

КРИТЕРІЇ ОЦІНКИ ЯКОСТІ ЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ДЛЯ ЗАДАЧІ ВИЗНАЧЕННЯ ВАРТОСТІ НЕРУХОМОСТІ

Вінницький національний технічний університет

Анотація

Розкрито поняття регресійного аналізу та якості регресійної моделі. Розглянуто такі критерії оцінки якості парної та множинної лінійної регресійної моделі, як коефіцієнт кореляції, коефіцієнт детермінації, середньоквадратична похибка та показник середньої похибки апроксимації.

Ключові слова: лінійна регресія, регресійний аналіз, оцінка якості регресійної моделі, критерії оцінки якості регресійної моделі.

Abstract

The concept of regression analysis and regression model quality are revealed. The following criteria for assessing the quality of a pair and multiple linear regression model, such as correlation coefficient, determination coefficient, mean square error and average approximation error, are considered.

Keywords: linear regression, regression analysis, estimation of the quality of the regression model, criteria for assessing the quality of the regression model.

Визначення вартості нерухомості є досить актуальною проблемою в наш час. Кожного дня з'являється велика кількість пропозицій по продажу чи оренді житла. Багато власників нерухомості витрачають час та кошти на співпрацю з ріелторами або визначають ціну самостійно, але при цьому не завжди правильно і, як наслідок, втрачають певні кошти при недооцінці житла, або час – у випадку завищеної ціни.

Регресійний аналіз є основним статистичним методом побудови математичних моделей об'єктів або явищ по експериментальним даним при вирішенні задачі визначення вартості нерухомості. Ці моделі зв'язують кількісні змінні – результуючу і пояснюючі. Основним завданням регресійного аналізу являється отримання оцінок параметрів регресії, які були б оптимальними в певному сенсі. Отримані оцінки дозволяють вирішувати задачу оцінки (відновлення) регресії і її прогнозу [1].

Якістю моделі регресії називається адекватність побудованої моделі вихідним (спостережуваним) даним.

Якість лінійної моделі парної регресії характеризується за допомогою таких критеріїв: парний лінійний коефіцієнт кореляції та коефіцієнт детермінації.

Парний лінійний коефіцієнт кореляції характеризує ступінь щільності зв'язку між пояснюючими змінними. Він розраховується тільки для кількісних змінних. Чим ближче модуль значення коефіцієнта кореляції до одиниці, тим щільнішим є зв'язок між досліджуваними змінними. Даний коефіцієнт змінюється в межах $[-1; +1]$. Якщо значення коефіцієнта кореляції знаходиться в межах $[-0; 1]$, то зі збільшенням незалежної змінної збільшується і залежна змінна, і навпаки. Якщо коефіцієнт кореляції знаходиться в межах $[-1; 0]$, то зі збільшенням незалежної змінної зменшується залежна змінна, і навпаки. Якщо коефіцієнт кореляції дорівнює нулю, то зв'язок між змінними відсутній [2].

Парний лінійний коефіцієнт кореляції розраховується за формулою [3]:

$$r_{yx} = \frac{\overline{xy} - \bar{x} * \bar{y}}{G(x) * G(y)}$$

де x – пояснююча (незалежна) змінна, y – результуюча (залежна) змінна, $G(x)$ – середньоквадратичне відхилення незалежної змінної, $G(y)$ – середньоквадратичне відхилення залежної змінної.

Ще одним критерієм, за допомогою якого також вимірюється щільність зв'язку між двома або більше показниками та перевіряється адекватність побудованої регресійної моделі дійсності є коефіцієнт детермінації. Коефіцієнт детермінації показує, чи дійсно зміна значення у лінійно залежить саме від зміни значення x , а не відбувається під впливом різних випадкових факторів [2].

Коефіцієнт детермінації (R^2) – найбільш поширена статистика для оцінки якості моделі. Коефіцієнт детермінації розраховується за такою формулою [3]:

$$R^2 = \frac{\sigma_{\text{регр}}^2}{\sigma_y^2},$$

де $\sigma_{\text{регр}}^2$ – дисперсія, що пояснює регресію, σ_y^2 – дисперсія помилок.

Якість лінійної моделі множинної регресії характеризується за допомогою показників, побудованих на основі теореми про розкладання дисперсій (загальна дисперсія залежної змінної може бути розкладена на дві частини: дисперсію, що пояснює регресію, та дисперсію помилок).

Якість лінійної моделі множинної регресії характеризується за допомогою таких критеріїв:

- множинний коефіцієнт кореляції;
- множинний коефіцієнт детермінації;
- середньоквадратична похибка моделі регресії;
- показник середньої похибки апроксимації.

Властивості множинного коефіцієнта кореляції та детермінації аналогічні властивостям парного коефіцієнта кореляції та детермінації.

Середньоквадратична похибка моделі регресії (MSE) розраховується за формулою [3]:

$$MSE = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n - h}},$$

де h – кількість параметрів моделі, n – кількість елементів вибірки, x_i – i -ий елемент вибірки.

Якщо показник середньоквадратичної похибки виявиться меншим за показник середньоквадратичного відхилення спостережуваних значень залежної змінної від модельних значень – то модель регресії можна вважати якісною.

Показник середньої похибки апроксимації розраховується за формулою [3]:

$$\bar{A}_y = \frac{1}{n} \sum_{i=1}^n \frac{|y - \tilde{y}_i|}{y_i},$$

де n – кількість елементів вибірки, y – результуюча (залежна) змінна.

Якщо величина даного показника становить менше 6–7%, то якість побудованої моделі регресії вважається хорошою, а максимально припустимим відхиленням є 12–15% [3].

В ході аналізу критеріїв для оцінки якості регресійної моделі було розглянуто критерії якості моделі для випадку парної та множинної регресії. Оскільки вартість нерухомості не оцінюється лише за одним параметром, доцільно використати критерії оцінки якості моделі саме для випадку множинної регресії, а саме: середньоквадратичну похибку та показник середньої похибки апроксимації, а також коефіцієнт детермінації, який перевіряє адекватність (відповідність) побудованої регресійної моделі дійсності.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Літнароч Р. М. Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу / Р. М. Літнароч // Навчальний посібник, МЕНУ, Рівне. – 2011. – с. 6.
2. Коефіцієнт кореляції та коефіцієнт детермінації – [Електронний ресурс]. – Режим доступу: <http://www.prognoz.ru/blog/platform/kriterii-otsenki-modeli/>
3. Характеристика якості моделі регресії – [Електронний ресурс]. – Режим доступу: https://www.e-reading.club/chapter.php/1002275/18/Yakovleva_Angelina_-_Otvety_na_ekzamenacionnye_biletu_po_ekonometrike.html

Переродов Артемій Олексійович — студент групи 2КН–146, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця.

Науковий керівник – **Петришин Сергій Іванович**, – асистент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

Pererodov Artemii O. — Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia.

Supervisor — **Petrishyn Sergiy I.** — assistant of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia.