanizowanej (osoby/km**²): Zwiększenie gęstości zaludnienia o 1% względem poziomu z 2010 roku prowadzi powoduje, przy pozostałych czynnikach niezmienionych, spadek indeksu cen mieszkań o 0,71%. Wynik ten może odzwierciedlać negatywne skutki przeludnienia i urbanizacji dla atrakcyjności mieszkaniowej, takie jak pogorszenie warunków życia lub przeciążenie infrastruktury.

**Produkt krajowy brutto na 10 mieszkańców województwa:** Wzrost PKB per capita o 1% względem poziomu bazowego prowadzi do spadku indeksu cen mieszkań o 0,51%. Choć rezultat może wydawać się sprzeczny z intuicją, może on wynikać z występowania wyższej podaży mieszkań w bardziej rozwiniętych regionach, co ogranicza presję cenową, bądź z migracji mieszkańców do tańszych obszarów.

**Nowo utworzone miejsca pracy:** Wzrost liczby nowo utworzonych miejsc pracy o 1% względem poziomu z 2010 roku prowadzi skutkuje przeciętnym spadkiem cen mieszkań o 0,011%. Wpływ ten jest jednak bardzo niewielki i najprawdopodobniej nieistotny ekonomicznie, co sugeruje brak bezpośredniego przełożenia tej zmiennej na ceny nieruchomości.

**Nakłady inwestycyjne na 1 mieszkańca w cenach bieżących:** Zwiększenie nakładów inwestycyjnych o 1% względem poziomu z 2010 roku prowadzi powoduje wzrost indeksu cen mieszkań o 0,016%. Efekt ten jest dodatni, lecz słaby, co może wskazywać na ograniczony bezpośredni wpływ ogólnych inwestycji infrastrukturalnych na ceny mieszkań.

**Przeciętne miesięczne wydatki na 1 mieszkańca:** Wzrost konsumpcji w ujęciu miesięcznych wydatków per capita o 1% względem poziomu z 2010 roku prowadzi przekłada się na wzrost indeksu cen mieszkań o 0,0126%. Sugeruje to, że wyższy poziom wydatków konsumpcyjnych może sprzyjać wzrostowi cen mieszkań, prawdopodobnie poprzez zwiększony popyt na dobra trwałego użytku, w tym nieruchomości.

**Diagnostyka**