

УДК 004.045, 311.218

*АВТОМАТИЗАЦИЯ СТАТИСТИЧЕСКОГО АНАЛИЗА ФАКТОРОВ  
ЭФФЕКТИВНОСТИ ГОСУДАРСТВЕННОЙ ПОДДЕРЖКИ СЕЛЬСКОГО  
ХОЗЯЙСТВА С ПРИМЕНЕНИЕМ ЯЗЫКА ПРОГРАММИРОВАНИЯ PYTHON*

**В.В. ДЕМИЧЕВ, А.А. НЕСТРАТОВА**

*Российский государственный аграрный университет – МСХА им. К.А. Тимирязева,  
127550, Российская Федерация, г. Москва, ул. Лиственничная аллея, 4,  
электронная почта: vadi.demiche@mail.ru, nastyashka-jka@mail.ru*

Целью настоящей статьи является – автоматизация ключевых фрагментов корреляционно-регрессионного анализа эффективности государственной поддержки сельского хозяйства Российской Федерации, а также выявление ключевых факторов эффективности субсидирования сельского хозяйства России. В качестве инструмента реализации выбран высокоуровневый язык программирования Python. В результате построения матрицы парных коэффициентов корреляции и регрессионных моделей, выявлены ключевые зависимости между факторными и результативными показателями эффективности государственной поддержки сельского хозяйства.

**Ключевые слова:** автоматизация, Python, субсидии, эффективность, сельское хозяйство, корреляционно-регрессионный анализ.

Согласно ведомственному проекту «Цифровое сельское хозяйство» в России к 2024 году должна быть создана национальная платформа цифрового государственного управления сельским хозяйством на региональном и муниципальном уровнях [5]. Среди первоочередных задач проекта выделяются: 100% наличие данных об объектах сельскохозяйственных ресурсов в цифровой платформе, а также данных об объемах и эффективности государственной поддержки сельского хозяйства регионов. Разработанная цифровая платформа позволит осуществлять учет, мониторинг и аналитику развития сельского хозяйства регионов, а также принимать управленческие решения о предоставлении государственной поддержки сельскохозяйственным товаропроизводителям. В связи с этим особую актуальность приобретает вопрос автоматизации статистического анализа эффективности государственной поддержки сельского хозяйства, включающий разработку модулей корреляционно-регрессионного анализа, <http://ntk.kubstu.ru/file/3068>

способствующего в дальнейшем принятию взвешенных управленческих решений.

В 2006 году государство принимает федеральный закон о развитии сельского хозяйства и приступает к реализации приоритетного национального проекта «Развитие АПК», а позже Государственных программ развития сельского хозяйства и регулирования рынков сельскохозяйственной продукции, сырья и продовольствия. За период 2006-2018 гг. в сельское хозяйство было направлено порядка 1,6 трлн. руб. федеральных субсидий, что составляет около 2% ВВП 2019 года.

В настоящее время, государственная программа продлена до 2025 года, однако, уже сейчас представляет интерес оценка и выявление ключевых факторов эффективности государственной поддержки сельского хозяйства регионов России [1,6].

Для проведения корреляционно-регрессионного анализа были отобраны следующие 8 факторных и 3 результативных показателя (табл. 1).

Таблица 1 – Показатели для кластерного анализа

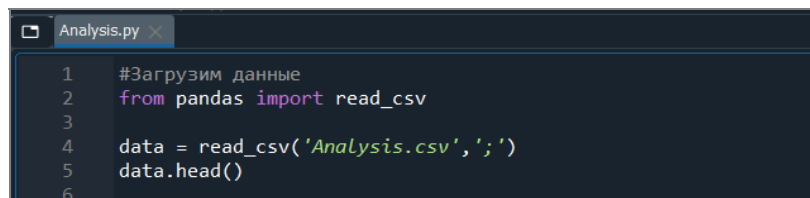
Переменная	Описание
X1	Внесение минеральных удобрений на 1 га посева с.-х. культур в 2018 г.
X2	Расход кормов в расчете на одну условную голову крупного скота в сельскохозяйственных организациях в 2018 году, ц. к. ед.
X3	Энергетические мощности в сельскохозяйственных организациях, в расчете на 1 работника, в 2018 году, л.с.
X4	Отношение субсидий в регионе к среднему по совокупности регионов в среднем за период 2006-2018 гг.
X5	Объем капитальных вложений федерального бюджета на 1 рубль капитальных вложений субъектов федерации, руб.
X6	Удельный вес животноводства в производстве продукции сельского хозяйства в 2018 году, %
X7	Средний балл продуктивности климата
X8	Средний объем субсидий на 1 руб. стоимости продукции сельского хозяйства, руб.
Y1	Рентабельность (убыточность) субсидий в 2018 г., %
Y2	Рентабельность производства (без субсидий) в 2018 г., %
Y3	Средняя урожайность зерновых за период 2006-2018 гг., ц/га

Источник: составлено автором.

Здесь X1-X8 – факторные (объясняющие) переменные, Y1, Y2, Y3 – результативные (объясненные) переменные.

Совокупность субъектов РФ, участвующих в анализе, численно равна 78 регионам.

1 этап. На первом этапе необходимо загрузить данные (рис. 1), используя библиотеку Pandas [2].

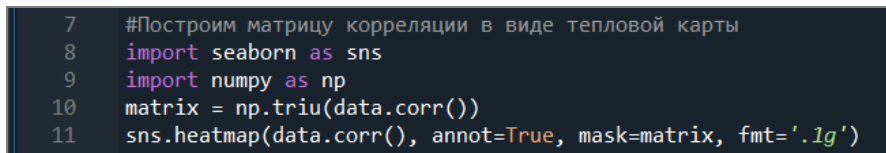


```
Analysis.py x
1 #Загрузим данные
2 from pandas import read_csv
3
4 data = read_csv('Analysis.csv', ';')
5 data.head()
6
```

Рисунок 1 – Загрузка данных в среду разработки Spyder (Python 3.7)

Источник: составлено автором.

2 этап. Следующим этапом необходимо построить матрицу парных коэффициентов корреляции, чтобы проверить взаимосвязь между всеми представленными переменными [3]. Для этого импортируем библиотеки Seaborn и Numpy. Для наглядности, используя функцию `sns.heatmap()` и построим матрицу парных коэффициентов корреляции в виде тепловой карты [4]. Реализация будет выглядеть следующим образом (рис. 2):



```
7 #Построим матрицу корреляции в виде тепловой карты
8 import seaborn as sns
9 import numpy as np
10 matrix = np.triu(data.corr())
11 sns.heatmap(data.corr(), annot=True, mask=matrix, fmt='.1g')
```

Рисунок 2 – Реализация функции тепловой карты

Источник: составлено автором.

В итоге получим следующую матрицу парных коэффициентов корреляции:

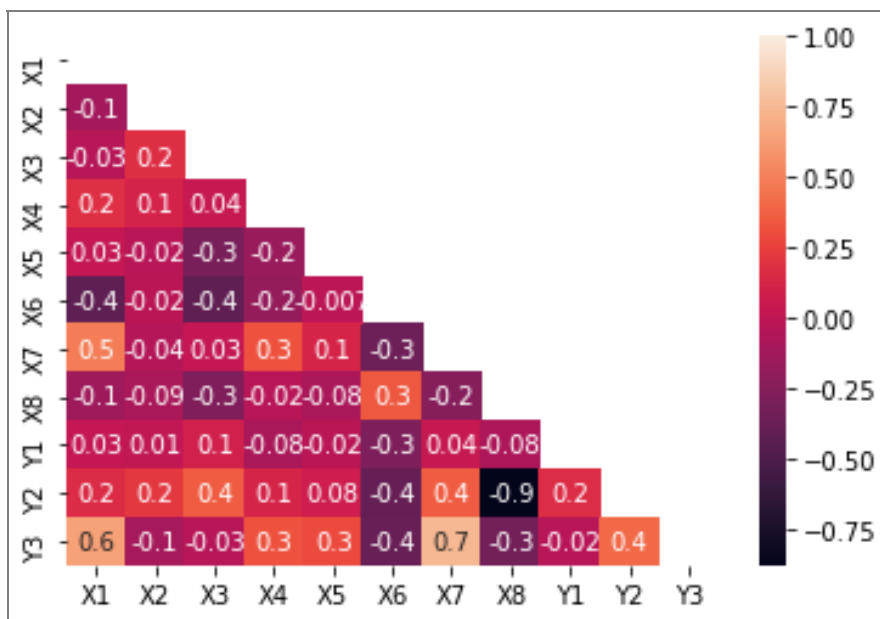


Рисунок 3 – Матрица парных коэффициентов корреляции

Источник: составлено автором.

Анализируя матрицу (рис. 3), отметим, что коррелируют между собой следующие пары переменных (коэффициент больше или равен значения 0,3):

Y1 – X6;

Y2 – X3, X6, X7, X8

Y3 – X1, X4, X5, X6, X7, X8

Также анализ матрицы парных коэффициентов регрессии показал, что на результативную переменную Y1 практически все переменные оказывают влияние ниже среднего, только переменная X6 оказывает средний уровень влияния. Поэтому далее построим модель регрессии между этими переменными.

3 этап. На этом этапе определим переменные и построим диаграмму рассеивания (рис. 4 и 5).

```

6
7 #определим переменные x и y
8 x = data['X6']
9 y = data['Y1']
10
11 #импортируем библиотеку для построения диаграммы рассеивания
12 import matplotlib.pyplot as plt
13
14 #построим диаграмму рассеивания
15 plt.plot(x, y, 'o', color = 'green')
16 plt.show()
    
```

Рисунок 4 – Реализация построения диаграммы рассеивания

Источник: составлено автором.

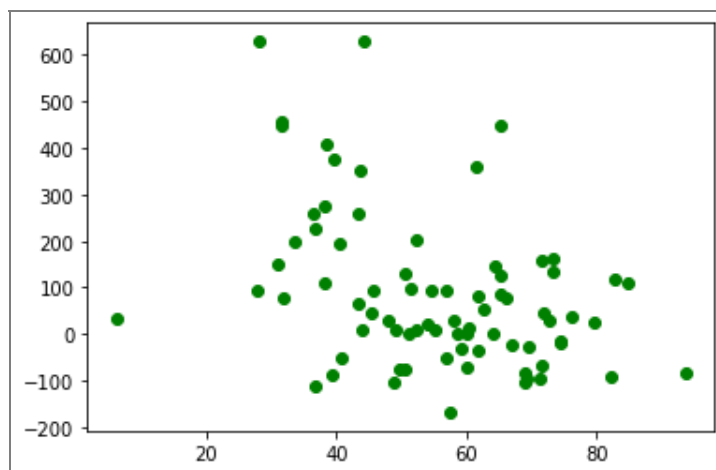


Рисунок 5 – Диаграмма рассеивания (зависимости) рентабельности субсидий (%) от удельного веса животноводства в регионе (%)

Источник: составлено автором.

Теснота связи находится на среднем уровне и подтверждается обратная зависимость между переменными.

4 этап. Далее выведем регрессионную модель  $Y1 = a + bX6$  и другие статистические параметры корреляционно-регрессионного анализа (рис. 6) [7].

```

13 #Построение модели регрессии Y1 и X6
14 import statsmodels.api as sm
15
16 #определим переменные x и y
17 x = data['X6']
18 y = data['Y1']
19 #Добавим в модель условное начало
20 x = sm.add_constant(x)
21
22 # Построим модель на основе метода наименьших квадратов
23 model = sm.OLS(y, x).fit()
24 predictions = model.predict(x)
25
26 #Сформируем вывод
27 model.summary()
28 print(model.summary())
    
```

Рисунок 6 – Реализация модели парной линейной регрессии методом наименьших квадратов

Источник: составлено автором.

Вывод итогов, содержаний модель регрессии, коэффициент детерминации, F- и t-тесты и другие статистические характеристики выглядит следующим образом (рис. 7).

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Y1      R-squared:                0.153
Model:                  OLS      Adj. R-squared:           0.141
Method:                 Least Squares      F-statistic:              13.50
Date:                   Fri, 05 Jun 2020    Prob (F-statistic):       0.000445
Time:                   19:57:10          Log-Likelihood:           -497.23
No. Observations:      77      AIC:                      998.5
Df Residuals:          75      BIC:                      1003.
Df Model:               1
Covariance Type:       nonrobust
=====
                        coef      std err      t      P>|t|      [0.025      0.975]
-----
const                 311.6666    63.465     4.911    0.000    185.238    438.095
X6                    -4.0608     1.105    -3.674    0.000    -6.262    -1.859
=====
Omnibus:              13.839    Durbin-Watson:           2.030
    
```

Рисунок 7 – Результаты регрессионного анализа для переменных Y1 и X6

Источник: составлено автором.

Построенная модель регрессии подтвердила обратную зависимость между рентабельностью субсидий и удельным весом животноводства. При увеличении удельного веса животноводства в региональном производстве продукции сельского хозяйства на 1%, рентабельность субсидий сократиться на 4.1%.

5 этап. Построение регрессии для переменной Y2. Стоит отметить, что при построении модели регрессии, значимость переменной X6 составила 66,4%, что значительно превышает допустимый уровень 5%. Поэтому указанная переменная не включена в дальнейшие этапы построения регрессии (рис. 8).

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Y2      R-squared:                0.810
Model:                  OLS      Adj. R-squared:           0.803
Method:                 Least Squares      F-statistic:              104.0
Date:                   Fri, 05 Jun 2020    Prob (F-statistic):       2.72e-26
Time:                   20:29:18          Log-Likelihood:           -302.75
No. Observations:      77      AIC:                      613.5
Df Residuals:          73      BIC:                      622.9
Df Model:               3
Covariance Type:       nonrobust
=====
                        coef      std err      t      P>|t|      [0.025      0.975]
-----
const                 -17.3159    7.793     -2.222    0.029    -32.848    -1.784
X3                     0.1523     0.061     2.500    0.015     0.031     0.274
X7                     0.1411     0.045     3.127    0.003     0.051     0.231
X8                    -266.3003   18.129   -14.689    0.000   -302.431   -230.170
=====
    
```

Рисунок 8 – Результаты регрессионного анализа для переменных Y2 – X3, X7, X8

Источник: составлено автором.

В прямой зависимости с рентабельностью производства находятся такие факторные показатели как энерговооруженность и средний балл продуктивности климата. Первый фактор отражает интенсификацию производства. То есть, наибольший прирост эффективности получают те регионы, которые имеют возможность вкладываться в обновление техники, производительность труда, энерговооруженность.

Показатель X8 (средний объем субсидий на 1 руб. стоимости продукции сельского хозяйства) находится в обратной зависимости к рентабельности производства. Во многом это связано с тем, что субсидии направлялись в регионы для развития животноводства, рентабельность которого, как показала рассмотренная выше модель – значительно ниже.

6 этап. Построение модели регрессии для переменной Y3 (рис. 9).

OLS Regression Results						
Dep. Variable:	Y3	R-squared:	0.744			
Model:	OLS	Adj. R-squared:	0.726			
Method:	Least Squares	F-statistic:	41.22			
Date:	Fri, 05 Jun 2020	Prob (F-statistic):	1.11e-19			
Time:	20:37:02	Log-Likelihood:	-231.36			
No. Observations:	77	AIC:	474.7			
Df Residuals:	71	BIC:	488.8			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-0.7171	2.379	-0.301	0.764	-5.461	4.026
X1	0.0739	0.014	5.335	0.000	0.046	0.102
X4	1.1623	0.568	2.048	0.044	0.030	2.294
X5	0.0276	0.007	3.742	0.000	0.013	0.042
X7	0.1354	0.022	6.223	0.000	0.092	0.179
X8	-15.8981	7.024	-2.263	0.027	-29.904	-1.892

Рисунок 9 – Результаты регрессионного анализа для переменных Y3 – X1, X4, X5, X7, X8

Источник: составлено автором.

На показатель производственной эффективности – средняя урожайность зерновых, прямое воздействие оказывают такие факторы как внесение минеральных удобрений, объем предоставляемых субсидий по отношению к среднему по совокупности и объем инвестиций из федерального бюджета в сельское хозяйство региона. Отрицательное влияние оказывает показатель субсидии в расчете на 1 руб. стоимости

продукции сельского хозяйства, что также связано с тем, что относительно стоимости продукции большой объем субсидий предоставлялся в животноводческие регионы, где урожайность зерновых значительно ниже.

Таким образом, положительное влияние на изучаемые показатели эффективности оказали энерговооруженность и климатический потенциал в регионе, внесение минеральных удобрений (энерговооруженность и внесение минеральных удобрений, фактически, отражают интенсификацию производства), объем предоставляемых субсидий и соотношение инвестиций из федерального бюджета к инвестициям региональных бюджетов. Отрицательное воздействие оказали – удельный вес животноводства в объеме производства сельскохозяйственной продукции и объем субсидий на 1 руб. стоимости продукции сельского хозяйства.

#### ЛИТЕРАТУРА

1. Демичев, В.В. Рейтинг инклюзивного развития экономики сельского хозяйства регионов России / В.В. Демичев // Российский экономический интернет журнал. – 2018. – № 3. – с. 29.
2. Любанович, Б. Простой Python. Современный стиль программирования. / Б. Любанович - СПб.: Питер, 2019. - 480 с.: ил.
3. Пример решения задачи множественной регрессии с помощью Python. [Электронный ресурс]. - Режим доступа: <https://habr.com/ru/post/206306/> (Дата обращения: 26.05.2020)
4. Тринадцать способов настроить визуализацию матрицы корреляции. [Электронный ресурс]. - Режим доступа: <https://datastart.ru/blog/read/seaborn-heatmaps-13-sposobov-nastroit-vizualizaciyu-matricy-korrelyacii> (Дата обращения: 30.04.2020)
5. Цифровая трансформация сельского хозяйства России: офиц. изд. – М.: ФГБНУ «Росинформагротех», 2019 – 80 с.
6. Demichev, V.V. Sustainable Development of Agriculture in Russian Regions on the Basis of Inclusiveness / V.V. Demichev // Proceedings of the <http://ntk.kubstu.ru/file/3068>



international scientific conference Hradec Economic Days 2020 (Hradec Králové, Czech Republic, April 2–3, 2020). - Vol. 10(1). – pp. 85-94.

7. Simple and Multiple Linear Regression in Python. Available online: <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9> (accessed on 16 May 2020).

#### REFERENCES

1. Demichev, V.V. Rejting inklyuzivnogo razvitiya ekonomiki selskogo khozyaystva regionov Rossii / V.V. Demichev // Rossiyskiy ekonomicheskii internet zhurnal. – 2018. – № 3. – s. 29.

2. Lyubanovich, B. Prostoy Python. Sovremennyy stil programmirovaniya. / B. Lyubanovich - SPb.: Piter, 2019. - 480 s.: il.

3. Primer resheniya zadachi mnozhestvennoy regressii s pomoshchyu Python. [Elektronnyy resurs]. - Rezhim dostupa: <https://habr.com/ru/post/206306/> (Data obrashcheniya: 26.05.2020)

4. Trinadtsat sposobov nastroit vizualizatsiyu matritsy korrelyatsii. [Elektronnyy resurs]. - Rezhim dostupa: <https://datastart.ru/blog/read/seaborn-heatmaps-13-sposobov-nastroit-vizualizatsiyu-matricy-korrelyatsii> (Data obrashcheniya: 30.04.2020)

5. Tsifrovaya transformatsiya selskogo khozyaystva Rossii: ofits. izd. – M.: FGBNU «Rosinformagrotekh», 2019 – 80 s.

6. Demichev, V.V. Sustainable Development of Agriculture in Russian Regions on the Basis of Inclusiveness / V.V. Demichev // Proceedings of the international scientific conference Hradec Economic Days 2020 (Hradec Králové, Czech Republic, April 2–3, 2020). - Vol. 10(1). – pp. 85-94.

7. Simple and Multiple Linear Regression in Python. Available online: <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9> (accessed on 16 May 2020).

*AUTOMATION OF STATISTICAL ANALYSIS OF EFFICIENCY FACTORS  
OF STATE SUPPORT FOR AGRICULTURE IN PYTHON*

**V.V. DEMICHEV, A.A. NESTRATOVA**

*Russian State Agrarian University – Moscow Timiryazev Agricultural Academy,  
4, Listvennichnaya alleya, Moscow, Russian Federation, 127550,  
e-mail: vadi.demiche@mail.ru, nastyashka-jka@mail.ru*

The purpose of this article is to automate key fragments of the correlation-regression analysis of the effectiveness of state support for agriculture in the Russian Federation, as well as to identify key factors in the effectiveness of subsidizing agriculture in Russia. The high-level programming language Python was chosen as the implementation tool. As a result of constructing a matrix of correlation coefficients and regression models, the key dependencies between the factorial and effective indicators of the effectiveness of state support for agriculture were identified.

**Key words:** automation, Python, subsidies, efficiency, agriculture, correlation and regression analysis