

Вестник БИСТ (Башкирского института социальных технологий). 2024. № 4(65). С. 141–147
Vestnik BIST (Bashkir Institute of Social Technologies). 2024;4(65):141–147

Научная статья

УДК 338

doi: 10.47598/2078-9025-2024-4-65-141-147

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ЗАВИСИМОСТИ ПЕРЕРАБОТКИ НЕФТИ ОТ ЕЕ ДОБЫЧИ

Риф Забирович Гильмутдинов^{1✉}, Любовь Юрьевна Мельник²,
Ляйсан Ринатовна Тухватуллина³

¹Башкирский институт социальных технологий (филиал) Академии труда и социальных отношений, Уфа, Россия

^{2,3}Уфимский государственный нефтяной технический университет, Уфа, Россия

¹gilrif@rambler.ru✉

²lubaleb@mail.ru

³lyaisan_78@mail.ru

Аннотация. Целью работы является выявление статистической зависимости между добычей нефти и ее переработкой и определение наилучшего способа исследования данной зависимости. Хотя экономически понятно, что такое влияние является взаимным, но в данном случае важна точная степень зависимости, а значит, и каким именно уравнением связаны количество переработки нефти и размеры ее добычи. В качестве методов исследования используются алгоритмы линейной, параболической регрессий. Уравнение регрессии позволяет выяснить, каково было бы действие на зависимую переменную главных факторов при неизменности других факторов и при исключении случайных элементов. Далее с помощью уравнения регрессии методами интерполяции и экстраполяции можно воспроизвести значения зависимой переменной как внутри интервала заданных значений независимых переменных, так и вне заданного интервала. Данные за 2007–2018 годы взяты осознанно. В этот период экономика развивалась без резких экономических, политических и других всплесков, что дает больше возможности проследить зависимость добычи нефти от объема переработки.

Ключевые слова: добычи нефти, переработки нефти, функции регрессии, коэффициент и индекс корреляции, линейная и нелинейная корреляция, критерий Фишера, средняя относительная ошибка, прогнозирование

Для цитирования: Гильмутдинов Р. З., Мельник Л. Ю., Тухватуллина Л. Р. Корреляционный анализ зависимости переработки нефти от ее добычи // Вестник БИСТ (Башкирского института социальных технологий). 2024. № 4 (65). С. 141–147. <https://doi.org/10.47598/2078-9025-2024-4-65-141-147>.

Research article

CORRELATION ANALYSIS OF THE DEPENDENCE OF OIL REFINING ON OIL PRODUCTION

Rif Z. Gilmutdinov^{1✉}, Lyubov Yu. Melnik², Laysan R. Tukhvatullina³

¹Bashkir Institute of Social Technologies (branch) of the Academy of Labor and Social Relations, Ufa, Russia

^{2,3}Ufa State Petroleum Technological University, Ufa, Russia

¹gilrif@rambler.ru✉

²lubaleb@mail.ru

³lyaisan_78@mail.ru

Abstract. The aim of the work is to identify the statistical relationship between oil production and its refining and to determine the best way to study this relationship. Although it is economically clear that such an influence is mutual, in this case the exact degree of dependence is important, and therefore what equation

connects the amount of oil refining and the size of its production. Linear and parabolic regression algorithms are used as research methods. The regression equation allows you to find out what the effect of the main factors on the dependent variable would be if other factors remain unchanged and random elements are excluded. Further, using the regression equation, interpolation and extrapolation methods can reproduce the values of the dependent variable both within the interval of specified values of independent variables and outside the specified interval. Data for 2007–2018 are taken deliberately. During this period, the economy developed without sharp economic, political and other surges, which provides more opportunities to trace the dependence of oil production on the volume of refining.

Keywords: oil production, oil refining, regression functions, coefficient and correlation index, linear and nonlinear correlation, Fischer criterion, average relative error, forecasting

For citation: Gilmutdinov R. Z., Melnik L. Yu., Tukhvatullina L. R. Correlation analysis of the dependence of oil refining on oil production. *Vestnik BIST (Bashkirskogo instituta social`ny`x texnologij) = Vestnik BIST (Bashkir Institute of Social Technologies)*. 2024;(4(65)):141–147. (In Russ.). <https://doi.org/10.47598/2078-9025-2024-4-65-141-147>.

Анализ объектов исследования

В экономике нашей страны сырьевые отрасли многие годы определяют бюджет, объемы экспорта, темпы роста промышленности, уровень цен на ресурсы, уровень инфляции [1].

В стране постоянно ведется разведка и поиск новых месторождений, в частности, в Сибири, на дне океанов и т.д. Около 30 заводов занимаются переработкой нефти. Добыча нефти постоянно растет, в нашем случае рассматривается рост с 491,3 млн т в 2007 г. до 555,7 млн т в 2018 году. Соответственно, количество переработки выросло с 229 млн т до 287 млн т.

Иногда рост добычи нефти уменьшается в связи с обязательствами со странами ОПЕК.

Из года в год усиливается модернизация нефтеперерабатывающей промышленности. Но на производство в нефтепереработке и нефтехимии, естественно, отрицательное влияние оказывают санкции.

Рассмотрим, как изменялись объемы добычи нефти в стране в период с 2007 по 2018 год, и как велась в тот же самый период переработка нефти, иные внутренние и внешние факторы при расчете не рассматриваем (табл. 1).

Таблица 1 — Объемы добычи нефти и переработки нефти, млн т*

Год	Добыча нефти, X	Переработка нефти, Y
2007	491,3	229,0
2008	488,0	235,7
2009	494,0	235,9
2010	504,9	248,8
2011	509,4	256,5
2012	516,8	265,4
2013	522,9	274,5
2014	525,0	288,9
2015	533,2	282,9
2016	548,6	280,6
2017	546,5	279,9
2018	555,7	287,0

* Составлено авторами по данным Росстата.

Из таблицы 1 видно, что перерабатывается примерно половина добываемой нефти. Можно считать, что это не лучшее использование добываемой нефти, остается много переработанной нефти.

В качестве методов анализа используются алгоритмы линейной и параболической регрессий, которые в дальнейшем и будем рассматривать.

Линейная регрессия

Согласно полю корреляции на рисунке 1 выдвигаем гипотезу, что связь между всеми данными X и Y имеет линейный характер. Напомним, что множество точек результативного и факторного признаков называется полем корреляции [2].

Из расчетов методом наименьших квадратов [2] получается линейное уравнение регрессии:

$$y = 0,862x - 184,331.$$

Рассчитаем показатель тесноты связи. Таким показателем является коэффициент линейной корреляции:

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{s(x)s(y)} = \frac{137495,173 - 519,692 \times 263,758}{22,127 \times 20,914} = 0.912.$$

Таким образом, в нашем случае связь между признаком Y и фактором X достаточно высокая и прямая.

Для проверки нулевой гипотезы о равенстве нулю генерального коэффициента корреляции

нормальной двумерной случайной величины при конкурирующей гипотезе $H_1 \neq 0$ (при уровне значимости α) вычисляется наблюдаемое значение критерия (величина случайной ошибки) [3]:

$$t_{набл} = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}.$$

По таблице критических значений распределения Стьюдента (уровень значимости α , число степеней свободы $k = n - 2$) находим критическую точку $t_{крит}$ двусторонней критической области. Если $|t_{набл}| > t_{крит}$ — нулевую гипотезу отвергают [3]. Здесь

$$t_{набл} = 0,912 \times \frac{\sqrt{10}}{\sqrt{1-0,912^2}} = 7,042.$$

По критерию Стьюдента ($\alpha = 0.05$, степень свободы $k = 10$) находим $t_{крит}$:

$$t_{крит} = \left(n - m - 1; \frac{\alpha}{2}\right) = (10; 0.025) = 2.23,$$

где $m = 1$ — количество объясняющих переменных.

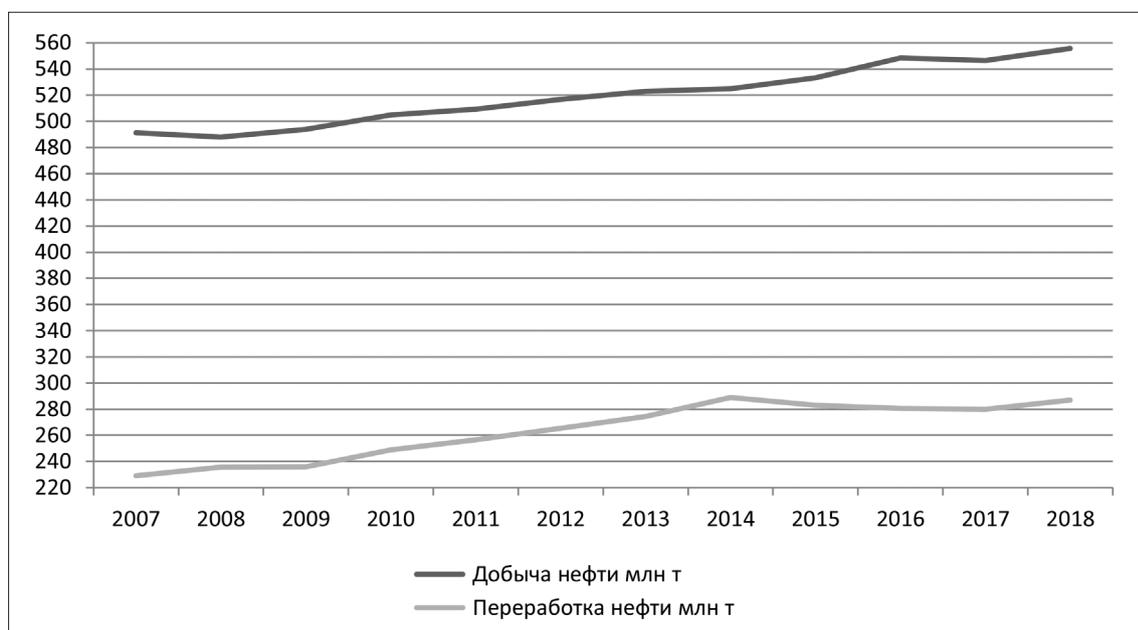


Рисунок 1 — Динамика исследуемых значений в период с 2007–2018 гг.

Figure 1 — Dynamics of the studied values in the period from 2007–2018

Так как $|t_{набл}| > t_{крит}$, то отклоняем гипотезу о равенстве нулю коэффициента корреляции, значит коэффициент корреляции статистически значим [2]. Тогда проверка гипотез о значимости коэффициентов линейной регрессии и корреляции в линейной регрессии $t_r^2 = t_b^2$ соответствует проверке гипотезы о существенности линейного уравнения регрессии.

Рассмотрим коэффициент эластичности E [4]:

$$E = b \frac{\bar{x}}{\bar{y}} = 0,862 \times \frac{519,692}{263,758} = 1,699.$$

В нашем примере коэффициент эластичности больше 1, значит X существенно влияет на Y .

Находим значение коэффициента детерминации:

$$R^2 = (0,912)^2 = 0,8322.$$

Он достаточно близок к единице.

Средняя относительная ошибка аппроксимации равна

$$\bar{A} = \frac{\sum |(y_i - y_x) : y_i|}{n} \times 100\% = \frac{0,311}{12} \times 100\% = 2,59\%.$$

Так как ошибка аппроксимации 2,59% меньше 7%, то данное уравнение хорошо описывает статистические данные, и ее можно использовать в качестве регрессии.

С помощью критерия Фишера можно оценить качество регрессионной модели в целом и по параметрам.

Вычислим значение критерия Фишера:

$$F = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m} = \frac{0,8322}{1-0,8322} \times \frac{12-1-1}{1} = 49,59,$$

где $m = 1$ для парной линейной регрессии.

Табличное значение критерия со степенями свободы $k_1 = 1$ и $k_2 = 10$, $F_{табл} = 4,96$. Поскольку фактическое значение $F > F_{табл}$, значит, у нас коэффициент детерминации статистически значим (найденная оценка уравнения регрессии статистически надежна).

Таким образом получили показатели, способные отразить зависимость добычи нефти и переработки нефти в период с 2007 по 2018 год. В таблице 2 приведены исследуемые показатели, а на рисунке 2 — график линейной регрессии.

Таблица 2 — Значения показателей линейной регрессии

Показатель	Значение
Коэффициент корреляции	0,912
Коэффициент детерминации	0,8322
Коэффициент эластичности	1,699
Средняя ошибка аппроксимации	2,59
Критерий Фишера	49,59

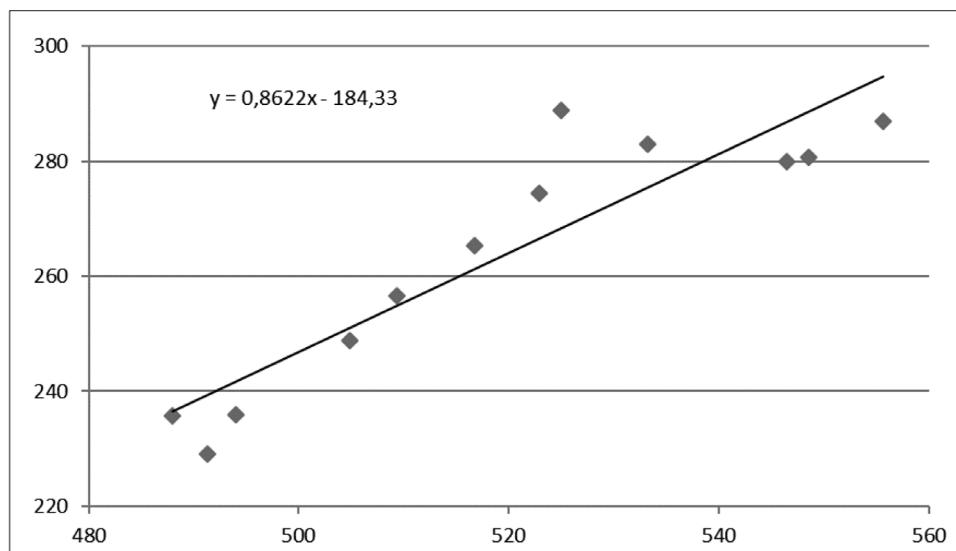


Рисунок 2 — Линейная регрессия

Figure 2 — Linear Regression

Параболическая регрессия

Рассмотрим прогнозирование с помощью модели параболической регрессии (по-другому — квадратичная регрессия), которая является простейшей криволинейной моделью. Параболическая регрессия описывается уравнением:

$$\hat{y} = ax^2 + bx + c.$$

В данном случае уравнение параболической регрессии будет иметь вид:

$$y = -0,0152x^2 + 16,754x - 4317,689.$$

Индекс корреляции равен:

$$\rho = \sqrt{1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}} = \sqrt{1 - \frac{378,867}{5248,889}} = 0,928.$$

Коэффициент детерминации, соответственно, равен:

$$R^2 = 0,928^2 = 0,861.$$

То есть 86,12% случаев изменения x приводят к изменению y , качество подбора уравнения регрессии — высокое.

Средняя ошибка аппроксимации в данном случае равна:

$$\bar{A} = \frac{\sum |y_i - y_x| : y_i}{n} \cdot 100\% = \frac{0,2093}{12} \cdot 100\% = 1,74\%.$$

Так как относительная ошибка меньше 7%, то данное уравнение очень хорошо аппроксимирует заданные значения. Данный показатель лучше, чем в линейной регрессии.

В нашем случае критический (табличный) критерий

$$F_{табл} = F(\alpha, k_1, k_2) = F(2; 9; 0,05) = 4,26.$$

Вычисленный критерий равен

$$F_{факт} = \frac{R^2}{1 - R^2} \times \frac{k_2}{k_1} = \frac{0,9278}{1 - 0,9278} \times \frac{9}{2} = 57,8438.$$

Так как фактическое значение $F > F_{табл}$, то согласно теории коэффициент детерминации и уравнение регрессии статистически значимы.

В таблице 3 приводим полученные показатели, а на рисунке 3 — график параболической регрессии.

Таблица 3 — Значения показателей параболической регрессии

Показатель	Значение
Индекс корреляции	0,963
Коэффициент детерминации	0,928
Коэффициент эластичности	1,735
Средняя ошибка аппроксимации	1,74
Критерий Фишера	57,8438

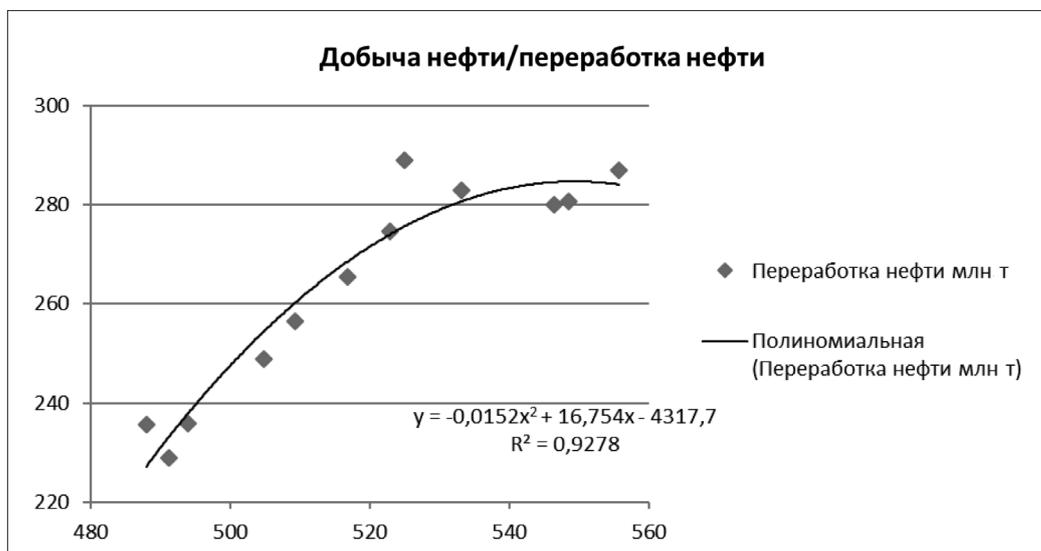


Рисунок 3 — Параболическая регрессия
Figure 3 — Parabolic Regression

Изучена зависимость объема переработки нефти Y от ее количества X . На этом этапе спецификации был выбран параболический тренд. Оценены его параметры методом наименьших квадратов. Статистическая значимость уравнения проверена с помощью коэффициента детерминации и критерия Фише-

ра. Установлено, что в исследуемой ситуации 92,78% общей вариабельности Y объясняется изменением временного параметра. Установлено также, что параметры модели статистически значимы. Полученные оценки уравнения регрессии позволяют использовать его для прогноза.

Заключение

В данной работе исследовалась статистическая зависимость переработки нефти (млн т) от добычи нефти и найдена соответствующая модель. В рассмотренных случаях (линейная и параболическая регрессии) связь между признаком Y и фактором X весьма высокая и прямая, коэффициент корреляции статистически значим, изменение X существенно влияет на Y , точность подбора уравнения регрессии высокая.

Графики регрессий с наибольшей степенью точности отражают зависимость добычи нефти от переработки нефти, значит, данную зависимость необходимо исследовать именно с помощью линейной и параболической регрессии. Коэффициент (индекс) корреляции позволил определить тесноту линейной связи между исследуемыми экономическими показателями

и в том числе правильно определить тип связи — прямая или обратная, и далее — принять решение, связанное с выбором анализа тех или иных показателей.

Параболическая регрессия более точно отражает зависимость, так как переработка нефти ограничивается не только количеством добычи, но и наличием необходимых производственных мощностей и других экономических условий для ее переработки. Отсюда естественный вывод — при наращивании добычи нефти следует заранее планировать развитие производства для ее переработки, а это является выгодным не только экономически, но и политически, исходя из современных реалий.

Аналогично, методами эконометрики и математической статистики нами исследовались и другие экономические процессы [5].

Список источников

1. Корчагин Ю. А. Современная экономика России. Ростов н/Д. : Феникс, 2007. 543 с. : ил. (Серия «Высшее образование»).
2. Гильмутдинов Р. З., Нурисламова Л. Ф. Эконометрика : учебное пособие. Уфа : Уфимский государственный университет экономики и сервиса. 2016. 98 с.
3. Тихомиров Н. П., Тихомирова Т. М., Ушмаев О. С. Методы эконометрики и многомерного статистического анализа : учебник для студентов, обучающихся по экономическим специальностям и направлениям. Москва : Экономика, 2010. 636 с. : ил.
4. Гильмутдинов Р. З. Математические методы в экономике : курс лекций. Уфа : УИКиП, 2006. 53 с.
5. Гильмутдинов Р. З., Тухватуллина Л. Р. Исследование моделей налоговых поступлений в зависимости от господдержки // Вестник УГНТУ. 2022. № 4 (42). С. 50–56.

References

1. Korchagin Yu. A. Modern Economy of Russia. Rostov on Don: Phoenix; 2007. 543 p.: ill. (Series "Higher Education"). (In Russ.).
2. Gilmutdinov R. Z., Nurislamova L. F. Econometrics: a tutorial. Ufa: Ufa State University of Economics and Service; 2016. 98 p. (In Russ.).
3. Tikhomirov N. P., Tikhomirova T. M., Ushmaev O. S. Methods of econometrics and multivariate statistical analysis: a textbook for students studying in economic specialties and areas. Moscow: Economy, 2010. 636 p.: ill. (In Russ.).
4. Gilmutdinov R. Z. Mathematical methods in economics: a course of lectures. Ufa: UIKIP, 2006. 53 p. (In Russ.).
5. Gilmutdinov R. Z., Tuxvatullina L. R. Study of models of tax revenues depending on state support. *Vestnik UGNTU = Bulletin of Ufa State Petroleum Technical University*. 2022;(4(42)):50–56. (In Russ.).

Информация об авторах

Р. З. Гильмутдинов — кандидат физико-математических наук, доцент кафедры экономики и информационных технологий;

Л. Ю. Мельник — кандидат физико-математических наук, доцент кафедры цифровых технологий и моделирования;

Л. Р. Тухватуллина — кандидат физико-математических наук, доцент кафедры цифровых технологий и моделирования.

Information about the authors

R. Z. Gilmutdinov — Candidate of Science (Physical and Mathematical), Associate Professor of the Department of Economics and Information Technology;

L. Yu. Melnik — Candidate of Science (Physical and Mathematical), Associate Professor of the Department of Digital Technologies and Modeling;

L. R. Tukhvatullina — Candidate of Science (Physical and Mathematical), Associate Professor of the Department of Digital Technologies and Modeling.

Статья поступила в редакцию 06.12.2024; одобрена после рецензирования 16.12.2024; принята к публикации 23.12.2024.

The article was submitted 06.12.2024; approved after reviewing 16.12.2024; accepted for publication 23.12.2024.