

«Сейфуллин оқулары – 18: « Жастар және ғылым – болашаққа көзқарас» халықаралық ғылыми -практикалық конференция материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 18: « Молодежь и наука – взгляд в будущее» - 2022.- Т.І, Ч.ІV. - С. 3-6

ОСОБЕННОСТИ ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ СТОИМОСТИ ЖИЛЬЯ

Ақан І. магистрант І курса

Казахский агротехнический университет им. С.Сейфуллина , г. Нур-Султан

Введение.

Известно, что жилье является очень значимым благом для каждого члена общества. Оно позволяет ему жить достойно, успешно работать, всесторонне развиваться. Аспекты наличия жилья у каждого человека который желает иметь как его собственность чрезвычайно важны и они непрерывно развиваются.

В то же время возможность приобретения или построения жилья человеку является показателем развитости экономики государства и социальной обеспеченности его жителей. Здесь очень важную роль играет стоимость жилья. Именно стоимость квартиры(жилья) является индикатором ее доступности на которого влияют множество факторов часть которых можно измерять и тех факторов влияние которых очень трудно фиксировать. Тем не менее выявлять прогнозных оценок стоимости жилья важно и для покупателей и для строительного бизнеса. На это ориентируется социальная политика правительства. Динамичное изменение экономических и социальных процессов диктует более тщательного анализа и прогнозирования состояния рынка жилья.

Целью данной статьи является изложение основных принципов построения статистических моделей стоимости жилья. Для этого рассматриваются особенности корреляционно-регрессионного и факторного анализа, возможности их использования для исследования рынка жилья и вычисления параметров статистической модели для прогнозных оценок стоимости жилья.

В данной работе произведен анализ стоимости квартир из газеты “Из рук в руки” за период с декабря 2016 г. по сентябрь 2017 г. Была выбрана Юго-Западная часть крупного мегаполиса, в которой высок спрос на жилые площади.

На стоимость квартир (Y) влияет ряд факторов, наиболее важными из которых являются: удаленность от центра (X_1), общая площадь (X_2), площадь кухни (X_3), площадь комнаты (X_4), категория дома(кирпичный или нет) (X_5).

Для анализа была сделана выборка из 55 квартир. Исходные данные представлены в табл. 1 [1]

Таблица 1. Исходные данные выбранных факторов

N	Улица	Удаленность от центра, км, X1	Общая площадь, м/кв, X2	Площадь кухни, м/кв, X3	Площадь комнаты, кв/м, X4	категория, 1-кирпич, 0-некирпич, X5	Цена, тыс \$, Y
1	Фрунзенская	4	34	7,5	19	1	54
2	Ленинский проспект	5,7	36	10	20	0	35
3	Ленинский проспект	5,7	45	13	20	1	59
4	Академическая	7,6	35,3	10	20	0	35
5	Университет	8,7	33	5,5	22	0	33
.....
54	Юго-Запад	13,3	30	7	19,8	0	34
55	Юго-Запад	13,3	34	9	19	0	42

Оцениваемая модель носит линейный характер и описывается уравнением:

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 + a_4 \cdot X_4 + a_5 \cdot X_5 + \varepsilon t.$$

Работа предполагает решение следующих задач:

- анализ коэффициентов корреляции между результирующим признаком Y и каждым из факторов X_j и оценка статистической значимости коэффициентов корреляции;
- расчет параметров регрессий для всех факторов X ;
- выбор лучшей модели с помощью оценки качества каждой модели через коэффициент детерминации;

1. Анализ коэффициентов корреляции между результирующим признаком Y и каждым из факторов X_j и оценка статистической значимости коэффициентов корреляции.

Используя функцию Анализ данных/корреляция из вкладки «Данные», получим матрицу коэффициентов парной корреляции между всеми имеющимися переменными [1]:

$|r(Y, X1)|=0,479 < 0,6$, следовательно, эта зависимость средняя.
 $|r(Y, X2)|= 0,726 > 0,6$, следовательно, эта зависимость тесная.
 $|r(Y, X3)|= 0,465 < 0,6$, следовательно, эта зависимость средняя.
 $|r(Y, X4)|= 0,545 < 0,6$, следовательно, эта зависимость средняя.
 $|r(Y, X5)|= 0,64 > 0,6$, следовательно, эта зависимость тесная.

2. Отбор факторов в модель множественной регрессии

Отбор факторов, включаемых в регрессию – один из важнейших этапов построения модели регрессии.

Для построения модели используем функция Анализ данных/регрессия во вкладке «Данные» [1]. В качестве входного интервала покажем значение факторов X.

Из полученной таблицы[1] извлекаем коэффициенты и составляем уравнение модели: $Y=-9+(-0,1 * X1)$. Коэффициент регрессии $a1=0,1$ следовательно, при увеличении удаленности от центра на 1 км. стоимость квартиры в среднем уменьшается на 100 долларов.

Аналогичные расчеты проводим для остальных факторов.

Коэффициент регрессии модели X2: $a2=0,637$.

Коэффициент регрессии модели X3: $a3=1,58$.

Коэффициент регрессии модели X4: $a4=0,6$.

Коэффициент регрессии модели X5: $a5=10,8$,

Для проверки значимости найденных коэффициентов корреляции используем критерий Стьюдента. Для каждого коэффициента $r(Y, Xj)$ вычислим t-статистику.

С помощью функции СТЬЮДРАСПОБР при уровне значимости $\alpha = 5\%$ и числе степеней свободы $k = n - 2 = 55 - 2 = 52$ определим критическое значение $t_{кр.} = 2$. Сопоставим фактические значения статистик с критическим значением $t_{кр.}$ и сделаем соответствующие выводы:

$t(r(Y, X1)) = 0,35 < t_{кр.} = 2$, т.е. не является значимым.

$t(r(Y, X2)) = 2,34 > t_{кр.} = 2$, т.е. является значимым.

$t(r(Y, X3)) = 2,65 > t_{кр.} = 2$, т.е. является значимым.

$t(r(Y, X4)) = 1,57 < t_{кр.} = 2$, т.е. не является значимым.

$t(r(Y, X5)) = 4,54 > t_{кр.} = 2$, является значимым.

Таким образом, можно сделать вывод:

Переменные X2, X3, X5 являются значимыми переменными. Следовательно переменные X1 и X4 можно исключить.

На основе полученных данных, можно составить модель:
$$Y = -9 + 0,637 * X2 + 2,65 * X3 + 4,54 * X5$$

3. Оценка качества модели

Проверка значимости уравнения регрессии через коэффициент детерминации.

Коэффициент детерминации и коэффициент множественной корреляции выдается протоколом в регрессионной статистике.

Множественный R	0,87
R-квадрат	0,75

В многофакторной регрессии добавление дополнительных объясняющих переменных увеличивает коэффициент детерминации. Следовательно, он должен быть скорректирован с учетом числа независимых переменных.

$$R2 = 1 - (1 - R2) * (n - 1) / (n - k - 1)$$

Скорректированный R-квадрат также выдается в протоколе в регрессионной статистике:

Нормированный R-квадрат	0,74
-------------------------	------

Коэффициент детерминации показывает долю вариации результативного признака, находящего под воздействием изучаемых факторов, т.е. определяет какая доля вариации признака Y учтена в модели и обусловлена влиянием на него факторов. Чем ближе R-квадрат к 1, тем выше качество модели.

В итоге анализ эконометрической модели показал, что на формирование стоимости квартиры наибольшее влияние оказывают категория дома, общая площадь и площадь кухни. Было доказано, что эти факторы являются значимыми. Проверка осуществлялась при помощи коэффициентов корреляции и детерминации. Для оценки значимости использовалась t-статистика Стьюдента. Также был приведен пример составления прогноза по модели. И данная модель может быть использована заинтересованными лицами для прогнозирования стоимости желаемой квартиры.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием

пакетов STATISTICA и EXCEL: учебное пособие. — 2-е изд., испр. и доп. — М.: ФОРУМ, 2008. — 464 с. (Высшее образование).

2. Бакушева Г.В. Основы эконометрики: решение задач шаг за шагом [Текст]: учебное пособие / Г. В. Бакушева. — Йошкар-Ола : Стринг, 2013. — 413 с.

3. Бывшев В.А. Эконометрика: учеб. пособие / В.А. Бывшев. — М.: Финансы и статистика, 2008. — 480 с.

4. Шалабанов А.К., Роганов Д.А. Практикум по эконометрике с применением MS EXCEL. Казань: Академия управления «ТИСБИ», 2008. — 53 с.

5. Айвазян С.А. Методы эконометрики: учебник. — М.: Магистр: ИНФРА-М, 2010.

6. Дайитбегов Д.М. Компьютерные технологии анализа данных в эконометрике: Монография. — М.: ИНФРА-М; Вузовский учебник, 2008.