**Квартиры. Трубчик А.И.**

Исходный код: <https://github.com/course-bd-and-pa-bsuir/r-statistics/blob/master/programmerby/05_flats.R>

Необходимо исследовать зависимость стоимости жилья г.Минска от следующих параметров: общей площади, площади квартиры, кухни, количества комнат, тип дома, этаж, месторасположения квартиры, год постройки, элитность, наличие телефона и лоджии.

Исходные данные были дополнены гео информацией (спасибо за это Сергею Кушмару), используя адреса квартир, что позволило ввести два новых параметра — расстояние до центра и расстояние до ближайшей станции метро.

Наилучшего результата удалось добиться при создании модели для каждого отдельного района города и использовании в качестве параметров жилую площадь ln(PlZ), площадь кухни ln(PlK), остальную площадь квартиры ln(PlOb - PlZ - PlK) и расстояние до центра Минска ln(CenterDistance):

Таблица 1 — Модели регрессии

|  |  |
| --- | --- |
| Район | Модель |
| Заводской | ln(Cena) = 9,61 + 0,0767ln(PlOb - PlZ - PlK) + 0,384ln(PlZ) + 0,516ln(PlK) - 0,13ln(CenterDistance) |
| Ленинский | ln(Cena) = 11,6 + 0,326ln(PlZ) + 0,804ln(PlK) - 0,384ln(CenterDistance) |
| Московский | ln(Cena) = 8,96 + 0,338ln(PlOb - PlZ - PlK) + 0,461ln(PlZ) + 0,526ln(PlK) - 0,153ln(CenterDistance) |
| Октябрьский | ln(Cena) = 8,99 + 0,11ln(PlOb - PlZ - PlK) + 0,585ln(PlZ) + 0,78ln(PlK) - 0,209ln(CenterDistance) |
| Первомайский | ln(Cena) = 8,43 + 0,278ln(PlOb - PlZ - PlK) + 0,458ln(PlZ) + 0,573ln(PlK) - 0,0764ln(CenterDistance) |
| Советский и Партизанский | ln(Cena) = 9,49 + 0,214ln(PlOb - PlZ - PlK) + 0,597ln(PlZ) + 0,338ln(PlK) - 0,185ln(CenterDistance) |
| Фрунзенский | ln(Cena) = 8,29 + 0,14ln(PlOb - PlZ - PlK) + 0,497ln(PlZ) + 0,359ln(PlK) |
| Центральный | ln(Cena) = 9,71 + 0,466ln(PlOb - PlZ - PlK) + 0,38ln(PlZ) + 0,361ln(PlK) - 0,202ln(CenterDistance) |

В процессе создания модели были выявлены выбросы и удалены из выборки. В них входили квартиры с чрезмерно завышенными или заниженными ценами или, возможно, неверными исходными данными (площадь кухни 2 или 20 м2 в однокомнатной квартире и так далее).

Исходя из полученных моделей, во всех районах, кроме Фрунзенского, расстояние от центра Минска отрицательно сказывается на стоимости квартиры.

Статистически незначимыми факторами оказались количество комнат, тип дома (кирпичный или нет), этаж (первый/последний или нет), элитность, расстояние до метро, год постройки, наличие телефона и лоджии. Проверка на мультиколлинеарность выявила, что количество комнат, в том числе введение дискретных фиктивных переменных, сильно связаны с площадью квартиры (VIF > 10) и поэтому не использовались при моделировании.

При попытке объединения данных нескольких районов уточнить модели не удалось, даже с вводом фиктивном переменной для каждого из районов. Модель только ухудшала свои показатели.

Все указанные модели соответствуют критериям качества: все коэффициенты статистически значимы на 5% уровне, т. е. их остатки нормально распределены, не автокоррелированы, не гетероскедастичны — таблица 2.

Таблица 2 — Критерии качества моделей

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Район | *R2* | *F* | *pF* | *pDurbin-Watson* | *pχ2* | *pBreusch-Pagan* |
| Заводской | 0.8147 | 33.97 | 0.000 | 0.568 | 0.2675 | 0.3794 |
| Ленинский | 0.9341 | 171.0 | 0.000 | 0.692 | 0.0540 | 0.7957 |
| Московский | 0.9038 | 76.13 | 0.000 | 0.494 | 0.1588 | 0.8833 |
| Октябрьский | 0.8772 | 52.79 | 0.000 | 0.594 | 0.5304 | 0.4335 |
| Первомайский | 0.9666 | 333.6 | 0.000 | 0.320 | 0.3616 | 0.0628 |
| Советский и Партизанский | 0.9179 | 157.5 | 0.000 | 0.274 | 0.9333 | 0.2453 |
| Фрунзенский | 0.8717 | 173.2 | 0.000 | 0.336 | 0.1062 | 0.5397 |
| Центральный | 0.8505 | 82.53 | 0.000 | 0.486 | 0.4384 | 0.0569 |