

Po co wymyślono statystykę?

Mieczysław Prystupa

Wstęp

Początki statystyki sięgają czasów starożytnych, gdyż w wielu państwach dokonywano spisów powszechnych ludności. Na terenie imperium rzymskiego dokonywano spisów dość systematycznie. Termin „statystyka” pochodzi od łacińskiego słowa „status”, który oznacza stan rzeczy, państwo. Metody statystyczne stosuje się wszędzie tam, gdzie chodzi o poznanie prawidłowości w zakresie zjawisk masowych, w szczególności tam, gdzie bada się problemy demograficzne, ekonomiczne, socjologiczne, ale również w innych dziedzinach. Metody statystyczne należą do metod badań naukowych. Wraz z rozwojem gospodarki posiadanie podstawowych danych stało się niewystarczające. Konieczne stało się nie tylko ulepszenie metod pozyskiwania danych, ale również ich opisu i analizy. Zbiegło się to w czasie z szybkim rozwojem metod matematycznych, szczególnie teorii prawdopodobieństwa. Przede wszystkim wynika to z losowej natury badanych wielkości. Uważa się, że na pewnym poziomie obserwacji wszystkie zjawiska mają charakter losowy. Ważniejszy jest jednak fakt, że czasami brak jest możliwości lub środków do tego, aby przebadać całą populację. Badanie np. preferencji wyborczych wszystkich ludzi w Polsce z ekonomicznego punktu widzenia byłoby zbyt kosztowne, lepiej zatem przebadać losowo wybraną grupę osób, zakładając, że reszta populacji będzie miała podobny rozkład preferencji wyborczych. Losowość zjawisk jest niejako wpisana w definicję metod statystycznych. **Dlatego właśnie statystykę łączy bardzo ścisły związek z teorią prawdopodobieństwa, działem matematyki dzięki któremu jesteśmy w stanie poradzić sobie z niepewnością.** Z tego krótkiego opisu zadań statystyki wynika, że zbadanie i wycena jednostkowego przypadku, jakim jest np. konkretna nieruchomość, nie jest domeną statystyki.

1. Dlaczego stosowanie metod statystycznych do wyceny konkretnych nieruchomości jest niewskazane

Pewna część rzeczoznawców próbuje zastępować podejście porównawcze metodami statystycznymi. Wiarygodność takich wycen, jeśli były poddane ocenie komisji

arbitrażowych bądź Komisji Odpowiedzialności Zawodowej, była we wszystkich znanych mi przypadkach kwestionowana. Metody statystyczne usiłuje się stosować do określania wag cech rynkowych nieruchomości. Uzasadnia się to tym, że trudno jest znaleźć dwie nieruchomości różniące się jedną cechą by wagę cechy oszacować stosując zasadę *ceteris paribus*. Próbuje się więc zastosować wzory na korelację. Korelacja jest wykorzystywana przy założeniu, że zmienną zależną Y są ceny nieruchomości, a zmiennymi niezależnymi X są cechy rynkowe np. lokalizacja, stan techniczny, sąsiedztwo itp., którym nadaje się odpowiednie stany. W konsekwencji określa się siłę związku pomiędzy cenami a poszczególnymi cechami. Jak już wielokrotnie pisałem problem polega jednak na tym, że badając statystyczne zależność pomiędzy cenami, a poszczególną cechą rynkową, także zakłada się, że inne cechy nie wpływają na poszczególne ceny. **To założenie wynika wprost z metody korelacji.** W rzeczywistości jednak inne cechy – oprócz określonej – wpływają na ceny i wyniki określające wagi cech rynkowych są zupełnie niewiarygodne. Niejednokrotnie więc zachodzą przypadki, że cecha, która powinna mieć niewielki wpływ na ceny w wyniku zastosowania wzorów korelacyjnych jest dominująca. Jest jednak dodatkowa psychologiczna pułapka, w którą wpadają rzeczoznawcy twierdząc, że skoro tak wyszło na podstawie wzorów – to musi być to dobre i mądre rozwiązanie. Niestety nie jest – bo stosowany model nie spełnia założeń i nie jest w stanie oszacować wag cech rynkowych. **Zdecydowanie lepiej jest ustalić wagi cech rynkowych na podstawie własnego doświadczenia i obserwacji niż zaufać wzorom, które dają całkowicie niewiarygodne wyniki.**

Jeszcze większe pułapki czyhają na rzeczoznawcę, który usiłuje jednym wzorem w postaci równania regresji wielorakiej określić wartość nieruchomości.

Podstawowe równanie ma postać:

$$C_i = \alpha_0 + \alpha_1 \cdot X_{1i} + \alpha_2 \cdot X_{2i} + \dots + \alpha_k \cdot X_{ki} + U_i$$

gdzie:

C_i – ceny nieruchomości,

α_0 – wyraz wolny,

$X_{1i}, X_{2i}, \dots, X_{ki}$ – zmienne objaśniające,

α_k – współczynniki, które interpretuje się jako wagi cech,
 U_i – składnik losowy.

W przypadku wyceny nieruchomości tworzy się cały szereg równań na podstawie znanych transakcji – by program komputerowy wygenerował jedno równanie, dzięki któremu usiłuje się określać wartość nieruchomości opisaną podobnie jak nieruchomości o znanych cenach, cechach i ich stanach.

Równanie to przybiera postać:

$$Y = b_0 + b_1X_1 + b_2X_2 + + \dots + b_kX_k + U_i$$

gdzie:

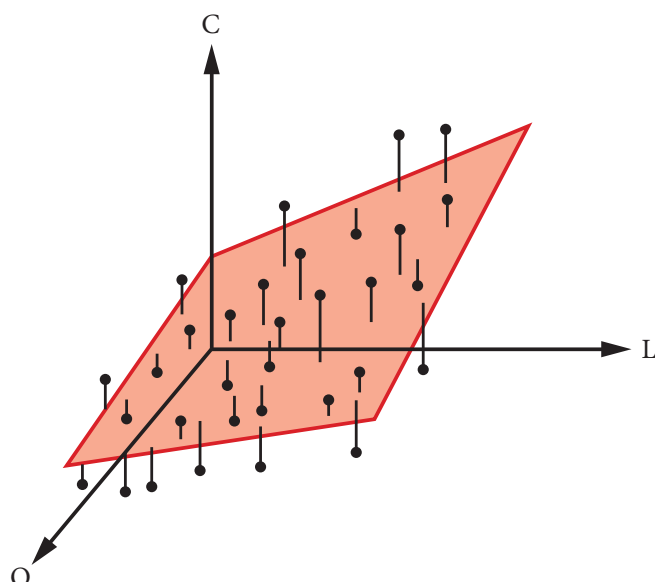
- Y- wartość nieruchomości,
- X_1, X_2, \dots, X_k – zmienne objaśniające – cechy rynkowe,
- b_0 – wyraz wolny,
- b_i – współczynniki, które można interpretować jako wagi cech,
- U_i – składnik losowy.

Aby zrozumieć istotę regresji wystarczy prześledzić przykład określania wartości rynkowej najpierw prostym modelem regresji liniowej jak to zaprezentowano na rys.1. Przykład dotyczy określenia wartości nieruchomości przy uproszczającym założeniu, że na określonym rynku ceny jednostkowe zależą wyłącznie od odległości nieruchomości niezabudowanej do centrum miasta. Linia regresji jest wyznaczona na podstawie

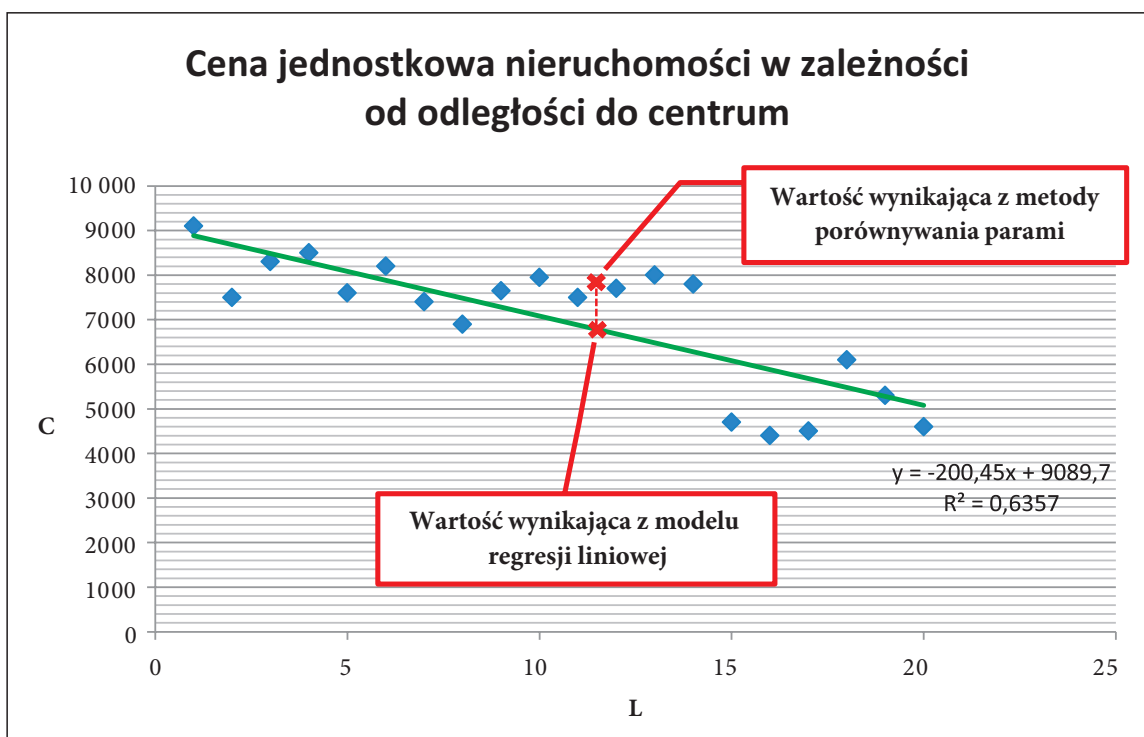
metody najmniejszych kwadratów. Oznacza to, że suma kwadratów odległości od linii regresji jest najmniejsza. W przykładzie tym pokazano, że wartość jednostkowa nieruchomości w odległości 10 km od centrum wyznaczona linią regresji, odbiega od wartości rynkowej określonej w podejściu porównawczym. W takiej odległości od centrum porównywalne nieruchomości leżą znacznie ponad linią regresji.

Z kolei na rysunku 2. pokazano, że w przypadku dwóch zmiennych objaśniających – lokalizacji i otoczenia, szacujemy równanie płaszczyzny, dla której suma

Rysunek 2.



Rysunek 1.



kwadratów jest najmniejsza dla wszystkich danych: zarówno lokalizacji jak i otoczenia. W przypadku k zmiennych powstaje hiperpłaszczyzna, która musi także spełniać warunek minimum kwadratów odległości dla wszystkich danych. Program komputerowy takie obliczenia przeprowadzi, lecz nie zapewni wiarygodnej wyceny.

Aby model statystyczny mógł dawać wiarygodne wyniki musi spełniać cały szereg założeń. **Wyznaczenie wartości nieruchomości przy pomocy metody regresji wielorakiej uwarunkowane jest zwłaszcza spełnieniem założeń metody najmniejszych kwadratów.**

Jak wynika z wielu publikacji w tym zakresie, założenia tej metody są następujące:

- 1) zmienne objaśniające (cechy rynkowe) są wielkościami nielosowymi i nie zachodzi między nimi współliniowość,
- 2) składnik losowy jest zmienną losową, której nadzieja matematyczna jest równa zero, a wariancja jest stałą,
- 3) obserwacje są niezależne,
- 4) składnik losowy jest nieskorelowany ze zmiennymi objaśniającymi,
- 5) liczba zmiennych objaśniających musi być mniejsza od liczby obserwacji (cen transakcyjnych),
- 6) nie występują współzależności między składnikami losowymi poszczególnych równań modelu.

Nie wszystkie założenia modelu można spełnić w przypadku badania rynku nieruchomości w procesie wyceny. W szczególności, trudno spełnić jest założenie, że pomiędzy cechami rynkowymi na rynku nieruchomości nie występuje współliniowość. Taka zależność pomiędzy cechami występuje nagminnie. Nieruchomości położone w dobrej lokalizacji mają na ogół dobre otoczenie, wysoki standard, zapewnioną ochronę osiedli, itp. W dobrych lokalizacjach nabywają nieruchomości ludzie bogaci, których stać na ochronę czy wysoki standard wykończenia. W lokalizacjach o złej renomie cechy rynkowe nieruchomości rzadko kiedy posiadają inne cechy na wysokim poziomie. Rozkład cen na rynkach nieruchomości nie zawsze zbliżony jest do rozkładu normalnego. Wnioskowanie statystyczne opiera się na rozkładach teoretycznych. Dokładność obliczeń zależy od tego jak teoretyczne modele odległe są od rzeczywistości, którą modelujemy. Jak wynika z książki Aczela „Statystyka w zarządzaniu” – do ważniejszych skutków nie spełnienia założeń modelu w metodzie regresji wielorakiej zaliczyć należy:

- 1) wariacje estymatorów i standardowe błędy ocen współczynników są nadmierne,
- 2) wartości współczynników regresji mogą bardzo różnić się od oczekiwanych,

- 3) znaki współczynników regresji są odmienne od oczekiwanych,
- 4) włączenie lub wyłączenie jednej zmiennej objaśniającej z modelu powoduje wielkie zmiany w ocenach współczynników regresji lub zmianę znaków,
- 5) usunięcie pewnych wyników obserwacji powoduje duże zmiany w wartościach ocen współczynników regresji lub zmianę ich znaków.

Zastosowane formalne testy, jak np. R^2 nie zawsze wystarczają do oceny modelu, co więcej – jeśli liczba obserwacji (cen transakcyjnych) niewiele przekracza liczbę cech rynkowych – test R^2 będzie zawsze zbliżony do jedności. Wartość ta jednak nie wystarcza do oceny, że model daje zadawalające wyniki obliczeń. Jak wynika ponadto z książki pt. „Przewodnik po finansach”¹ (str. 317) w punkcie dotyczącym testów stosowanych w regresji wielorakiej, istotnym zabiegiem pozwalającym ocenić przygotowane równania jest zadawanie pytań o sens otrzymanego modelu. Autorzy piszą wprost cyt.: „**Równanie, które nie ma sensu z intuicyjnego czy teoretycznego punktu widzenia należy odrzucić**”.

Zacząłem krytycznie pisać o zastosowaniu regresji wielorakiej do wyceny konkretnych nieruchomości wówczas, gdy przy pomocy matematyków przeprowadziłem wiele symulacji wycen na konkretnych danych. Wyniki wycen symulacyjnych oraz przykłady wycen dla konkretnych celów wielokrotnie potwierdzały, że model regresji wielorakiej nie nadaje się do wyceny nieruchomości. Właściwie wszystkie skutki nie spełnienia założeń modelu w metodzie regresji wielorakiej występują, gdy usiłujemy oszacować wartość nieruchomości metodą regresji wielorakiej:

- to, że wartości współczynników regresji mogą bardzo różnić się od oczekiwanych oznacza, że cecha nieistotna okazuje się decydująca,
- to, że znaki współczynników regresji są odmienne od oczekiwanych oznacza, że im lepszą dajemy ocenę nieruchomości tym mniejszą otrzymujemy wartość i odwrotnie,
- włączenie lub wyłączenie jednej zmiennej objaśniającej z modelu powoduje wielkie zmiany w ocenach współczynników regresji lub zmianę znaków, co oznacza diametralnie inną wartość,
- podobnie, gdy usuniemy pewne wyniki obserwacji może zajść całkowita zmiana równania.

Jest również i tak, że jeśli do modelu użyjemy większej liczby zmiennych objaśniających (cech nieruchomości), to z równania regresji wielorakiej otrzymamy wyniki całkowicie sprzeczne ze stanem rynku. Wynika to także z faktu, że w modelu znajdzie się większa liczba cech ze sobą skorelowanych.

2. Przykłady nietrafnych wycen

Poniżej zamieszczono przykłady wycen, z którymi zetknąłem się bezpośrednio. Świadczą one dobitnie, że programy komputerowe potrafią obliczyć wartości, ale wiarygodność zależy od tego czy zastosowane wzory spełniają założenia modelu danego zjawiska.

■ Wycena z roku 1997 – 14 domów wycieczkowych FWP z rejonu Polanicy Zdrój i Kudowy na podstawie 38 transakcji według wzoru:

$$Y = 93,66 + 26,91X_1 + 1,77X_2 + 97,5X_3 + 16,6X_4 - 76,3X_5$$

gdzie:

Y – wartość w tys. zł.,

X_1 – lokalizacja; (1 średnia, 2 dobra, 3 b. dobra),

X_2 – stan techniczny; (1 do remontu, 2 średni, 3 dobry, 4 b. dobry),

X_3 – standard wykończenia; (1 zły, 2 średni, 3 dobry),

X_4 – kategoria pensjonatu; (1 zła, 2 średnia, 3 dobra),

X_5 – działka (tytuł prawny); (1 inny tytuł prawny, 2 użytkowanie wieczyste, 3 własność).

Zastosowany wzór wskazuje, że wyższą wartość będą miały pensjonaty z gorszym prawem do działki niż te, które mają prawo własności. Tak określone nieruchomości w przeliczeniu na 1m² budynków otrzymały ok. 45 zł, gdy średnia z bazy porównawczej wynosiła ponad 800 zł/m².

■ Opinia sądowa wyceny lokalu z roku 2010 przy pomocy następującego wzoru:

$$Y = -289840 + 8144,61 \times \text{powierzchnia lokalu} + 26915,41 \times \text{stan techniczny} + 57935,08 \times \text{standard} + 36689,63 \times \text{lokalizacja} - 27254,92 \times \text{garaż}$$

gdzie:

Y – wartość lokalu;

Powierzchnia lokalu w m²;

Stan techniczny: 1 – zły, 2 – słaby, 3 – średni, 4 – dobry.

Standard: 1 – niski, 2 – przeciętny, 3 – wysoki.

Lokalizacja; (1 najgorsza – 4 najlepsza).

Garaż: 0 – brak, 1 – jest miejsce garażowe jako pomieszczenie przynależne do lokalu.

Zastosowany wzór wskazuje, że wyższą wartość otrzymają takie same mieszkania bez miejsca garażowego niż mieszkania z miejscem garażowym. Jednocześnie podstawiając do wzoru parametry mieszkania o pow. do 24 m² o złych ocenach otrzymamy wartość ujemną.

■ Ostatnim zamieszczonym tu przykładem niech będzie wycena nieruchomości zabudowanej w Strzelcach Opolskich zamieszczona w Internecie – sporządzona na podstawie, jak napisano: arkuszy kalkulacyjnych opartych o mechanizm regresji wielorakiej autorstwa Tomasa Kotrańskiego. Na stronie 27 operatu podano cechy rynkowe jako zmienne objaśniające:

X_1 – stan techniczny; (od 1 – zły do 5 bardzo dobry),

X_2 – lokalizacja; (od 1 – słaba do 3 bardzo dobra),

X_3 – powierzchnia działki (cecha ilościowa),

X_4 – powierzchnia użytkowa (cecha ilościowa),

X_5 – rodzaj praw do nieruchomości (użytkowanie wieczyste – 1, własność – 2).

Na podstawie 22 transakcji określono równanie regresji:

$$Y = -304,65 - 0,01 \times \text{powierzchnia użytkowa} + 0,01 \times \text{powierzchnia działki} + 116,77 \times \text{rodzaj prawa} + 462,2 \times \text{stan techniczny}$$

Równanie to rozpatrywane z punktu widzenia zasad kształtowania się cen jednostkowych jest dość nielogiczne. Z równania wynika, że im większa powierzchnia działki, tym wartość jednostkowa rośnie. Jest na ogół odwrotnie. Typowa zamiana znaków. Ponadto zmiana stanu technicznego o jeden stopień w skali od 1 do 5 powoduje przyrost wartości 1 m² aż o 4622 zł. Pobieżna analiza 22 transakcji przeczy takiemu wpływowi tej cechy na ceny.

Niestety, błędne wzory bez ich oceny merytorycznej znaleźć można także w kilku publikacjach. Należy zauważyć także, że modele stochastyczne są bardziej pracochłonne. Z opisu programu Statistica, który jest dostępny i stosowany w metodzie regresji wielorakiej wynika, że minimalna liczba obserwacji (transakcji), powinna wynosić od 10 do 20 razy więcej niż liczba zmiennych objaśniających, czyli cech rynkowych. Taki wymóg wyraźnie jednak ogranicza możliwości stosowania metody regresji wielorakiej na rynkach lokalnych.

Wnioski

Przeprowadzona krytyka modeli stochastycznych do wyceny nieruchomości dotyczy przede wszystkim wycen konkretnych nieruchomości dla większości celów wyceny. Istnieje na szczęście wiele obszarów możliwych zastosowań modeli stochastycznych w analizie rynków nieruchomości, w tym także jako narzędzie pomocnicze w procesie wyceny. Problem wykorzystywania modeli statystycznych do wyceny nieruchomości, jako zabezpieczenia wiarygodności, był jednym z punktów Rozporządzenia Parlamentu Europejskiego i Rady (UE) nr 575/2013 z dnia 26 czerwca 2013 r. w sprawie wymogów ostrożnościowych dla in-

stytucji kredytowych i firm inwestycyjnych. Europejski Urząd Nadzoru Bankowego (EUNB) wydał odpowiedzi dotyczące wykorzystywania modeli statystycznych do wycen nieruchomości na potrzeby zabezpieczenia wierzytelności. W jednym z punktów odpowiedzi stwierdzono, że cyt.: *Artykuł 208(3)(b) CRR nie pozwala na wykorzystanie modelu statystycznego jako jedyne go sposobu wykonywania aktualizacji wyceny nieruchomości.*

W trakcie Zgromadzenia Ogólnego TEGoVA w Mar-sylii 28 października 2017 r. delegaci, reprezentujący 71 stowarzyszeń rzeczoznawców (w tym PFSRM) z 36 krajów zatwierdzili, nowy Europejski Standard Wyceny (EVS 6), dotyczący wykorzystania opartych na statystyce Automatycznych Modeli Wyceny (AVM).

Główne przesłanie standardu jest następujące:

Automatyczne Modele Wyceny nie mogą być wykorzystywane do sporządzania raportów z wyceny zgodnych z EVS w oderwaniu od procesu wyceny opartego, między innymi, na oględzinach nieruchomości oraz ocenach dokonywanych przez rzeczoznawcę.

W przypadkach, gdy jest wykorzystywany Automatyczny Model Wyceny, nie jest nigdy niczym więcej, jak tylko narzędziem wspomagającym szacowanie wartości przez rzeczoznawcę, który jest za ten szacunek odpowiedzialny.

Bibliografia

- Aczel A.D., Statystyka w zarządzaniu, PWN, Warszawa 2000.
- Bitner A., Statystyka opisowa w wycenie nieruchomości, Część I – wyznaczanie miar zbioru danych, Rzeczoznawca majątkowy Nr 1 (77) Wyd. PFSRM, Warszawa 2013.
- Bitner A., Statystyka opisowa w wycenie nieruchomości część II – graficzna prezentacja danych, Rzeczoznawca majątkowy Nr 2 (78) Wyd. PFSRM, Warszawa 2013.

Streszczenie

Niektórzy rzeczoznawcy, którzy z uporem godnym lepszej sprawy usiłują wyceniać nieruchomości metodami statystycznymi, mogą odpowiedzieć, że pytanie w tytule artykułu jest retoryczne. Mogliby odpowiedzieć także, że statystykę wymyślono m. in. po to, by można było łatwo i przyjemnie wyceniać nieruchomości. Są jednak w błędzie. Nie po to wymyślono statystykę. Postaram się to uzasadnić.

Słowa kluczowe:

korelacja, podejście porównawcze, regresja prosta, regresja wieloraka, statystyka.

- Prystupa M., Wycena nieruchomości i przedsiębiorstw w podejściu porównawczym, Replika, Zakrzewo 2014.
- Prystupa M., Rzeczoznawstwo majątkowe czy zabawa liczbami, Rzeczoznawca majątkowy Nr 2(90) Wyd. PFSRM, Warszawa 2016.
- Siegel J.G., Shim J.K., Hartman S. W., Przewodnik po finansach PWN, Warszawa 1995.

Przypisy:

1. Siegel J.G., Shim J.K., Hartman S.W., Przewodnik po finansach PWN, Warszawa 1995.

Prof. dr hab. inż. Mieczysław Prystupa jest szefem Warszawskiego Stowarzyszenia Rzeczoznawców Majątkowych i redaktorem naczelnym kwartalnika „Wycena Nieruchomości i Przedsiębiorstw”.

Artykuł naukowy recenzowany.



Summary

To some appraisers, who are obstinate enough to try to price the property just with statistical methods, the question in the title of the article may sound rhetorical. They could also answer that statistics was invented, among other reasons, to make valuation easy and pleasant. But they are wrong. This is not the case for statistics and I will attempt to justify it.

Key words:

correlation, comparative approach, simple regression, multiple regression, statistics.