

Anna Hajduk*

Weryfikacja modelu regresji wielorakiej na przykładzie nieruchomości rekreacyjnych gmin: Gródek nad Dunajcem i Łososina Dolna

1. Wprowadzenie

Celem głównym badań było opracowanie modelu ekonometrycznego oraz zbadanie trendów i zależności panujących na lokalnym rynku nieruchomości rekreacyjnych [1].

Przebadano następujące gminy: Gródek nad Dunajcem i Łososina Dolna.

Biorąc pod uwagę wyżej wymienione cele, postawiono następujące hipotezy badawcze:

- zebrane dane, analiza istotności, a także normalizacja ich wartości utworzą model regresji wielorakiej;
- analiza statystyczna przeprowadzona za pomocą programów Statistica i Excel pozwala na stwierdzenie istotności cech i czynników kształtujących rynek nieruchomości rekreacyjnych;
- zarówno utworzony model, jak i procedura badania istotności cech mogą mieć charakter aplikacyjny.

2. Badanie struktury zbiorowości

Rynek nieruchomości nie posiada właściwych tylko dla niego metod badawczych. Wydaje się, że jeżeli zachodzą w nim pewne prawidłowości statystyczne, można wykorzystać tu klasyczne metody badania rozkładów w czasie i przestrzeni. W rzeczywistości praktyka badawcza pokazuje pewne problemy związane ze stosowaniem metod statystycznych [2], takie jak:

- bardzo utrudniony dostęp do danych o transakcjach, a także do większości podstawowych atrybutów nieruchomości; instytucje rejestrujące transakcje zwykle nie udostępniają informacji;

* Wydział Geodezji i Kartografii, Wyższa Szkoła Przedsiębiorczości, Nowy Sącz

- niejednorodność danych – oznacza to, że pochodzą one z różnych rynków niejednorodnych przestrzennie, jakościowo i ilościowo;
- nieprawdziwe dane pojawiające się w dokumentach – ogranicza to ich wiarygodność;
- prowadzone w czasie obserwacji na obiektach podobnych a nie tych samych.

W toku analiz na lokalnym rynku nieruchomości rekreacyjnych zebrano informacje o cenach rynkowych działek. Wszystkie transakcje miały miejsce w okresie od października 2002 do grudnia 2004 roku. Nie stwierdzono dodatkowych uwarunkowań i okoliczności mogących mieć wpływ na uzyskane ceny. Z całego zbioru informacji do badania struktury zbiorowości wybrano 102 przypadki. Są to działki budowlane przeznaczone pod budownictwo jednorodzinne, niezabudowane oraz nieruchomości rekreacyjne.

Według Stanisza [3] podstawowa analiza danych powinna doprowadzić do zwięzłego przedstawienia ogólnej charakterystyki istotnych właściwości badanej zbiorowości. Analiza struktury przestrzennej ma na celu stworzenie podstaw do badania wielkości podaży oraz ustalenia metod analizy przestrzennego rozmieszczenia cen i wartości nieruchomości.

Statystyczną analizę lokalnego rynku rozpoczęto od następujących miar położenia:

- średnich klasycznych,
- przeciętnych pozycyjnych.

Ze średnich klasycznych wybrano średnią arytmetyczną. Na podstawie zebranych informacji obliczono, że wynosi ona 18,25 zł za 1 m² gruntu i przyjęto jako przeciętną cenę nieruchomości rekreacyjnej. Z pośród parametrów tzw. pozycyjnych wybrano dwie miary:

- 1) modalną, czyli wartość najczęściej występującą,
- 2) medianę, czyli wartość środkową.

Wśród podanych wartości ceny nie zaobserwowano wielkości występujących częściej niż inne, z czego wynika, że nie można obliczyć miary modalnej. Rozważa się parzystą liczbę obserwacji tj. 102 pozycje, a więc medianę należy wyznaczyć ze wzoru (1)

$$M = \frac{x_k + x_{k+1}}{2} \quad (1)$$

Z obliczeń otrzymano wartość równą 14,44, co oznacza, że w 51 przypadkach średnie ceny nieruchomości były mniejsze od 14,44 zł na 1 m², a w pozostałych 51 były co najmniej równe tej cenie. W przypadku rozkładów symetrycznych (lub do nich podobnych), średnie klasyczne arytmetyczne oraz średnie pozycyjne (modalna i mediana) są do siebie zbliżone.

Miary położenia nie wystarczają do pełnego opisu struktury zbiorowości pod względem wybranej cechy statystycznej. Niezbędnym uzupełnieniem są miary zmienności. Jedną z podstawnych charakterystyk tej grupy jest rozstęp, czyli różnica pomiędzy najmniejszą a największą ceną jednostkową, który w przykładzie badanej zbiorowości wynosi 44,33 zł/m². Bardziej dokładną informację o zmienności badanej cechy przedstawia odchylenie standardowe. W pierwszej kolejności wyznaczono wariancję, a na jej podstawie odchylenie $s = 12,26$ zł/m². Aby ocenić stopień zróżnicowania, należy obliczyć współczynnik zmienności. Zmienność średnich cen kształtuje się na poziomie 67%, co oznacza, iż badana cecha charakteryzuje się dużą zmiennością.

W celu uzyskania lepszego opisu badanej struktury należy zbadać typ rozkładu empirycznego. Ważną informacją jest określenie, czy badana cecha w rozważanej zbiorowości charakteryzuje się rozkładem symetrycznym czy też asymetrycznym. W tym celu wykorzystuje się współczynnik asymetrii oparty na momencie centralnym rzędu trzeciego.

Wartość momentu oblicza się według następującego wzoru (2)

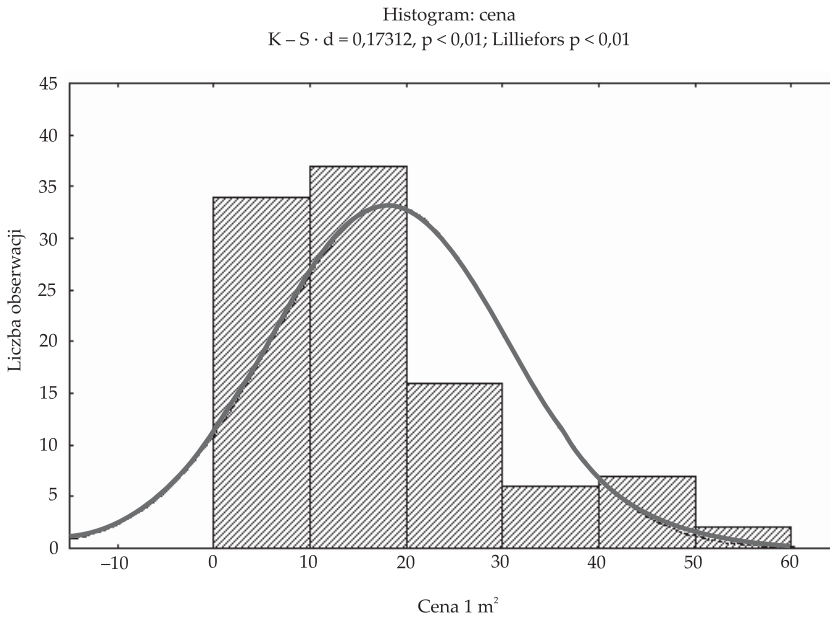
$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (2)$$

i wynosi on 239,29.

Współczynnik asymetrii – zwany też współczynnikiem skośności A_s – określa nie tylko kierunek, lecz także siłę asymetrii. Wynosi 0,19, jest dodatni, a więc asymetria jest prawostronna. Na histogramie (rys. 1) przedstawiano zależność ceny od liczby transakcji potwierdzającą asymetrię prawostronną. Ten sam wynik uzyskujemy z nierówności, jaka zachodzi między medianą, a wartością średnią dla tego rozkładu, czyli: $x = 18,25$ zł/m² > $M = 14,44$ zł/m².

Są sytuacje, w których badanie średniego poziomu zmiennej i rozproszenia jej wartości nie wskazuje na istnienie różnic między badanymi zbiorowościami. Obserwacja rozkładów cech wyklucza podobieństwo rozważanych zbiorowości [3].

W przypadku występowania asymetrii rozkładów zmiennych istotne jest badanie dodatkowo tzw. koncentracji zjawisk. Przez koncentrację rozumie się nierównomierne rozłożenie ogólnej sumy wartości cechy pomiędzy poszczególne jednostki zbiorowości. Przy opisie wartości zmiennej doskonałym uzupełnieniem jej miary są poznane parametry, ponieważ opisują rozkład wartości danej cechy wokół średniej. Najbardziej znaną miarą takiego skupienia obserwacji jest kurtoza oznaczana jako K . Im jest ona wyższa, tym krzywa liczebności jest bardziej wypukła. Jeśli $K < 0$, to rozkład jest bardziej spłaszczony od normalnego, a jeżeli $K > 0$ to rozkład jest bardziej wysmukły niż normalny. Kurtoza dla analizowanej zbiorowości wynosi $K = 1,33$, co świadczy o bardzo dużej koncentracji wokół średniej.



Rys. 1. Liczba transakcji w przedziale cenowym

3. Budowa modelu regresji wielorakiej

3.1. Zdefiniowanie zmiennych do modelu

W celu utworzenia modelu przebadanych zostało 200 nieruchomości, w skład których wchodziły nieruchomości rolne, rekreacyjne, budowlane oraz zabudowane. Największą grupę wśród nich stanowiły działki rekreacyjne i budowlane niezabudowane z przeznaczeniem pod zabudowę jednorodziną (53%). Drugą pod względem wielkości były nieruchomości zabudowane (40%) i najmniej, bo 7% stanowiły nieruchomości rolne.

Do wstępnej analizy danych oraz do budowy modelu ekonometrycznego wykorzystano 5 zmiennych objaśniających opartych na 92 przypadkach. Stanowią one (wg planu zagospodarowania przestrzennego) nieruchomości przeznaczone pod rekreację oraz zabudowę jednorodziną. Pozostałe 108 działek nie podlega badaniu, ze względu na ich klasyfikację tj. działki rolne i zabudowane.

W badanym modelu wykorzystano:

- zmienną zależną (objaśnianą), którą jest cena 1 m² gruntu (Y);
- zmienną niezależną (objaśniającą), którymi są:
 - data zawarcia transakcji kupna-sprzedaży (X_1),
 - lokalizacja (X_2),

- dostępność do jeziora (X_3),
- uzbrojenie terenu (X_4),
- powierzchnia działki (X_5).

Zmienne ilościowe występujące w modelach mogą przyjmować nieskończenie wiele wartości. Znaczna część zjawisk ekonometrycznych i gospodarczych ma charakter jakościowy. Do modelu najczęściej używa się zmiennych zwanych dychotomicznymi, które przyjmują dwie wartości 0 lub 1. Tak zakodowane są wszystkie zmienne objaśniające z wyjątkiem daty. Nie jest istotne, która zmienna zostanie pominięta, ale musi być ona jednoznaczna do interpretacji. Wykluczenie jednej zmiennej pomocniczej zabezpiecza model przed pełną współliniowością i umożliwia obliczenie równania. Jeśli zmienna jakościowa albo ilościowa posiada n poziomów, należy jej przyporządkować $(n - 1)$ zmiennych zerojedynkowych.

Pierwsza z rozpatrywanych zmiennych, czyli data (X_1), została wprowadzona do bazy bezpośrednio, a jej format zmieniono na liczbowy w programie Statistica. Drugą jest lokalizacja (X_2) zakodowana binarnie i przyjmuje wartości: jeden, gdy badana działka należy do danego przedziału, oraz zero w przeciwnym wypadku. Do modelu przyjęto 6 najbardziej reprezentatywnych miejscowości zlokalizowanych w obrębie Jeziora Rożnowskiego. Są to: Bartkowa, Rożnów, Gródek nad Dunajcem, Tabszowa, Znamirowice, Tęgorozę. Pominiętą zmienną jest Bartkowa.

Z lokalizacją jest związana zmienna: „dostępność do jeziora” (X_3). Przyjmuje ona wartość 1 w sytuacji, gdy działka leży bezpośrednio przy linii brzegowej jeziora lub w odległości do 500 m (dostępność do jeziora jest dobra).

Kolejną z analizowanych cech jest uzbrojenie terenu (X_4), dla którego przyjęto następujące podziały i wartości:

- pełna sieć uzbrojenia terenu o wartości 1 (nieruchomość posiada trzy lub cztery rodzaje sieci, np.: elektryczność, wodę, kanalizację, gaz);
- częściowe uzbrojenie terenu o wartości 0 (nieruchomość posiada jedną lub dwie sieci, np.: elektryczność i wodę);
- brak sieci uzbrojenia, co oznacza, że nieruchomość nie posiada żadnej sieci.

Ostatnią analizowaną cechą jest powierzchnia działki. Ma ona wartość jeden, jeśli jej wielkość jest mniejsza od 600 m², a w pozostałych przypadkach wynosi zero. Gdy jej powierzchnia przekracza 1000 m², nie występuje ona w modelu. Działki z pierwszego przedziału mają najbardziej optymalną powierzchnię i dlatego otrzymały wartość 1.

Na wartość rynkową nieruchomości rekreacyjnych wpływają czynniki związane zarówno z cechami samej nieruchomości, jak i jej otoczenia. Dokładny opis sprzedanej działki pozwala wyodrębnić liczny zbiór cech, które mogą wpływać na jej wartość. Niestety tylko niewielka liczba atrybutów ma istotny wpływ na cenę.

Bardzo ważną cechą jest lokalizacja, czyli przestrzenne usytuowanie, która wpływa na dostępność usług handlowych, usługowych, czy głównych tras komunikacyjnych oraz walorów krajobrazowych. Innym istotnym aspektem nieruchomości jest jej stopień wyposażenia w urządzenie infrastruktury technicznej, co wpływa na standard i cenę. Z badań wywnioskowano, że im bogatsze sąsiedztwo naturalnych walorów krajobrazowych danej nieruchomości (jeziora, lasy, góry itp.), tym większa jest jej atrakcyjność i wartość.

W modelu statystycznym można uwzględnić wiele innych czynników nieujętych w badaniach, choć równie istotnych przy wartościowaniu terenów, a są to m.in.: strona sprzedająca, kształt działki czy dostępność do ośrodków wypoczynkowych. Na wartość rynkową mogą wpływać też czynniki makroekonomiczne, jak np.: rozwój gospodarki, zmiany koniunktury gospodarczej czy poziom bezrobocia. W dobrze skonstruowanym modelu uwzględnia się jednak najbardziej istotne cechy określające zmienność opisywanego zjawiska, pomijając mniej ważne czynniki w stosunku do wybranych.

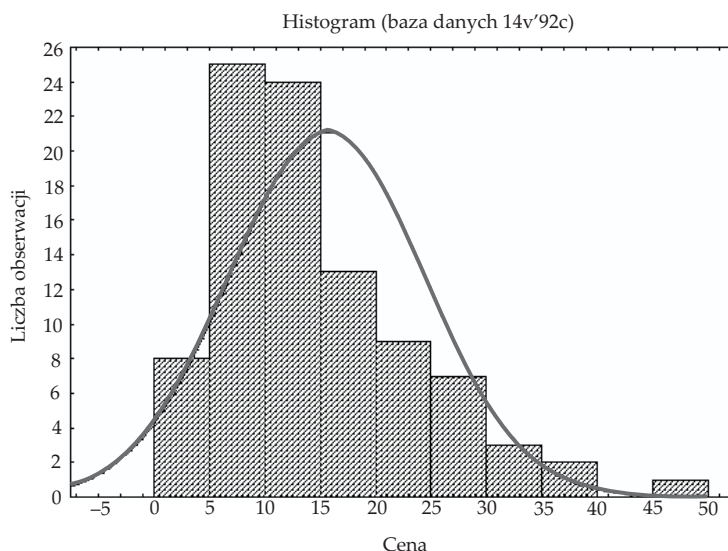
3.2. Wstępne wyniki badań przeprowadzonych na lokalnym rynku nieruchomości rekreacyjnych

Na początku analizy sprawdzono, jak przedstawia się liczba sprzedanych działek w poszczególnych miejscowościach. Największy obrót nieruchomościami był w obrębie gminnym – Gródku nad Dunajcem (31%) i Znamierowicach (22%), w Tęgorborzy, Rożnowie i Bratkowej odpowiednio po 14%, 12% i 12%, najmniej sprzedaży zanotowano w Tęgorborzy – około 9%.

Kolejnym etapem analizy było utworzenie tabel zawierających podstawowe zestawienia i porównanie procentowe. Dla pierwszej zmiennej (X_1), jaką jest lokalizacja, ustalono średnie ceny za 1 m² w poszczególnych obrębach.

Ogólnie biorąc, rozrzut cenowy przyległych obszarów terenowych, jakimi są: Gródek, Rożnów i Bartkowa waha się w granicach 1 zł za m², co można uznać za prawidłowe zjawisko. W omawianym zbiorze działek 59% miało powierzchnię mniejszą niż 600 m², a średnia cena tych działek wyniosła 18,43 zł za m² gruntu. W zakresie od 600 do 1000 m² było około 28% analizowanych nieruchomości, a średnia cena w tym przedziale wynosiła 13,31 zł za m². Ostatnia grupa obejmowała zaledwie 11 nieruchomości i ich średnia cena wyniosła 7,49 zł za m². Zaobserwowana wcześniej prawidłowość potwierdziła, że im większa jest powierzchnia działki, tym cena jest mniejsza.

Na histogramie (rys. 2) przedstawiono liczbę sprzedanych działek w określonym przedziale cenowym. Oś odciętych przedstawia przedziały klasowe ceny 1 m² gruntu (wyodrębniono 9 przedziałów klasowych o szerokości 5 zł), natomiast oś rzędnych mówi o liczbie obserwacji w poszczególnych klasach.



Rys. 2. Liczba sprzedanych działek w zależności od ceny

Najwięcej działek osiągnęło ceny z przedziału od 5 do 10 zł (25 przypadków) oraz od 10 do 15 zł (24 obserwacje). W przedziałach od 15 do 20 zł było 15 obserwacji i ze wzrostem ceny liczba obserwacji malała. W przedziałach od 30 do 45 było bardzo mało transakcji, wynika z tego, że działki te nie cieszyły się bardzo dużym zainteresowaniem (6 przypadków).

W analizowanym zbiorze danych prawie wszystkie działki (98%) posiadają elektryczność. Około 82% działek będących przedmiotem kupna – sprzedaży posiada sieć wodociągową i dużo, bo 60% sieć telefoniczną. Sieć kanalizacyjna jest bardzo słabo rozwinięta, ponieważ posiada ją tylko 25 działek.

Ostatnią z analizowanych zmiennych był dostęp do jeziora. Wśród przebadanych transakcji kupna sprzedaży 54% znajdowało się przy linii brzegowej jeziora lub do 500 m od brzegu. Cena tych działek wynosiła 18,05 zł i była najwyższa w porównaniu z dostępnością innych działek. Złe położenie miało 7 działek, i ich średnia cena wynosiła 9,17 zł na m².

3.3. Budowa modelu regresji wielorakiej na podstawie opracowanej bazy danych

Budowa i analiza modelu regresji wielorakiej ma na celu zbadanie istnienia zależności pomiędzy zmiennymi i ocenę wpływu wartości zmiennych X_i na zmienną Y (cenę nieruchomości). W analizowanym modelu szukano wpływu najbardziej istotnych cech nieruchomości na ich cenę.

Zbudowano model regresji w postaci:

$$\text{Cena} = b_0 + b_1 \text{Data} + b_2 \text{Lokalizacja} + b_3 \text{Dostępność} + b_4 \text{Uzbrojenie} + b_5 \text{Powierzchnia}.$$

Wyjściowa baza danych zawierała zmienne zakodowane binarnie i posłużyła do estymacji ocen wpływu cech jakościowych. Uzyskane w wyniku regresji oceny liczbowe dla poszczególnych zmiennych wprowadzono do bazy danych, zastępując zera i jedynki odpowiednimi wartościami w zależności od przedziału klasowego. Uzyskane w ten sposób oceny cech niemierzalnych mają charakter obiektywny.

W tabeli 1 przedstawiono wyniki regresji wielorakiej. Zawiera ona współczynnik przy każdej zmiennej, błąd standardowy oraz test $t(86)$ pozwalający sprawdzić istotność równania. Jest również poziom p , którego wartość reprezentuje malejący wskaźnik poprawności wyniku.

Tabela 1. Zestawienie wyników regresji wielorakiej

Wyszczególnienie	Wyraz wolny	Data	Lokalizacja	Dostępność	Uzbrojenie	Powierzchnia
b_i	27,74846	-0,0009	1,02889	1,04410	0,96285	0,95593
błąd standardowy	134,7901	0,0035	0,1612	0,2507	0,1598	0,2206
$t(86)$	0,210550	-0,25824	6,383594	4,164294	6,024758	4,334026
poziom p	0,833736	0,796834	0,00000	0,00007	0,00000	0,00004

Zgodnie z tabelą 1 ostateczny model regresji będzie miał następującą postać:

$$\hat{Y} = 27,74846 - 0,0009 X_1 + 1,03 X_2 + 1,04 X_3 + 0,962 X_4 + 0,96 X_5,$$

gdzie:

\hat{Y} – teoretyczna (przewidywana) cena 1 m² gruntu,

X_1 – data zawarcia transakcji kupna – sprzedaży,

X_2 – lokalizacja,

X_3 – dostępność,

X_4 – uzbrojenie terenu,

X_5 – powierzchnia.

Analizując wyniki regresji, stwierdzono, że korelacja między datą zawarcia transakcji, a ceną nieruchomości jest bardzo niska (-0,02). Pozostawiono ją jednak w modelu w celu zbadania zmian cen na skutek czasu.

Kolejnym etapem analizy modelu regresji wielorakiej była interpretacja uzyskanych wyników (oszacowanych parametrów równania).

Pierwszą analizowaną zmienną była data zawarcia transakcji kupna – sprzedaży. Współczynnik dla tej zmiennej prowadzi do wniosków, że zmiana daty o jeden dzień

przynosi spadek ceny jednostkowej o 0,0009 zł (przy stałości pozostałych zmiennych). Cena nieruchomości w krótkim okresie czasu jest niezmienna. Rocznie cena 1 m² gruntu wzrasta średnio o 0,3285 zł (365 × 0,0009 zł). Przy zakupie nieruchomości o powierzchni 1000 m² w cenie 15 zł za 1 m² upływ czasu spowoduje nieznaczny spadek jej ceny. Ze względu na wartości współczynników analizowano wszystkie cechy.

Współczynniki regresji mają różne jednostki miary w zależności od typu rozważanych zmiennych. Nie można więc ich użyć do bezpośredniego porównania siły powiązania poszczególnych zmiennych niezależnych ze zmienną zależną. Można wykorzystać do tego celu standaryzowane współczynniki beta. W analizowanym przypadku najwyższy poziom beta osiągnęła lokalizacja (0,5) i uzbrojenie (0,5), następnie dostępność i powierzchnia (0,3). Najśłabsze powiązanie z ceną wykazuje data zawarcia transakcji.

3.4. Weryfikacja modelu

Poprawność statystyczną modelu testowano z zastosowaniem kilku mierników.

Pierwszym jest współczynnik determinacji (R^2), który jest miarą stopnia dopasowania zależności liniowych w modelu. Osiągnął on wartość 0,80, co oznacza, że zmienne objaśniające wyjaśniają zmienność ceny jednostkowej w 80%, jest to bardzo dobre dopasowanie.

Kolejnym jest standardowy błąd szacunku, który wynosi 5,59 (ocena wielkości losowych odchyłeń modelu) i oznacza przeciętne odchylenie ceny obserwowanej w próbie od ceny teoretycznej – wyznaczonej z modelu. Ceny rzeczywiste (zaobserwowane) różnią się od cen teoretycznych średnio o 5,59 zł na 1 m² gruntu.

Standardowy błąd szacunku wynosi 5,67 zł, a średnia cena $Y = 15,56$ zł za m².

Odchylenie standardowe reszt stanowi 36% średniej wartości cen jednostkowych nieruchomości.

Test F służy do weryfikacji hipotezy istotności całego modelu. Postawiona hipoteza zerowa: żadna i -ta zmienna niezależna w modelu nie ma istotnego wpływu na zmienną zależną. Hipoteza alternatywna natomiast oznacza, że istnieje co najmniej jedna zmienna istotnie związana ze zmienną Y .

Postawiono następujące hipotezy do testowania badanego modelu:

- H_0 : cechy badanych nieruchomości (zmienne niezależne X_1, X_2, X_3, X_4, X_5) nie mają istotnego związku z ceną;
- H_1 : cechy badanych nieruchomości (zmienne niezależne X_1, X_2, X_3, X_4, X_5) mają istotny związek z ceną.

Wartość statystyki F-Snedecora obliczona w programie Statistica wynosi: 2731,6. Wyznaczona wartość krytyczna F na podstawie tablic rozkładu F , przy poziomie istotności 0,05 wynosi 2,21 (dla $k = 5$ i $n - k - 1 = 86$ stopni swobody). Reguła decyzyjna mówi, że w przypadku gdy $F > F_\alpha$ ($2732,6 > 2,21$) hipotezę zerową nale-

ży odrzucić. Odrzucono hipotezę zerową na korzyść alternatywnej, co oznacza, że zmienne niezależne wywierają łącznie istotny wpływ na cenę nieruchomości [3].

Test F potwierdza albo odrzuca hipotezę o zachodzeniu związku między zmienną niezależną, a co najmniej jedną ze zmiennych objaśniających. W przypadku, gdy związek ten zachodzi należy przeprowadzić dodatkowe testy, w celu zbadania, które współczynniki kierunkowe β_1, \dots, β_k są różne od zera. Narzędziem pozwalającym na sprawdzenie istotności parametrów równania jest test t-Studenta o $t - (k + 1)$ stopniach swobody. Służy on do określenia, czy zmienna objaśniająca X_i ma moc wyjaśniania zmienności zmiennej zależnej Y . Mówi o tym, czy tę badaną zmienną należy odrzucić z równania regresji.

Dla wszystkich zmiennych sformułowano hipotezy istotności $H_0: \beta_i = 0$ oraz $H_1: \beta_i \neq 0$.

Postawiono następujące hipotezy:

H_0 – zmienne objaśniające nie mają istotnego wpływu na cenę;

H_1 – zmienne objaśniające mają istotny wpływ na cenę.

W analizowanym przykładzie wartość krytyczna statystyki t odczytana z tabeli dla poziomu istotności równego 0,05 oraz 86 stopni swobody wynosi 2,6. Empiryczna statystyka t dla zmiennej „powierzchnia” wynosi 3,11. Wartość należy do przedziału, który jest prawostronnym obszarem odrzucenia. Można wnioskować, że istnieją statystyczne podstawy do przyjęcia hipotezy, że parametr β_5 (współczynnik kierunkowy przy zmiennej X_5), jest różny od zera. Zmienna X_5 posiada moc wyjaśniania zmiennej zależnej Y . Badając pozostałe zmienne, można analogicznie wnioskować. Wartości statystyki t przy poziomie istotności równym 0,05 leżą w prawostronnym obszarze odrzucenia hipotezy zerowej (wartość p jest mniejsza od 0,05).

Test istotności zmiennej objaśniającej X_1 (data) pozwala wnioskować, że wartość t nie należy do obszaru odrzucenia hipotezy zerowej (wartość p przekracza poziom istotności). Hipotezę zerową należy przyjąć, czyli zmienne objaśniająca jaką jest czas nie ma istotnego wpływu na cenę nieruchomości. Zmienna ta nie została jednak usunięta z modelu, aby stworzyć możliwość zbadania zmian cen w czasie.

Wartość poziomu p przedstawia malejący wskaźnik rzetelności wyniku. Inaczej jest to osiągnięty poziom α związany z daną wartością sprawdzianu, przy którym można odrzucić hipotezę zerową. Jeżeli więc p przekracza poziom $\alpha = 0,05$, wtedy należy przyjąć hipotezę zerową. Jeżeli natomiast wartość p jest mniejsza od 0,05 należy ją odrzucić. Im wyższy poziom p , tym zaobserwowana zależność między zmiennymi w próbie jest mniejsza od zależności między odpowiednimi zmiennymi w populacji.

W pakiecie Statistica dostępna jest też wartość testu Durбина–Watsona (DW) oraz wartość estymatora współczynnika autokorelacji (ρ). Wartość statystyki DW

w analizowanym modelu wynosi 2,07. Odczytane z tablic rozkładu statystyki d przy poziomie istotności 0,05 wynoszą odpowiednio: $d_1 = 1,41$, $d_u = 1,64$. Ponieważ DW jest większe od d_u można wnioskować, że założenie o braku korelacji składnika losowego jest spełnione.

3.5. Wykorzystanie modelu regresji wielorakiej do prognozowania wartości nieruchomości

Regresja wieloraka znajduje zastosowanie w prognozowaniu czyli określaniu przysługujących wielkości zmiennej zależnej na podstawie równania.

Ostatnia transakcja w analizowanej bazie danych miała miejsce 2004 roku. Zebrano informacje dotyczące transakcji kupna-sprzedaży zawarte po tym okresie i wykorzystano je do prognozy *ex post*. Na podstawie zbudowanego modelu regresji dokonano prognozy cen rynkowych działek, których wartości zakodowanych odpowiednio zmiennych niezależnych wykorzystano w prognozie *ex post*.

Najwyższy błąd prognozy otrzymała nieruchomość położona w Znamierowicach, stanowi on 46% wartości prognozowanej. Najtrafniej cenę nieruchomości oszacowano dla nieruchomości położonej w Bartkowej. Błąd prognozy wynosi 0,49 zł, co stanowi 6% wartości prognozowanej. Średni błąd prognozy dla wszystkich dziesięciu nieruchomości wynosi 1,61, co oznacza, że rzeczywista cena nieruchomości będzie się przeciętnie różniła od prognozowanej o 1,61 zł za 1 m².

4. Weryfikacja hipotez i wnioski

Hipoteza pierwsza zakładała, że zebrane dane, analiza istotności, a także normalizacja ich wartości utworzą model regresji wielorakiej. Zbiór obserwacji przebadano pod względem występowania zależności pomiędzy nimi. Struktura zbiorowości cen pokazała, jakie jest ich zróżnicowanie, koncentracja czy też rozproszenie. Do modelu ekonometrycznego wśród analizowanych danych wybrano mniej obserwacji. Usunięto dane, które znacznie odbiegały od pozostałych (np. średnia cena wynosiła około 15 zł za m², a było kilka przypadków powyżej 50 zł). Do modelu dobrano zmienne objaśniające (powierzchnia, data, lokalizacja, dostępność do jeziora, uzbrojenie i zmienną zależną (cenę). Zostały one zakodowane binarnie. Utworzono model i zweryfikowano go odpowiednimi miernikami. Współczynnik determinacji świadczył o dobrym dopasowaniu modelu (80%), a korelacja była na wysokim poziomie (64%). Hipoteza ta została przyjęta.

Druga hipoteza zakładała, że analiza statystyczna, z wykorzystaniem programów: Statistica i Excel, pozwala na stwierdzenie istotności cech i czynników kształtujących rynek nieruchomości rekreacyjnych. Ta hipoteza również została zweryfikowana pozytywnie. Na podstawie pakietu komputerowego Statistica sprawdzono, czy dobrane zmienne są istotne, czy wywierają duży wpływ na cenę.

Ostatnia hipoteza zakładała, że zarówno model ekonometryczny jak i procedura badania istotności cech mają charakter aplikacyjny, dlatego mogą być stosowane do prognozowania przyszłej wartości nieruchomości. Zbudowany model posłużył do sprawdzenia nowej bazy danych utworzonej już po zbudowaniu modelu. Uzyskane ceny były zbliżone do uzyskanych na podstawie transakcji kupna-sprzedaży, ale obciążone pewnym błędem, którego nie można unikać, biorąc pod uwagę krótki okres, z którego pochodzą dane. Hipotezę tę zweryfikowano pozytywnie.

Literatura

- [1] Hajduk A.: *Dobór i analiza parametrów na potrzeby nieruchomości rekreacyjnych*. Geodezja. Półrocznik Akademii Górniczo-Hutniczej, t. 12, z. 2, 2006, Kraków.
- [2] Hozer J.: *Zastosowanie ekonometrii w wycenie nieruchomości*. XII Krajowa Konferencja Rzeczoznawców Majątkowych „Kierunki rozwoju w teorii i praktyce wyceny”, Koszalin 2003.
- [3] Stanisław A.: *Przystępny kurs statystyki na przykładach z medycyny*, StatSoft, 2003.