



Статистический анализ данных отчетности о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса

Уколова А.В., к.э.н., доцент, и.о. заведующего кафедрой статистики и кибернетики,

ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, Москва, Россия

Быков Д.В., ассистент,

ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, Москва, Россия

Витковская А.М., магистрант,

ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, Москва, Россия

Аннотация. В статье представлены результаты анализа данных отчетности о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса. Анализ данных осуществлялся посредством разработанного программного средства, позволяющего автоматизировать проведение кластерного анализа, включая формирование сводной информации по выделенным кластерам для проведения комплексной оценки выделенных групп объектов, а также корреляционно-регрессионного анализа для выявления взаимосвязи между показателями.

Ключевые слова: ведомственная отчетность, автоматизация анализа, R, кластерный анализ, корреляционно-регрессионный анализ

Statistical analysis of data on the financial and economic condition of agricultural producers

Ukolova A.V., Ph.D., associate professor, acting Head of the Department of Statistics and Cybernetics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia

Bykov D.V., assistant, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia

Vitkovskaya A.M., magister, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia

Annotation. The article presents the results of the analysis of data on the financial and economic condition of commodity producers of the agro-industrial complex. Data analysis is carried out using the developed software tool that allows cluster analysis, including the collection of information on selected clusters for a comprehensive assessment of selected groups of objects, as well as correlation and regression analysis to identify the relationship between indicators.

Key words: departmental reporting, automation, R, cluster analysis, correlation-regression analysis.

Изучение форм ведомственной отчетности необходимо для более полного представления о текущем состоянии сельского хозяйства и перспектив его развития.

В исследовании использованы данные отчета о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса за 2019 г., утвержденного приказом Минсельхоза России от 3 декабря 2019 г. № 669¹.

Целью исследования является анализ данных отчетности о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса.

Для автоматизации анализа было разработано программное средство с помощью языка R, среды разработки RStudio и специального пакета Shiny². Интерфейс приложения представлен ниже (Рис. 1).

¹ Об утверждении форм отчета о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса за 2019 год и срока его представления: Приказ Минсельхоза России от 03.12.2019 № 669. – Доступ из справ.-правовой системы «КонсультантПлюс». – Текст: электронный

² Бессонов, А. С. Интерфейсы автоматизированных систем. Методические указания: методические указания / А.С. Бессонов. – Москва: РТУ МИРЭА, 2020. – 61 с. – Текст: непосредственный

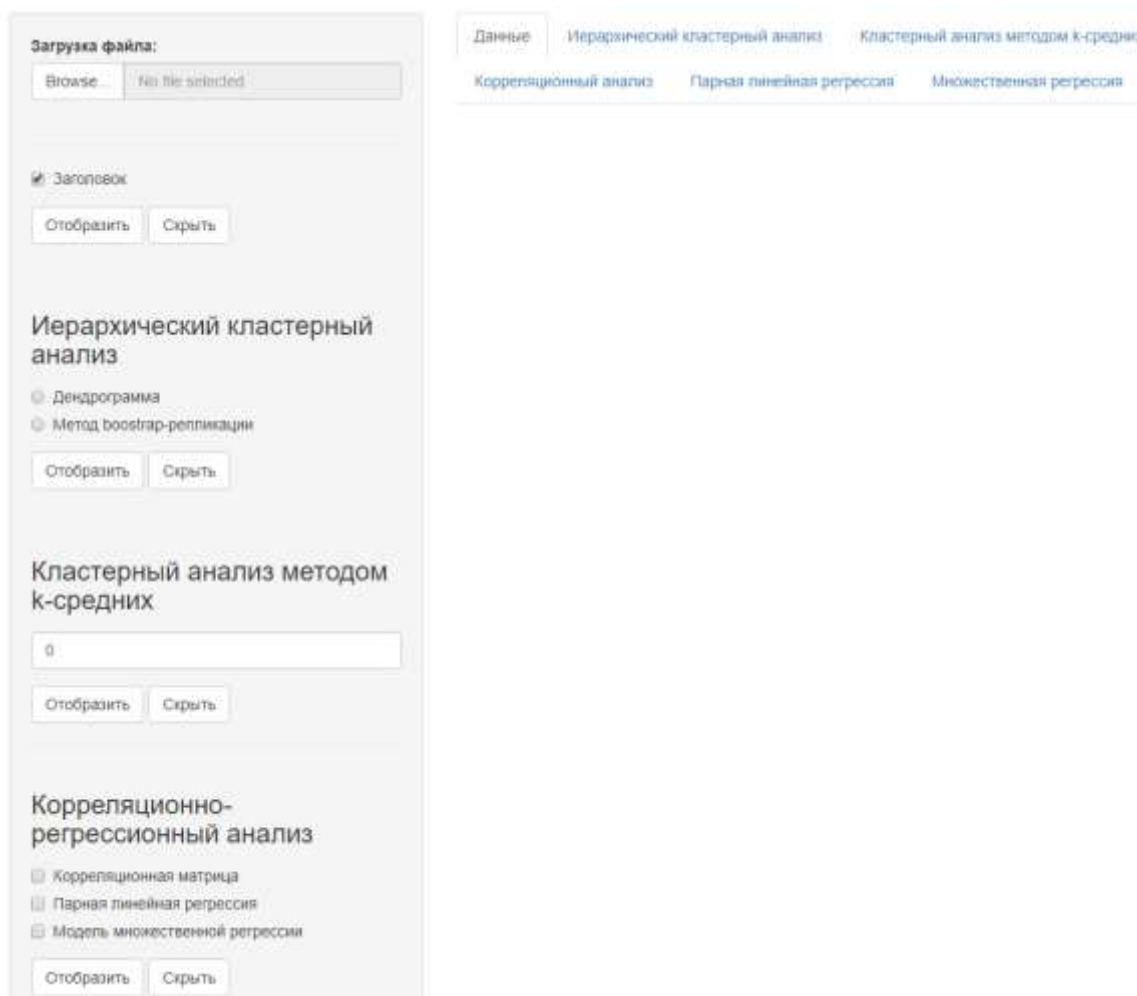


Рис. 1 – Интерфейс информационной системы

В качестве объекта исследования была выбрана статистическая совокупность 135 сельскохозяйственных организаций Белгородской области. Для наиболее существенных признаков, используемых в работе, введены обозначения:

- x_1 – Площадь сельскохозяйственных угодий (с.-х.), га;
- x_2 – Выручка от реализации продукции в расчете на 100 га с.-х. угодий, руб.;
- x_3 – Среднегодовая численность работников в расчете на 100 га с.-х. угодий, чел.;
- x_4 – Среднегодовая стоимость основных средств производства на 1 га с.-х. угодий, руб.;
- x_5 – Обеспеченность энергоресурсами на 100 га с.-х. угодий, л.с.;

- x_6 – Затраты на оплату труда в расчете на 100 га с.-х. угодий, руб.;
- x_7 – Материальные затраты на минеральные удобрения в расчете на 100 га с.-х. угодий, руб.;
- x_8 – Рентабельность реализации продукции, %;
- x_9 – Выручка от реализации продукции растениеводства в расчете на 100 га с.-х. угодий, руб.;
- x_{10} – Производственные затраты на 100 га с.-х. угодий, руб.

Для реализации кластерного анализа были выбраны признаки x_2 , x_3 , x_4 и x_5 . На первом этапе был реализован иерархический кластерный анализ, в результате которого строится дендрограмма (Рис. 2). Она показывает все возможные кластеры, поэтому по ней можно определить, на сколько групп делить совокупность при проведении кластерного анализа методом к-средних.

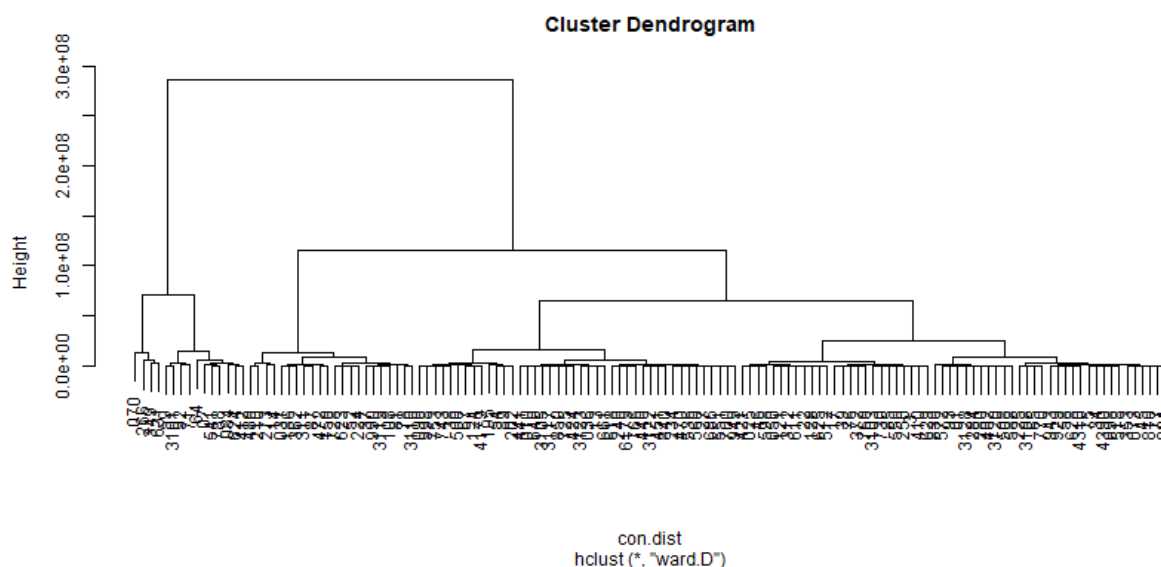


Рис. 2 – Дендрограмма иерархического кластерного анализа

Далее был проведен кластерный анализ методом к-средних. Для этого в боковой панели, в области «Кластерный анализ методом к-средних» необходимо вписать количество кластеров (в настоящем исследовании организации делились на 3 кластера) (Рис. 3).

На рис. 3 показано, сколько предприятий вошло в каждый кластер: 83, 37, 15. Если атрибут «between_SS / total_SS» близок к 100%, то объекты

кластеризуются достаточно точно³. В данном случае внутрикластерная сумма квадратов составляет 77,2% от общего объема вариации, следовательно, качество кластеризации довольно высокое.

Результаты кластеризации

```
K-means clustering with 3 clusters of sizes 83, 37, 15

Cluster means:
      x2      x3      x4      x5
1 3011862 1.841566 63714.79 434.5843
2 6983344 3.112703 106362.49 1039.7919
3 20146937 15.862000 1017940.89 4644.5867

Clustering vector:
 [1] 1 2 1 2 1 3 1 2 1 1 1 2 3 1 1 1 1 3 1 2 2 2 2 2 1 1 1 1 1 3 3 1 2 1 1
 [37] 1 3 1 1 1 2 2 1 1 1 3 1 1 1 1 1 1 1 1 3 1 1 2 1 1 1 1 1 2 1 2 3 2 2 2
 [73] 1 1 2 1 3 1 2 2 1 1 2 1 1 1 1 1 3 1 1 2 1 2 2 1 3 1 1 1 2 1 1 1 1 1 1 1
 [109] 3 1 1 2 1 2 2 1 2 2 1 1 2 1 1 2 1 1 1 1 1 1 2 3 1 3 2 1

Within cluster sum of squares by cluster:
 [1] 1.017435e+14 1.123095e+14 9.027994e+14
 (between_SS / total_SS = 77.2 %)

Available components:

 [1] "cluster"      "centers"      "totss"       "withinss"
 [5] "tot.withinss" "betweenss"    "size"        "iter"
 [9] "ifault"
```

Рис. 3 – Результаты кластеризации методом k-средних

Средние значения каждого показателя по кластерам отличаются (матрица «Cluster means» на Рис. 2). В первый кластер попали организации с наименьшей выручкой, численностью работников, стоимостью основных средств производства и обеспеченностью энергоресурсами в расчете на единицу площади, во второй кластер – средними значениями по перечисленным ранее показателями, а в третий кластер – наибольшими значениями.

Кроме того, был проведен дисперсионный анализ⁴ (Рис. 4).

³ Analysis of Variance (ANOVA) in R. [электронный ресурс] – Режим доступа: <https://www.instructables.com/Analysis-of-Variance-ANOVA-in-R/> (дата обращения: 01.11.2022). – Текст: электронный

⁴ Кластерный анализ в R. [электронный ресурс] – Режим доступа: <https://mymagictools.blogspot.com/2015/07/r.html> (дата обращения: 01.11.2022). – Текст: электронный

Дисперсионный анализ

 	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x2	1	3.487e+09	3.487e+09	38.89	5.812e-09
x3	1	3.566e+09	3.566e+09	39.78	4.097e-09
x4	1	1.325e+09	1.325e+09	14.78	0.0001879
x5	1	2.64e+09	2.64e+09	29.45	2.709e-07
Residuals	130	1.165e+10	89648670	NA	NA

Рис. 4 – Дисперсионный анализ

Как видно из рис. 4, различия по всем показателям являются существенными на уровне значимости 1%. Наибольшим образом повлиял на различия между кластерами признак x_3 – среднегодовая численность работников в расчете на 100 га с.-х. угодий, чел.

Также в разработанном приложении автоматизирован процесс сводки данных по выделенным кластерам, что является отличительной особенностью данного программного средства в сравнении с аналогами. На основе сводных данных можно провести комплексную оценку выделенных кластеров.

Сначала рассмотрим показатели интенсификации и эффективности сельскохозяйственного производства (табл. 1).

Распределение хозяйств по кластерам имеет правостороннюю скошенность, показатели в целом планомерно растут от низшей – к высшей группе.

В расчете на 100 га с.-х. угодий выручка от реализации продукции выше в 4 раза в III группе по сравнению с I, но производственные затраты и затраты на реализованную продукцию – в 6 раз, поэтому рентабельность в III группе ниже на 44 п.п. Высокий уровень интенсификации производства (численность

работников в 5 раз выше, чем в первой группе, обеспеченность энергоресурсами – в 6 раз) при высоком уровне производства и меньшем уровне рентабельности, что может говорить о различиях в специализации производства. Организации 2 группы ведут также рентабельное производство и занимают промежуточное положение.

Таблица 1

Показатели интенсификации и эффективности сельскохозяйственного производства, в расчете на 100 га сельскохозяйственных угодий

Показатель	Типические группы			В среднем по совокупности	Отношение III группы к I, %
	I (низшая)	II (средняя)	III (высшая)		
Число организаций	83	37	15	135	x
Выручка от реализации продукции, тыс. руб.	3773	7067	16208	4739	429,6
Затраты на реализованную продукцию, тыс. руб.	2433	4793	14610	3172	600,5
Прибыль, тыс. руб.	1340	2277	1590	1567	119,3
Рентабельность реализации продукции, %	55	48	11	49	- 44 п.п.
Производственные затраты, тыс. руб.	2981	6315	18824	4005	631,5
в т.ч. на минеральные удобрения	385,1	425,1	579,2	397,5	150,4
Численность работников, чел.	1,0	2,3	5,3	1,4	507,9
Обеспеченность энергоресурсами, л.с.	229	929	3994	451	1743
Среднегодовая стоимость основных средств производства, тыс. руб.	5119	12720	58690	7707	1146

Далее в приложении выводится таблица с показателями размера и специализации сельскохозяйственного производства (табл. 2).

Предприятия в I и II группе специализируются на растениеводстве, а III группа – на животноводстве. Во II группе наблюдается наибольшая выручка в расчете на 1 предприятие, равная 601,4 млн. руб. Энергоресурсы и производственные затраты больше во II группе, чем в двух других. Предприятия II группы по всем показателям размеров больше I и III групп, в том числе по площади с.-х. угодий, численности работников, выручке и производственным затратам.

**Показатели размера и специализации сельскохозяйственного производства,
расчете на 1 предприятие**

Показатель	Типические группы			В среднем по совокупности	Отношение III группы к I, %
	I (низшая)	II (средняя)	III (высшая)		
Число организаций	83	37	15	135	
Площадь сельскохозяйственных угодий, га	11878	8507	1273	9776	10,7
Выручка от реализации продукции - всего, млн. руб.	448,1	601,4	206,3	463,3	46,0
в т.ч. от реализации продукции:					
растениеводства	382,2	326,2	61,8	331,3	16,2
животноводства	65,8	275,2	144,6	132,0	219,5
Удельный вес в общей выручке от реализации продукции, %:					
растениеводства	85,3	54,2	29,9	71,5	35,1
животноводства	14,7	45,8	70,1	28,5	476,8
Производственные затраты, млн. руб.	354,0	537,2	239,6	391,5	67,7
в т.ч. на минеральные удобрения	45,7	36,2	7,4	38,9	16,1
Среднегодовая численность работников, чел.	123,7	195,9	67,3	137,2	54,4
Энергетические мощности, л.с.	27219	79035	50839	44045	186,8
Среднегодовая стоимость основных средств производства, млн. руб.	608,1	1082,1	747,0	753,4	122,9

В программном средстве реализован модуль корреляционно-регрессионного анализа, который состоит из 3 частей: корреляционного анализа, парной линейной регрессии и множественной регрессии.

Корреляционно-регрессионный анализ позволяет определить среднее изменение результативного признака под влиянием одного или нескольких факторов, а также найти меру тесноты связи результативного признака с одним или всем комплексом включенных в анализ факторов.

Для начала была рассчитана корреляционная матрица для показателей $x_2 - x_{10}$, а также построена тепловая карта⁵ (рис. 5).

На тепловой карте насыщенность цвета указывает на тесноту связи между парой переменных: чем более насыщенный цвет, тем ближе значение

⁵ How to create an interactive correlation matrix heatmap in R. [электронный ресурс] – Режим доступа: <https://www.datanovia.com/en/blog/how-to-create-an-interactive-correlation-matrix-heatmap-in-r/> (дата обращения: 01.11.2022). – Текст: электронный

коэффициента корреляции к 1 (красный цвет) или к -1 (синий цвет). При этом красный цвет указывает на положительную связь, а синий – на отрицательную.

Корреляционная матрица

	x2	x3	x4	x5	x6	x7	x8	x9	x10
x2	1.0000000	0.6758811	0.5501043	0.60199367	0.6937575	0.33793095	0.39214506	0.45417903	0.9182680
x3	0.6758811	1.0000000	0.6568941	0.48748210	0.9659085	0.21761137	0.55734726	0.31311656	0.7530789
x4	0.5501043	0.6568941	1.0000000	0.55518404	0.6919936	0.61370515	0.19019679	0.50365201	0.7109503
x5	0.6019937	0.4874821	0.5551840	1.00000000	0.5194322	0.58219552	-0.03650636	0.41801619	0.6991000
x6	0.6937575	0.9659085	0.6919936	0.51943219	1.0000000	0.28049486	0.58443535	0.30599972	0.7989132
x7	0.3379309	0.2176114	0.6137051	0.58219552	0.2804949	1.00000000	-0.06523678	0.69093956	0.4757520
x8	0.3921451	0.5573473	0.1901968	-0.03650636	0.5844354	-0.06523678	1.00000000	-0.08979962	0.4073656
x9	0.4541790	0.3131166	0.5036520	0.41801619	0.3059997	0.69093956	-0.08979962	1.00000000	0.4379449
x10	0.9182680	0.7530789	0.7109503	0.69909998	0.7989132	0.47575197	0.40736555	0.43794492	1.0000000

Интерактивная корреляционная тепловая карта

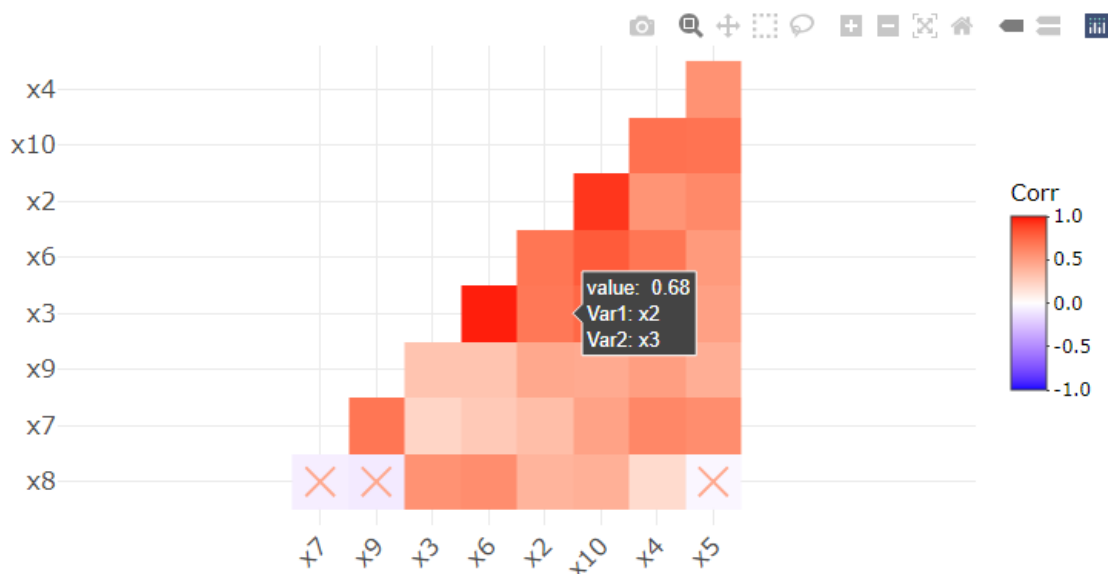


Рис. 5 – Результаты корреляционного анализа

В качестве результативного признака был взят x_2 , он имеет тесноту связи, выше средней с переменными $x_3, x_4, x_5, x_6, x_{10}$. Для построения парной линейной регрессии, в качестве независимой переменной был выбран показатель x_3 .

Сначала в приложении отображается диаграмма рассеяния для выбранных показателей (Рис. 6).

По графику на рис. 6 можно предположить прямую линейную зависимость между переменными средней тесноты связи. Одно наблюдение резко выделяется (выделено на рисунке), его можно было бы исключить из дальнейшего изучения.

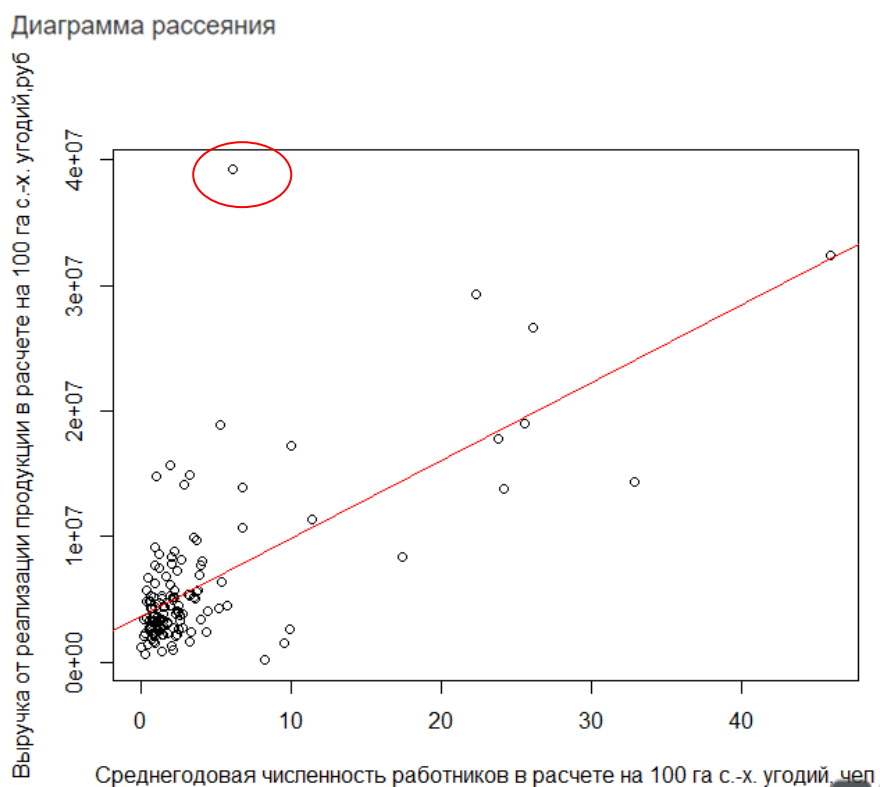


Рис. 6 – Диаграмма рассеяния

Далее выводятся результаты регрессионного анализа⁶ (Рис. 7).

Данные о полученной модели линейной регрессии

	Estimate	Std. Error	t value	Pr(> t)
** (Intercept) **	3677279	442619	8.308	9.849e-14
x3	620891	58707	10.58	2.397e-19

Observations	Residual Std. Error	R^2	Adjusted R^2
135	4462363	0.4568	0.4527

Table: Fitting linear model: x2 ~ x3

Рис. 7 – Результаты регрессионного анализа

⁶ Кабаков Р. И. R в действии. Анализ и визуализация данных в программе R / пер. с англ. П. А. Волновой. – М.: ДМК Пресс, 2014. – 588 с. – Текст: непосредственный

Уравнение регрессии имеет вид: $y = 3677279 + 620891x$. Коэффициент корреляции равен 0,68, что свидетельствует о наличии связи выше среднего уровня. Коэффициент полной регрессии показывает, что при увеличении среднегодовой численности работников на 1 человека в расчете на 100 га с.-х. угодий выручка увеличится на 621 тыс. руб. в расчете на 100 га с.-х. угодий. В модели коэффициент детерминации R^2 равен 0,46, т.е. 46% вариации выручки объясняется вариацией среднегодовой численности работников.

Также были выведены результаты дисперсионного анализа и доверительные интервалы для параметров уравнения (Рис. 8).

Дисперсионный анализ

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x3	1	2.227e+15	2.227e+15	111.9	2.397e-19
Residuals	133	2.648e+15	1.991e+13	NA	NA

Table: Analysis of Variance Table

NULL

Доверительные интервалы

	2.5 %	97.5 %
(Intercept)	2801796.8	4552761.5
x3	504769.8	737011.7

Рис. 8 – Результаты дисперсионного анализа и доверительные интервалы для парной линейной регрессии

Из рис. 8 следует, что фактическое значение критерия F-Фишера = 111,9 в данном случае получено достоверное уравнение регрессии при уровне значимости менее 0,01%, поскольку фактическая значимость критерия F-Фишера близка к нулю.

Судя по рассчитанным доверительным интервалам для параметров, коэффициент полной регрессии будет находиться в пределах от 504769,8 до

737011,7 руб. в расчете на 100 га с.-х. угодий при уровне вероятности суждения 97,5%.

Далее в приложении можно рассчитать множественную регрессию. Для этого необходимо исключить мультиколлинеарность факторов, поэтому в качестве результативного признака примем также x_2 , а в качестве факторов x_3 и x_5 .

Между переменными x_3 и x_5 мультиколлинеарность отсутствует, т.к. коэффициент корреляции (0,49) меньше 0,7.

В результате в веб-приложении можно посмотреть данные о полученной модели, результаты дисперсионного анализа и доверительные интервалы параметров (Рис. 9).

Данные о полученной модели множественной регрессии

	Estimate	Std. Error	t value	Pr(> t)
** (Intercept) **	3087392	417203	7.4	1.42e-11
x3	460813	61141	7.537	6.847e-12
x5	1114	207.4	5.371	3.429e-07

Observations	Residual Std. Error	SR ² S	Adjusted SR ² S
135	4057760	0.6542	0.6475

Table: Fitting linear model: $x_2 \sim x_3 + x_5$

Рис. 9 – Результаты множественной регрессии с факторными признаками x_3 , x_5

Была получена модель множественной линейной регрессии:

$$y = 3087392 + 460813x_1 + 1114x_2.$$

Коэффициент детерминации равен 0,65, т.е. 65% вариации уровня выручки объясняется вариацией обеспеченности работниками и энергоресурсами в расчете на единицу площади.

Получено уравнение регрессии со всеми достоверными параметрами (фактическая значимость критерия t-Стьюдента по всем параметрам близка к нулю).

Коэффициент чистой регрессии b_1 показывает, что при увеличении среднегодовой численности работников на 1 человека в расчете на 100 га с.-х. угодий выручка увеличится на 460813 руб. в расчете на 100 га с.-х. угодий при прочих равных условиях.

Параметр b_2 показывает, что при увеличении обеспеченности энергоресурсами на 1 л.с. в расчете на 100 га с.-х. угодий выручка увеличится на 1114 руб. в расчете на 100 га с.-х. угодий при прочих равных условиях.

Ниже приведены рассчитанные доверительные интервалы (Ошибка! Источник ссылки не найден.).

Доверительные интервалы

	2.5 %	97.5 %
(Intercept)	2262123.1012	3912660.588
x3	339869.6290	581756.276
x5	703.5946	1524.051

Рис. 10 – Доверительные интервалы для оценок параметров уравнения регрессии

Автоматизация анализа данных отчетности приводит к сокращению временных затрат, упрощению проведения анализа и получению более точных результатов, минимизируя или полностью исключая ошибки, связанные с человеческим фактором. Полученные по итогу анализа сведения могут оказать помощь в развитии отрасли, например, при выявлении типичных групп сельскохозяйственных организаций или при необходимости построения прогноза результативной переменной по ряду факторов.

Библиографический список:

1. Бессонов, А.С. Интерфейсы автоматизированных систем. Методические указания: методические указания / А.С. Бессонов. – Москва: РТУ МИРЭА, 2020. – 61 с. – Текст: непосредственный

2. Об утверждении форм отчета о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса за 2019 год и срока его представления: Приказ Минсельхоза России от 03.12.2019 № 669. – Доступ из справ.-правовой системы «КонсультантПлюс». – Текст: электронный

3. Кабаков, Р.И. R в действии. Анализ и визуализация данных в программе R / пер. с англ. П. А. Волновой. – М.: ДМК Пресс, 2014. – 588 с. – Текст: непосредственный

4. Кластерный анализ в R. – URL: <https://mymagictools.blogspot.com/2015/07/r.html> (дата обращения: 01.11.2022). – Текст: электронный

5. Analysis of Variance (ANOVA) in R. – URL: <https://www.instructables.com/Analysis-of-Variance-ANOVA-in-R/> (дата обращения: 01.11.2022). – Текст: электронный

6. How to create an interactive correlation matrix heatmap in R. – URL: <https://www.datanovia.com/en/blog/how-to-create-an-interactive-correlation-matrix-heatmap-in-r/> (дата обращения: 01.11.2022). – Текст: электронный

References:

1. Bessonov, A.S. Interfaces of automated systems. Guidelines: Guidelines / A.S. Bessonov. – Moscow: RTU MIREA, 2020. – 61 p. – Text: direct

2. On approval of the forms of the report on the financial and economic condition of agricultural producers for 2019 and the deadline for its submission: Order of the Ministry of Agriculture of Russia dated 03.12.2019 № 669. – Access from the ConsultantPlus legal reference system. – Text: electronic

3. Kabakov, R. I. R in action. Analysis and visualization of data in the R program / tr. from English. P. A. Volnova. – М.: ДМК Пресс, 2014. – 588 p. – Text: direct

4. Cluster analysis in R. – URL: <https://mymagictools.blogspot.com/2015/07/r.html> (Accessed: 11/01/2022). – Text: electronic

5. Analysis of Variance (ANOVA) in R. – URL: <https://www.instructables.com/Analysis-of-Variance-ANOVA-in-R/> (Accessed: 11/01/2022). – Text: electronic

6. How to create an interactive correlation matrix heatmap in R. – URL: <https://www.datanovia.com/en/blog/how-to-create-an-interactive-correlation-matrix-heatmap-in-r/> (date of access: 01.11.2022). – Text: electronic

Для цитирования: Уколова А.В., Статистический анализ данных отчетности о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса/ Уколова А.В., Быков Д.В., Витковская А.М. // Российский экономический интернет-журнал. – 2022. – № 4. URL:
© Уколова А.В., Быков Д.В., Витковская А.М., Российский экономический интернет-журнал 2022, № 4.