

СТАТИСТИЧЕСКИЙ АНАЛИЗ УРОЖАЙНОСТИ В РЕСПУБЛИКЕ БЕЛАРУСЬ МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

Белько И.В.,

доктор физико-математических наук, профессор,

Криштапович Е.А.,

Сапун О.Л.,

кандидат педагогических наук, доцент,

Белорусский государственный аграрный технический университет

Урожайность сельскохозяйственных культур является важным показателем экономики страны. Целью нашей работы является статистический анализ динамики и прогноз урожайности на 2020 г. зерновых в Республике Беларусь за период с 2013 по 2019 г. Использовались данные за 2013–2019 гг. по урожайности и восьми основным показателям (табл. 1), которые коррелируют с урожайностью (табл. 2). Данные о погоде¹ разделены на три уровня: плохая, умеренная и хорошая. Удобрения NPK (азотные, фосфорные и калийные) объединены в один показатель – минеральные удобрения.

¹ URL: <http://www.hmn.ru/index.php?index=8&value=26850>.

Таблица 1

Исходные данные по основным экономическим показателям

Год	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	Y
2013	94,8	147,9	197,3	100,7	15,5	50	34	32	57	274	29,7
2014	64,9	111,3	199,3	107,5	15,5	55,4	60	28	36	236	36,7
2015	88,6	101,9	199	99,8	16,7	53,1	39	26	57	209	36,5
2016	76	109,4	236	106,5	18	54,1	58	29	35	158	31,5
2017	116,3	120,9	276	108,9	17	29,3	67	30	25	155	33,2
2018	98,9	126,9	328	99,2	15,6	28,4	48	45	29	168	26,8
2019	110,3	143,5	345	108,4	16,8	27,5	55	46	21	165	30,4

Источник: составлено по данным: Сельское хозяйство Республики Беларусь. 2020. Минск: Национальный статистический комитет Республики Беларусь.

Таблица 2

Матрица корреляций

	Инвестиции	Индекс	Средние цены	Производительность труда	Экспорт	Производство продукции	Трактора	Плохая погода	Умеренная погода	Хорошая погода	Минеральные удобрения	Урожайность
Инвестиции	1											
Индекс	0,54	1										
Средние цены	0,66	0,42	1									
Производительность труда	0,07	-0,02	0,20	1								
Экспорт	0,11	-0,36	0,13	0,45	1							
Производство продукции	-0,52	-0,49	-0,71	0,32	-0,08	1						
Трактора	-0,86	-0,49	-0,92	-0,17	0,01	0,63	1					
Плохая погода	0,06	-0,27	0,33	0,86	0,42	0,16	-0,30	1				
Умеренная погода	0,52	0,64	0,89	-0,06	-0,20	-0,75	-0,78	-0,01	1			
Хорошая погода	-0,37	-0,14	-0,81	-0,66	-0,26	0,33	0,72	-0,81	-0,58	1		
Минеральные удобрения	-0,41	0,22	-0,71	-0,39	-0,68	0,45	0,60	-0,64	-0,36	0,76	1	
Урожайность	-0,45	-0,67	-0,63	0,29	0,13	0,95	0,57	0,20	-0,77	0,29	0,23	1

Источник: протокол расчета в MS Excel.

Целям исследования отвечают также следующие показатели: инвестиции в основной капитал – X_1 ; индекс цен сельскохозяйственной продукции – X_2 ; средние цены – X_3 ; производительность труда – X_4 ; экспорт – X_5 ; трактора и комбайны – X_6 ; погода – X_7, X_8, X_9 ; минеральные удобрения – X_{10} ; урожайность – Y .

При этом показатель объема производства зерновых не использовался, поскольку его динамика почти идентична динамике урожайности (рис. 1, 2).

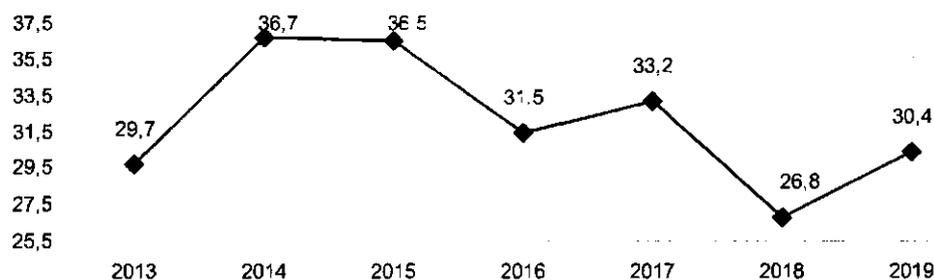


Рис. 1. Динамика урожайности зерновых в Республике Беларусь за 2013–2019 гг., ц/га

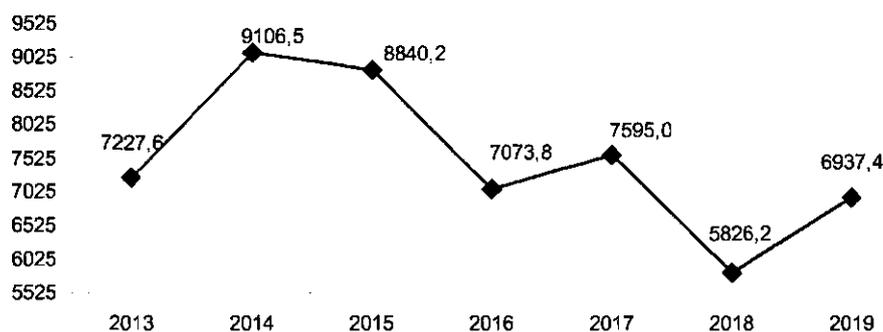


Рис. 2. Динамика валового сбора зерновых в Республике Беларусь за 2013–2019 гг., т

Особенностью использования показателей урожайности является их годовая периодичность, точный же учет факторов, которые могут влиять на урожайность, начал осуществляться позднее. Таким образом, имеется большое количество факторов с малым количеством наблюдений, и это обстоятельство оказывает влияние на качество статистических моделей.

Ранее нами проводился анализ динамики урожайности зерновых в Республике Польша (Белько, Криштапович, Лапко, 2013. С. 124). На первом этапе исследования строилась общая модель множественной регрессии с использованием всех факторов. Ввиду большого количества рассматриваемых факторов и малого – наблюдений качество модели было неудовлетворительным. Влияние большинства факторов не оценивается, и отсутствует значимость коэффициентов модели. Как и следовало ожидать, при проведении многофакторного регрессионного анализа возникла проблема мультиколлинеарности: распределение остатков модели не удовлетворяло нужным условиям. Для устранения этой проблемы существует классический метод главных компонент, использованный нами. Согласно этому методу (Сошникова, Тамашевич, Узбе, Шефер, 1999), по элементам матрицы корреляций строятся новые переменные – главные факторы, лишенные мультиколлинеарности.

При построении главных факторов (компонент) и при вычислениях прогнозов использовался компьютерный пакет SPSS и Microsoft Excel (Мидлтон, 2005). Главные факторы являются латентными переменными, они носят вербальный характер и строятся рекуррентным способом по максимуму объясненной дисперсии. Их недостаток – отсутствие непосредственной интерпретации по исходным факторам.

Для построения главных факторов применялась матрица корреляций исходных переменных с переходом к их нормированным значениям. Поиск главных компонент сводился к задаче последовательного выделения первой главной компоненты, обладающей максимальной дисперсией, второй главной компоненты, имеющей вторую по величине дисперсию, и т.д. Эта задача решалась в пакете SPSS. Главные компоненты находились как собственные векторы преобразования, заданного введенной выше матрицей корреляций.

Значения трех главных компонент, полученных по десяти переменным, приведены в табл. 3.

Таблица 3

Значения трех главных компонент и прогноз урожайности

Год	FAC1	FAC2	FAC3	Урожайность
2013	-1,0378	1,3827	0,4862	29,7
2014	-0,7387	-0,7004	1,7093	36,7
2015	-1,0804	-0,1063	-1,4303	36,5
2016	-0,1172	-1,3545	-0,6117	31,5
2017	0,9361	-0,6973	0,0543	33,2
2018	0,6698	1,0347	-0,5497	26,8
2019	1,3681	0,4412	0,3418	30,4
2020	1,72156	0,007816	-0,49522	28,84

Источник: протокол расчета в пакете SPSS.

В табл. 4 приведены собственные значения векторов главных компонент и объясненная дисперсия. Кумулятивная вариация выбранных трех первых главных компонент объясняет 87,995% всей вариации.

Таблица 4

Полная объясненная дисперсия для трех главных компонент

Комп- нента	Начальные собственные значения			Суммы квадратов нагрузок извлечения			Суммы квадратов нагрузок вращения		
	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %
1	4,946	49,459	49,459	4,946	49,459	49,459	4,261	42,61	42,61
2	2,826	28,255	77,715	2,826	28,255	77,715	2,631	26,305	68,915
3	1,028	10,28	87,995	1,028	10,28	87,995	1,908	19,08	87,995
4	0,725	7,252	95,247						
5	0,427	4,273	99,521						
6	0,048	0,479	100						
7	1,37E-15	1,37E-14	100						
8	-2,93E-17	-2,93E-16	100						
9	-1,14E-16	-1,14E-15	100						
10	-6,40E-16	-6,40E-15	100						

Источник: авторская разработка.

Выбор первых трех главных компонент для использования подтверждается данными рис. 3 с учетом того, что их собственные значения больше 1.

Проведенная по трем главным факторам регрессия на урожайность позволила получить прогноз, выявить его ошибку и установить интервал прогноза (24,83; 32,85).

Прогнозные значения факторов (см. табл. 3) заданы предсказанием по временным рядам. Прогнозное значение урожайности, равное 28,84 ц/га, получено по протоколу регрессии по этим факторам (рис. 4).

Для сравнительного анализа и сокращения количества факторов выбраны 6 показателей из десяти: X_1 , X_2 , X_3 , X_6 , X_8 , X_{10} , и вычисления повторяются. В этом случае для итога достаточно двух главных компонент. Значения полученных главных факторов, предсказания их значений для прогноза и прогноз урожайности по уравнению регрессии, равный 28,97 ц/га, приведены в табл. 5.

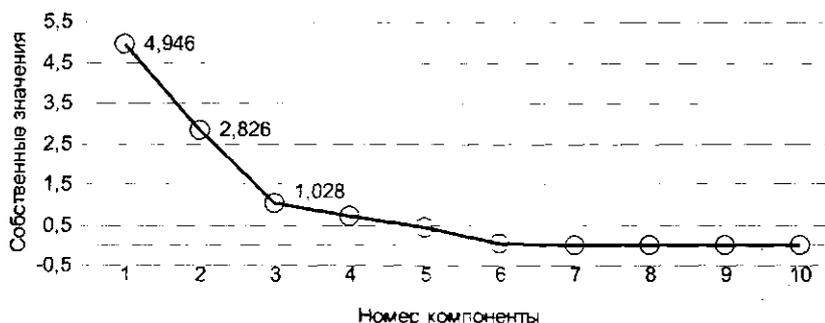


Рис. 3. Объясненная кумулятивная дисперсия для трех главных компонент

Источник: протокол расчета в пакете SPSS.

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0,7871639								
R Square	0,619627005								
Adjusted R Sq	0,23925401								
Standard Error	3,16144797								
Observations	7								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	3	48,84431163	16,28143721	1,628998413	0,349152043				
Residual	3	29,9842598	9,994753268						
Total	6	78,82857143							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>	
Intercept	32,1143	1,1949	26,8758	0,0001	28,3115	35,9170	28,3115	35,9170	
	-1,7395	1,2907	-1,3477	0,2705	-5,8469	2,3680	-5,8469	2,3680	
	-2,1994	1,2907	-1,7041	0,1869	-6,3068	1,9080	-6,3068	1,9080	
	0,5269	1,2907	0,4082	0,7105	-3,5806	4,6343	-3,5806	4,6343	

Рис. 4. Протокол регрессии по трем факторам

Источник: протокол расчета в пакете SPSS.

Таблица 5

Значения двух главных компонент и прогноз урожайности

Год	FAC_1	FAC_2	Урожайность
2013	1,0252	1,7843	29,7
2014	-0,7982	0,8212	36,7
2015	-1,0153	0,1807	36,5
2016	-1,1203	-0,3678	31,5
2017	0,0064	-0,8949	33,2
2018	0,6219	-0,8411	26,8
2019	1,2802	-0,6823	30,4
2020	0,6610	-1,6857	28,97

Источник: авторская разработка.

Выбор используемых двух факторов, значения которых приведены в табл. 5, подтверждается данными табл. 6 и рис. 5, 6. Эти два фактора объясняют 88,382% дисперсии. Прогноз урожайности равен 28,97, интервал прогноза – (24,95; 32,99).

Таблица 6

Полная объясненная дисперсия для двух главных компонент

Компонента	Начальные собственные значения			Суммы квадратов нагрузок извлечения			Суммы квадратов нагрузок вращения		
	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %	Всего	% дисперсии	Кумулятивный %
1	4,002	66,698	66,698	4,002	66,698	66,698	2,802	46,693	46,693
2	1,301	21,685	88,382	1,301	21,685	88,382	2,501	41,690	88,382
3	0,545	9,086	97,468						
4	0,114	1,902	99,370						
5	0,036	0,597	99,966						
6	0,002	0,034	100,000						

Источник: авторская разработка.



Рис. 5. Объясненная кумулятивная дисперсия для двух главных компонент

Источник: авторская разработка.

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0,746271								
R Square	0,556921								
Adjusted R Sq	0,335381								
Standard Error	2,954966								
Observations	7								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	2	43,90127	21,95063684	2,513866032	0,196319148				
Residual	4	34,9273	8,73182444						
Total	6	78,82857							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>	
Intercept	32,1143	1,1169	28,7538	0,0000	29,0134	35,2152	29,0134	35,2152	
F1	-2,5655	1,2064	-2,1266	0,1006	-5,9149	0,7839	-5,9149	0,7839	
F2	0,8574	1,2064	0,7107	0,5165	-2,4920	4,2068	-2,4920	4,2068	

Рис. 6. Протокол регрессии по двум факторам

Источник: авторская разработка.

В качестве недостатка отметим, что коэффициенты регрессий урожайности по главным факторам имеют низкую значимость: максимальная вероятность – 0,5–0,7. Коэффициенты детерминации R^2 равны 0,62 и 0,56 соответственно. Вместе с тем прогнозные результаты практически совпадают (рис. 7), и динамику урожайности в достаточной степени характеризуют главные компоненты.

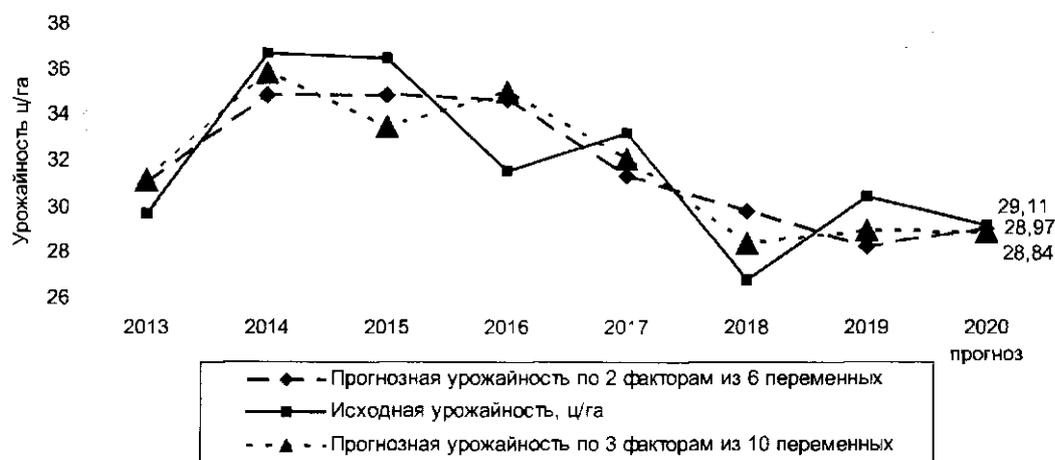


Рис. 7. Графики прогнозных урожайностей в сравнении с исходной, по предсказанию равной 29,11 ц/га

Источник: авторская разработка.

С помощью предложенных средств можно проводить статистический анализ основных факторов повышения урожайности зерновых в Республике Беларусь.

Литература

Белько И.В., Криштапович Е.А., Лапко В.В. 2013. Применение математических моделей для анализа показателей урожайности зерновых культур. *Экономика, моделирование, прогнозирование*. Сборник научных трудов. Минск: НИЭИ Минэкономки Республики Беларусь. Выпуск 7. С. 127–131.

Мидлтон М.Р. 2005. *Анализ статистических данных с использованием Microsoft Excel: практикум*. Москва: Бином. 199 с.

Сошникова Л.А., Тамашевич В.Н., Узбе Г., Шефер М. 1999. *Многомерный статистический анализ в экономике*. Москва: ЮНИТИ. 598 с.

Статья поступила 14. 01. 2021 г.

