УДК 004

Исследование зависимости жилых мест в приморье Лондона и загородом Лондона с использованием линейной регрессии в системе R

Беляев Алексей Андреевич
Приамурский государственный университет им. Шолом-Алейхема
студент

Баженов Руслан Иванович
Приамурский государственный университет им. Шолом-Алейхема
к.п.н., доцент, зав. кафедрой информационных систем, математики и правовой информатики

Аннотация
В данной статье исследуется зависимость жилых мест в приморье Лондона и загородом Лондона с использованием интеллектуального метода линейной регрессии в системе R. Зависимость использованных данных является плохой.
Ключевые слова: R, система, RStudio, регрессия, линейная, интеллектуальный метод, Лондон, приморье, загород, жилые места

The study of the dependence of residential areas in the coastal area of London and the city of London using linear regression in the system R

Belyaev Alexey Andreevich
Sholom-Aleichem Priamursky State University
student

Bazhenov Ruslan Ivanovich
Sholom-Aleichem Priamursky State University
Candidate of pedagogical sciences, associate professor, Head of the Department of Information Systems, Mathematics and Legal Informatics

Abstract
This article examines the dependence of residential areas in the coastal area of London and the city of London using the intellectual method of linear regression in the system R. The dependence of the data used is poor.
Keywords: R, system, RStudio, regression, linear, intellectual method, London, seaside, countryside, residential areas.

Большинство фирм, корпораций, учреждений, предприятий имеют огромное количество данных, которые каждый год меняются в лучшую или худшую сторону и чтобы спрогнозировать эти данные и узнать какими они будут в последующих годах, то используются интеллектуальные методы

статистического анализа данных. В зависимости, какие данные собираемся анализировать, то нам необходимо выбрать для этого эффективный интеллектуальный метод, который даст наиболее существенный и правдивый прогноз самих данных на будущий период и покажет в них зависимость, что лучше, а что хуже будет.

В статье В. В. Стрижкова и Р. А. Сологуба применили алгоритм выбора нелинейных регрессионных моделей [1]. И. Ю. Глухих разработал модель экспресс-анализа финансовой состоятельности организаций на базе методов многомерного регрессионного анализа [2]. В статье Л. М. Бугаевского и Г. Г. Прохорова рассматривается разработка методики составления карт взаимосвязи с использованием корреляционного и регрессионного анализов [3]. В. С. Баджанов и Е. А. Матушевская применили корреляционно-регрессионный анализ для анализа себестоимости продукции на примере ГУП АО "Севастопольский Винодельческий завод" [4]. В статье А. А Манцаева. проведен анализ долговременных тенденций производительности труда в РК: корреляционно-регрессионный анализ [5]. О. Ю. Легкодух, Д. Д. Капустина и Т. А. Кокодей провели анализ и прогноз динамики курса доллара, используя инструментарий регрессионного анализа [6]. В статье Н. Х. Рашитова был проведен анализ эффективности структуры экономики на основе корреляционно-регрессионного анализа [7]. R. Boukezzoula, S. Galichet и D. Coquin в своей статье рассматривали анализ параметрических интервальных регрессионных методологий в соответствии с онтологическими и эпистемическими видениями интервалов [8]. В статье F. O. de França исследуется математическое выражение, описывающее взаимосвязь набора объясняющих переменных с измеряемой переменной [9]. I. Georgiev, D. I. Harvey, Stephen J. Leybourne, A.M. и R. Taylor рассмотрели тесты на структурные изменения, основанные на статистических данных типа SupF и Cramer-von-Mises Andrews и Nyblom [10].

Целью данной статьи является исследование зависимости жилых мест в приморье и загородом с использованием метода линейной регрессии в системе R.

В нашем случае данными будут являться жилые места в приморье и загородом Лондона. Данные были взяты с официального сайта баз данных Лондона в период времени с 2013 по 2018 год включительно. В данной статье рассматриваются данные по переселению из приморья и загорода Лондона [11].

Для анализа и выявление зависимостей данных рабочих мест используется метод под названием «Метод линейной регрессии». Данный метод позволяет построить регрессионную модель, в которой можем увидеть зависимость этих данных за каждый год и как они менялись, а также построить саму линию регрессии, которая покажет, что с данными будет происходить дальше, то есть какой будет вероятный их прогноз и в каком промежутке.

Метод линейной регрессии использует уравнение линейной регрессии $y = a+bx$ (это уравнение похоже на уравнение прямой), где y – зависимая

переменная, a – свободный член (пересечение) линии оценки, b – угловой коэффициент или градиент оценённой линии, x – независимая переменная [12].

Чтобы узнать, насколько хороша связь между двумя переменными, то вводится коэффициент детерминации (R^2). Чтобы вычислить этот коэффициент вручную, то потребуется множество формул. Первая формула это вычисления средних значений каждой переменной x и y и сразу обеих. Она вычисляется суммой всех значений деленных на их количество. Далее вычисляется дисперсия: сумма каждого значения возведенного в квадрат делится на их количество и вычитается средним значением, возведенным в квадрат. Потом вычисляется среднеквадратичное отклонение равное квадратному корню дисперсии. После вычисляется сам коэффициент корреляции в виде дроби: в числитель дроби идет разница между средним значением двух переменных x и y и умноженных средних значений каждой переменной x и y, и в знаменатель идет умножение среднеквадратичных отклонений. Ну и напоследок при найденном коэффициенте корреляции, можно уже найти сам коэффициент детерминации R^2 он равен всего лишь коэффициенту корреляции, возведенному в квадрат [13].

Инструмент, который понадобится для реализации и использования метода линейной регрессии на данные, называется язык программирования R, который в основном предназначен для статистического анализа и обработки данных, а также в нем есть поддержка построение различных графиков, и он является свободной средой, в которой можно проводить различные вычисления. Программное обеспечение, на котором понадобится использовать данный метод называется «RStudio», сама программа обрела популярность, благодаря ее понятно интуитивному интерфейсу и в ней гораздо удобнее работать, так как она имеет ряд инструментов, которые понадобятся для статистического анализа. В язык программирования R разработано было большое количество пакетов (библиотек), которые добавляют новые методы обработки данных, а также внедряют новые функции и возможности.

Теперь приступим к использованию самой программы «RStudio» и начнем использовать метод линейной регрессии на данные жилых мест в приморье Лондона и загорода Лондона. Для начала выявим чему равен R^2 (коэффициент детерминации), используя функцию «lm» рабочих мест в первой группе.

```
Call:
lm(formula = New.applications ~ New.applications2, data = b)

Residuals:
      1       2       3       4       5
 -2.179 -25.235 -48.454  21.083  54.786

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       320.59554  124.17582   2.582   0.0817 .
New.applications2   0.07458    0.03742   1.993   0.1403
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 46.32 on 3 degrees of freedom
Multiple R-squared:  0.5697,    Adjusted R-squared:  0.4263
F-statistic: 3.972 on 1 and 3 DF,  p-value: 0.1403
```

Рис 1. Результат функции «lm»

Как видим на рисунке 1, показан результат использования функции «lm» зависимых данных о новых заявлениях на жилье в приморье и загородом Лондона, R^2 = 0,4263, это означает, что если R^2 приближен к 0,5 или выше его, то линейная регрессия этих зависимых данных имеет плохую связь.

К оставшимся данным были выданы результаты R^2 ниже:

- Поданные заявления на выдачу жилых мест в приморье и загородом Лондона =0.1745;
- Сдаваемы в аренду жилые дома в приморье и загородом Лондона = 0.4067.

Далее необходимо наглядно посмотреть зависимости заявлений на выдачу жилья в приморье и загородом с методом линейной регрессии, для этого необходимо построить график.

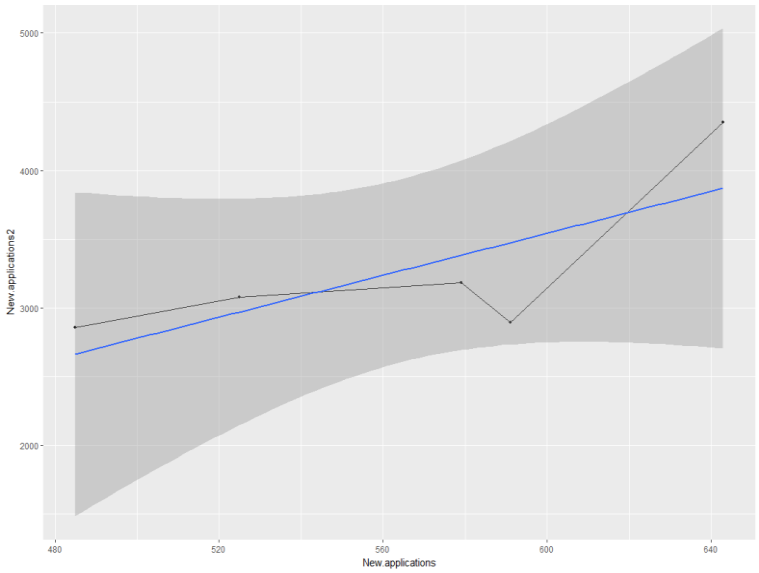

Рис. 2. График зависимости заявлений на выдачу жилья в приморье и загороде Лондона

На рисунке 2 изображен график, в нем видим зависимость данных: о заявлениях на выдачу жилья в приморье и загородом Лондона. Как видно на

рисунке в период подачи заявлений в размере 580 штук произошел спад выдаваемых мест. В дальнейшем спад прошел, и данная регрессия выровнялась, из-за этого спада связь ухудшилась до 0.4.

Далее рассмотрим график зависимости поданных заявлений на выдачу жилья в приморье и загородом Лондона.

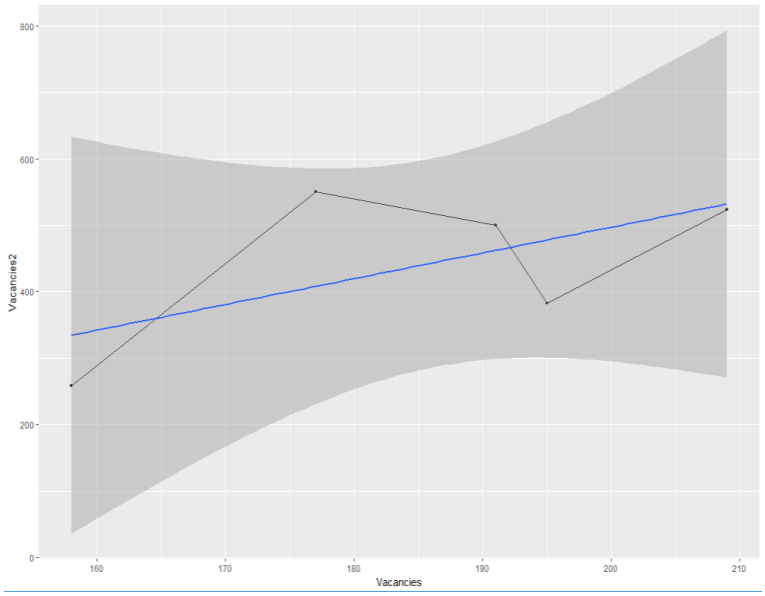

Рис. 3. График зависимости поданных заявлений на выдачу жилья в приморье и загородом Лондона

На рисунке 3 изображен график поданных заявлений на выдачу жилья в приморье и загородом Лондона. По графику видно, что он представляет из себя ломаную прямую, на которой видно резкое возрастание и падение. Это связано с тем, что многие люди перестали писать такого рода заявления. В следствии наша регрессия получилась такой плохой, а точнее почти не имеет связи, и она равна 0.17.

Аналогично построим график сдаваемого в аренду жилья в приморье и загородом Лондона.

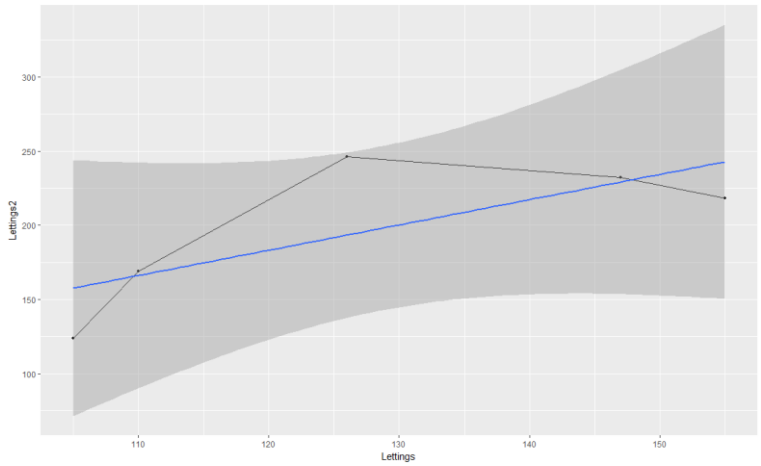

Рис. 4. График зависимости сдаваемого в аренду жилья в приморье и загородом Лондона

На рисунке 4 изображен график сдаваемого в аренду жилья в приморье и загородом Лондона. По графику видно, что он представляет из себя ломаную прямую, на которой видно резкое возрастание и падения потребности на снятие жилья в приморье Лондона. В следствии наша регрессия получилась такой плохой, и она равна 0.4.

Таким образом, в данной статье была достигнута цель с исследованием зависимостей данных о жилых местах в приморье и загородом Лондона с использованием метода линейной регрессии в системе R.Можно судя по графикам увидеть то, что потребности в жилье, а также сдаче жилья в загороде Лондона почти не требуется. Это можно заметить по графикам выше. Прогноз линейной регрессии дает вывод что в скором времени в приморье и загородом Лондона жильем будут обеспечены почти все люди.

Библиографический список

1. Стрижов В.В., Сологуб Р.А. Алгоритм выбора нелинейных регрессионных моделей с анализом гиперпараметров // Математические методы распознавания образов. 2009. Т. 14. № 1. С. 184-187.
2. Глухих И.Ю. Разработка моделей экспресс-анализа финансовой состоятельности организаций на базе методов многомерного регрессионного анализа // Управленческое консультирование. Актуальные проблемы государственного и муниципального управления. 2011. № 3 (43). С. 185-195.
3. Бугаевский Л.М., Прохоров Г.Г. Разработка методики составления карт взаимосвязи с использованием корреляционного и регрессионного анализов // Известия высших учебных заведений. Геодезия и аэрофотосъемка. 1990. № 6. С. 101-109.
4. Баджанов В.С., Матушевская Е.А. Применение корреляционно-регрессионного анализа для анализа себестоимости продукции на примере ГУП АО "Севастопольский Винодельческий завод" // Southern Almanac of Scientific Research. 2017. № 4 (4). С. 20-25.
5. Манцаева А.А. Анализ долговременных тенденций производительности труда в рк: корреляционно-регрессионный анализ // Вестник Института комплексных исследований аридных территорий. 2015. Т. 1. № 1 (30). С. 18-24.
6. Легкодух О.Ю., Капустина Д.Д., Кокодей Т.А. Анализ и прогноз динамики курса доллара, используя инструментарий регрессионного анализа // В сборнике: Развитие методологии современной экономической науки и менеджмента материалы I Всероссийской конференции студентов, аспирантов и молодых учёных. Севастопольский государственный университет. 2016. С. 57-58.
7. Рашитова Н.Х. Анализ эффективности структуры экономики на основе корреляционно-регрессионного анализа // В сборнике: Инновационное развитие российской экономики материалы X Международной научно-практической конференции. Российской Федерации Российский

экономический университет имени Г. В. Плеханова; Российский фонд фундаментальных исследований. 2017. С. 250-252.
8. Reda Boukezzoula, Sylvie Galichet, Didier Coquin From fuzzy regression to gradual regression: Interval-based analysis and extensions // Information Sciences, Volume 441, May 2018, Pages 18-40
9. Fabrício Olivetti de França A greedy search tree heuristic for symbolic regression // Information Sciences, Volumes 442–443, May 2018, Pages 18-32
10. Iliyan Georgiev, David I. Harvey, Stephen J. Leybourne, A.M. Robert Taylor Testing for parameter instability in predictive regression models // Journal of Econometrics, Volume 204, Issue 1, May 2018, Pages 101-118
11. База данных в Лондоне URL: https://data.london.gov.uk/dataset (дата обращения 07.05.2018)
12. Основы линейной регрессии URL: http://statistica.ru/theory/osnovy-lineynoy-regressii (дата обращения 07.05.2018)
13. Пример нахождения коэффициента детерминации URL: https://math.semestr.ru/corel/prim2.php (дата обращения 07.05.2018)
14. М.Е.Кочитов, Р.И. Баженов Исследование зависимости рабочих мест в Лондоне и Великобритании с использованием интеллектуального метода линейной регрессии в системе R // Постулат. 2018. №4