

РЕГРЕССИОННЫЕ МОДЕЛИ В ОПРЕДЕЛЕНИИ ХЛЕБОПЕКАРНЫХ СВОЙСТВ ЯРОВОЙ ПШЕНИЦЫ

Плеханова Л.В.¹, Шевцова Л.Н.²

¹ *Красноярский научно-исследовательский институт сельского хозяйства
ФИЦ КНЦ СО РАН, Красноярск, Россия*

² *Красноярский государственный аграрный университет, Красноярск, Россия*

В представленной работе была сделана попытка разработки регрессионных моделей для составления хлебопекарных оценок зерна мягкой яровой пшеницы. Исследования показали перспективы использования регрессионных моделей для выбора селекционных линий с потенциально высокими хлебопекарными свойствами зерна, максимально снизив субъективное мнение экспертов.

Ключевые слова: *зерно, мягкой яровой пшеницы, хлебопекарные свойства зерна, корреляция, моделирование, регрессионные модели.*

REGRESSION MODELS IN PREDICTING THE BAKING PROPERTIES OF SPRING WHEAT

Plekhanova L.V.¹, Shevtsova L.N.²

¹ *Krasnoyarsk Agricultural Research Institute FRC KSC SB RAS*

² *Krasnoyarsk state agrarian university, Krasnoyarsk, Russia*

The attempt to study technological qualities of grain using various direct and indirect methods, as well as dozens of indicators that require a large amount of time, labor and grain was done. The constructed linear regression models showed the adequacy of the description and high comparability with the actual data. Studies have shown the prospects of using regression models for the selection of breeding lines with potentially high baking properties of grain.

Key words: *grain, soft spring wheat, baking properties of grain, correlation, modeling, regression models.*

Введение. В настоящее время использование математических моделей при оценке хлебопекарных свойств пшеницы вызывает повышенное внимание исследователей [2, 4]. Актуальность этих вопросов связана с практической возможностью производить выбор селекционных линий с потенциально высокими технологическими свойствами уже на ранних стадиях селекции.

Цель исследований. Построение и сравнение регрессионных моделей для определения хлебопекарных свойств мягкой яровой пшеницы.

Материал и методы исследований. Для проведения исследований использовали селекционный материал конкурсного сортоиспытания (КСИ)

2008-2010 гг. лаборатории селекции пшеницы, выращенный на полях стационара «Минино» Красноярского НИИ сельского хозяйства.

Оценку качества зерна мягкой яровой пшеницы КСИ проводили в соответствии с методиками национальных стандартов Российской Федерации и методов ИСО, методическими рекомендациями по 15 показателям; хлебопекарные свойства муки оценивались методом прямой выпечки хлеба из муки 70%-ного выхода с добавлением сахара, разработанным центральной лабораторией государственной комиссии по сортоиспытанию сельскохозяйственных культур [3].

Для математического моделирования использовали статистический пакет «Пакет анализа» MS Excel.

Результаты исследований и их обсуждение. Общую хлебопекарную оценку ($Y_{\text{факт}}$, балл) устанавливали по лабораторной выпечке хлеба, основанной на органолептических показателях – это форма, цвет и поверхность корки, эластичность, пористость, цвет мякиша, вкус и аромат хлеба.

Для моделирования хлебопекарных свойств яровой пшеницы были построены линейные и нелинейные уравнения множественной регрессии с варьированием набора факторов. При сравнении и выборе моделей были использованы статистические критерии из итоговых таблиц инструмента «Регрессия» («Пакет анализа» в Excel): сумма квадратов остатков ($SSR \rightarrow \min$), коэффициент детерминации ($R^2 \rightarrow 1$) и F-тест.

Первая часть наших исследований явилась продолжением работы [4] по регрессионному моделированию хлебопекарных свойств на основании восьми лабораторных показателей (факторов): содержание белка (X_1 , %), сила муки (X_2 , е.а), время до начала разжижения теста (X_3 , мин), разжижение теста (X_4 , е.ф), валориметрическая оценка (X_5 , %), объем хлеба (X_6 , см³), стекловидность зерна (X_7 , %) \bar{x}_7 , % $\bar{}$, выход муки (X_8 , %).

Для проверки наличия коллинеарности были построены корреляционные матрицы для всех факторов и общей хлебопекарной оценкой (Y) по годам исследований. Была выявлена коллинеарность между факторами X_3 и X_5 . Поэтому были построены регрессионные уравнения как с включением этих факторов так и попеременным исключением каждого из этих факторов. Так, в таблице 1 представлены значения факторов, обобщенный показатель хлебопекарной оценки ($Y_{\text{факт}}$) и рассчитанный (Y_1) по уравнению множественной линейной регрессии следующего вида:

$$Y_1 = 0,39202 - 0,00365x_1 - 0,000251x_2 - 0,00961x_3 + 0,00124x_4 + 0,00353x_5 + 0,00305x_6 + 0,01034x_7 + 0,01195x_8.$$

Таблица 1 – Показатели технологических качеств зерна мягкой яровой пшеницы КСИ-2009

Название образцов / сортов	Y_{факт}	X₁	X₂	X₃	X₄	X₅	X₆	X₇	X₈	Y₁	Y₂	Y₃	Y₄	Y₅
Ветлужанка	4,00	11,970	334,80	10,50	30,00	80,00	760,00	50,00	65,00	4,09	4,093	4,114	3,9652	3,8920
Мана 2	3,90	14,260	461,00	6,00	60,00	63,00	680,00	46,00	65,00	3,79	3,790	3,794	3,8677	3,8789
Минуса	3,50	14,170	734,40	7,50	50,00	70,00	620,00	49,00	63,00	3,54	3,546	3,492	3,6626	3,7326
Таежная нива	3,90	13,640	528,40	14,50	30,00	89,00	690,00	51,00	64,00	3,82	3,833	3,869	3,9810	4,0152
Новосибирская 29	4,00	14,640	737,70	10,50	50,00	80,00	760,00	49,00	62,50	3,97	3,965	4,091	3,9325	3,9891
К-174-1	4,00	14,110	513,40	10,00	50,00	79,00	700,00	51,00	64,00	3,88	3,876	3,917	3,8525	3,8584
К-91-1	3,10	11,590	535,60	4,70	70,00	58,00	510,00	43,00	65,00	3,24	3,246	3,217	3,5565	3,3905
К-142-4	4,10	13,580	577,50	5,80	40,00	64,00	780,00	56,00	63,70	4,14	4,138	4,096	4,0239	4,1873
Омская 33	3,80	13,370	327,60	8,40	50,00	72,00	660,00	50,00	66,00	3,82	3,812	3,800	3,5988	3,4132
К-64-2	3,00	12,070	408,70	10,00	90,00	76,00	490,00	43,00	64,80	3,24	3,235	2,996	3,3781	3,1665
К-265-4	3,50	11,230	456,50	7,30	80,00	67,00	550,00	47,00	67,00	3,47	3,463	3,406	3,6461	3,4198
К-195-5	4,10	10,580	281,20	4,80	90,00	56,00	760,00	38,00	59,50	3,97	3,969	4,058	3,8612	4,0891
Омская 32	3,80	13,250	579,40	4,50	100,00	52,00	700,00	52,00	62,90	3,89	3,896	3,857	3,8368	3,9145
К-195-2	3,70	13,340	523,80	4,30	130,00	50,00	600,00	54,00	69,20	3,72	3,724	3,753	3,6419	3,7894
К-186-1	3,70	13,140	476,80	2,80	70,00	51,00	650,00	51,00	68,10	3,79	3,791	3,686	3,7438	3,5987
К-254-1	4,10	12,960	368,80	3,50	90,00	50,00	750,00	42,00	66,90	4,03	4,032	4,140	3,7240	3,7044
К-227-2	4,10	12,980	623,90	6,00	90,00	61,00	700,00	48,00	71,80	3,95	3,942	3,948	3,8128	4,1497
К-227-4	3,80	13,320	476,80	5,50	80,00	59,00	620,00	52,00	72,00	3,77	3,767	3,864	3,7521	3,7244
К-257-6	3,80	12,710	536,90	6,00	50,00	65,00	690,00	48,00	64,00	3,81	3,810	3,757	3,5617	3,5873
К-152-1	4,30	14,880	389,10	3,00	100,00	47,00	790,00	55,00	61,70	4,22	4,221	4,303	4,0708	4,1045
К-1	4,30	15,050	568,90	4,00	80,00	54,00	800,00	53,00	69,30	4,26	4,260	4,226	3,9615	4,1756
К-2	4,00	15,340	540,80	8,00	30,00	70,00	720,00	55,00	64,80	3,95	3,955	3,965	4,0148	4,1196
К-3	3,60	15,770	349,90	6,00	70,00	62,00	580,00	53,00	65,90	3,60	3,598	3,531	3,7457	3,7803
К-4	3,50	15,320	359,00	6,30	40,00	66,00	630,00	53,00	68,10	3,75	3,749	3,611	3,8718	3,7349
К-5	3,90	15,450	456,50	5,30	50,00	63,00	720,00	51,00	64,80	3,95	3,947	3,953	3,9671	4,0240
К-6	3,60	14,910	337,40	3,30	60,00	54,00	630,00	51,00	62,50	3,68	3,684	3,658	3,9004	3,8037
К-7	4,00	13,800	313,30	5,30	30,00	70,00	700,00	52,00	64,30	3,94	3,915	3,837	3,8298	3,8517
К-8	3,80	14,400	411,40	9,00	80,00	73,00	620,00	52,00	66,20	3,73	3,716	3,797	3,8652	3,7352
К-233-2	3,40	12,450	394,40	3,00	40,00	59,00	550,00	43,00	67,70	3,41	3,400	3,404	3,5257	3,4754
К-159-1	3,80	14,390	756,70	6,00	80,00	61,00	750,00	46,00	71,00	4,02	4,018	3,862	3,6971	3,7215
В-1	3,60	13,200	459,80	7,00	50,00	68,00	670,00	48,00	67,70	3,81	3,809	3,765	3,4275	3,6058
В-2	3,40	13,900	713,50	4,00	80,00	58,00	500,00	49,00	67,00	3,26	3,256	3,346	3,5830	3,5804
В-3	3,50	13,710	521,20	6,50	50,00	67,00	500,00	48,00	70,40	3,31	3,309	3,432	3,4493	3,5673
В-4	3,50	13,270	483,30	5,70	50,00	63,00	520,00	50,00	69,70	3,39	3,392	3,510	3,6073	3,5243
В-5	3,60	13,760	309,30	5,70	50,00	53,00	590,00	48,00	68,40	3,57	3,607	3,561	3,3294	3,6495
В-6	3,50	14,110	449,90	3,50	80,00	53,00	560,00	53,00	69,50	3,57	3,568	3,507	3,6687	3,5626
В-7	3,40	13,480	390,40	5,70	50,00	64,00	480,00	49,00	68,60	3,27	3,269	3,423	3,4858	3,5077

В-8	3,40	14,110	399,60	4,50	60,00	60,00	520,00	51,00	67,70	3,41	3,402	3,423	3,6840	3,6163
-----	------	--------	--------	------	-------	-------	--------	-------	-------	------	-------	-------	--------	--------

Аналогичные вычисления были выполнены для сортообразцов КСИ 2008 и 2010 годов. Построены линейные регрессионные модели, при этом из линейных моделей исключались коллинеарные факторы X_3 или X_5 .

Для построения нелинейной регрессионной модели была выбрана формула [1], учитывающая линейное, нелинейное и взаимное влияние факторов:

$$Y = a + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i=1}^k \sum_{j=1}^k b_{ij} x_i x_j + \dots$$

Составленное уравнение нелинейной множественной регрессии в общем виде представлено следующим выражением:

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_6x_6 + b_7x_7 + b_8x_8 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{44}x_4^2 + b_{66}x_6^2 + b_{77}x_7^2 + b_{88}x_8^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{16}x_1x_6 + b_{17}x_1x_7 + b_{18}x_1x_8 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{26}x_2x_6 + b_{27}x_2x_7 + b_{28}x_2x_8 + b_{34}x_3x_4 + b_{36}x_3x_6 + b_{37}x_3x_7 + b_{38}x_3x_8 + b_{46}x_4x_6 + b_{47}x_4x_7 + b_{48}x_4x_8 + b_{67}x_6x_7 + b_{68}x_6x_8 + b_{78}x_7x_8.$$

В данное уравнение не вошли фактор X_5 и сложные факторы взаимного влияния трех и более факторов. Далее было проведено переобозначение (линеализация) факторов и в MS Excel продолжен регрессионный анализ. При составлении нелинейной множественной регрессионной модели были устранены члены уравнений, вклад которых в результат оказался несущественный. Используя рассчитанные MS Excel коэффициенты регрессии, выведена третья регрессионная модель (Y_3):

$$Y_3 = - 1,92868 + 1,54688x_1 - 0,00025x_2 - 0,80988x_3 + 0,00427x_6 - 0,26741x_7 + 0,02951x_8 - 0,01335x_1^2 + 0,0000068x_6^2 - 0,00170x_7^2 + 0,000163x_1x_4 - 0,01796x_1x_8 + 0,00917x_3x_7 + 0,00555x_3x_8 - 0,00015x_6x_8 + 0,00590x_7x_8.$$

Сравнение оценочных параметров моделей представлено в таблице 2.

Таблица 2 – Сравнение статистических показателей регрессионных моделей по годам наблюдений

Регрессионная модель	Год	R ²	SSR	F-test
Y_1 – линейная, 8 факторов	2008	0,868	0,213	12,33 > 2,64
	2009	0,864	0,465	23,11 > 2,28
	2010	0,923	0,255	34,47 > 2,37
Y_2 - линейная, 7 факторов (без X_5)	2008	0,860	0,225	14,10 > 2,65
	2009	0,922	0,257	40,80 > 2,43
	2010	0,863	0,467	27,13 > 2,33
Y_2 - линейная, 7 факторов (без X_3)	2008	0,857	0,230	13,80 > 2,42
	2009	-	-	-
	2010	0,922	0,258	40,53 > 2,42
Y_3 – множественная регрессия (нелинейная)	2008	0,959	0,066	10,28 > 3,49
	2009	0,939	0,206	22,86 > 2,15
	2010	0,958	0,139	21,39 > 2,38

На основании F - статистики можно утверждать, что с вероятностью $p=0,95$ все полученные регрессионные уравнения являются достоверными. При этом третья модель имеет минимальную ошибку ($SSR \rightarrow \min$) и высокий коэффициент детерминации (R^2). Визуальное сопоставление показателей хлебопекарной оценки (Y , балл) по результатам пробной выпечки с показателями, полученными по регрессионным моделям (Y_1, Y_2, Y_3) также показывает достаточно устойчивое совпадение по образцам (табл. 1).

Из линейных моделей ($Y_1, 8$ факторов) лучшие статистические характеристики показала модель Y_1 2010 года (табл. 2). Поэтому она была выбрана для проверки составления прогнозов хлебопекарных свойств у сортообразцов питомников 2008 и 2009 гг. Прогноз хлебопекарных свойств образцов КСИ 2009 г. по линейному регрессионному уравнению $Y_1(2010) = 1,4092 - 0,0197x_1 + 0,000015x_2 - 0,0375x_3 + 0,000196x_4 + 0,00502x_5 + 0,00323x_6 + 0,00649x_7 + 0,00266x_8$ показал высокое сходство с фактическими показателями и представлен на рисунке 1.



Рисунок 1 – Сравнение графиков хлебопекарных оценок образцов КСИ 2009: фактически ряд данных и расчетный по модели Y_1 2010 г.

Таким образом, при использовании лабораторных показателей, включающих объем хлеба ($X_6, \text{см}^3$), расчет хлебопекарных оценок (Y) по регрессионным уравнениям позволяет избежать субъективного влияния на определение хлебопекарных баллов. Показана высокая достоверность линейных регрессионных моделей, простота и доступность их составления в MS Excel.

Следующий этап нашей работы включает исследования по составлению регрессионных моделей и возможности прогнозирования хлебопекарных свойств на основании лабораторных факторов, полученных без выпечки хлеба. Т.е. из регрессионных уравнений исключается фактор объем хлеба ($X_6, \text{см}^3$). По корреляционным матрицам было выявлено его высокое влияние на обобщенный показатель хлебопекарной оценки (коэффициенты корреляции больше 0,9). Для моделирования были отобраны следующие показатели: содержание белка ($X_1, \%$), сила муки ($X_2, e.a$), время до начала разжижения теста ($X_3, \text{мин}$), разжижение теста ($X_4, e.f$), валориметрическая оценка ($X_5, \%$), **масса 1000 зерен ($X_6, \text{г}$)**, стекловидность зерна ($X_7, \%$) $\bar{x}_7, \%$, выход муки ($X_8, \%$), натура ($X_9, \text{г-л}$), ВПС ($X_{10}, \%$). По результатам корреляционной матрицы с новыми факторами из регрессионного уравнения исключили факторы

X_2 , X_4 (слабо коррелированы с $Y_{\text{факт}}$) и X_5 (коллинеарность с X_3).

С оставленными факторами рассчитаны регрессионные уравнения: линейное Y_4 и нелинейное Y_5 . Сравнение этих моделей (табл. 3) показало низкую достоверность линейной регрессии и, следовательно, невозможность использования такой модели для прогнозирования хлебопекарных свойств.

Таблица 3 – Сравнение статистических показателей регрессионных моделей с новыми факторами (без выпечки хлеба) на примере КСИ 2009

Регрессионная модель	R^2	SSR	F-test
Y_4 – линейная	0,3503	2,227	2,31 < 2,32
Y_5 - нелинейная	0,671	1,603	2,20 > 2,12

Прогноз хлебопекарных свойств, выполненный по нелинейному регрессионному уравнению показал, что зависимость у фактических (эмпирических) и теоретических (расчетных) величин имеет схожую корреляцию. Наглядное графическое представление этой корреляции отображено на рисунке 2.

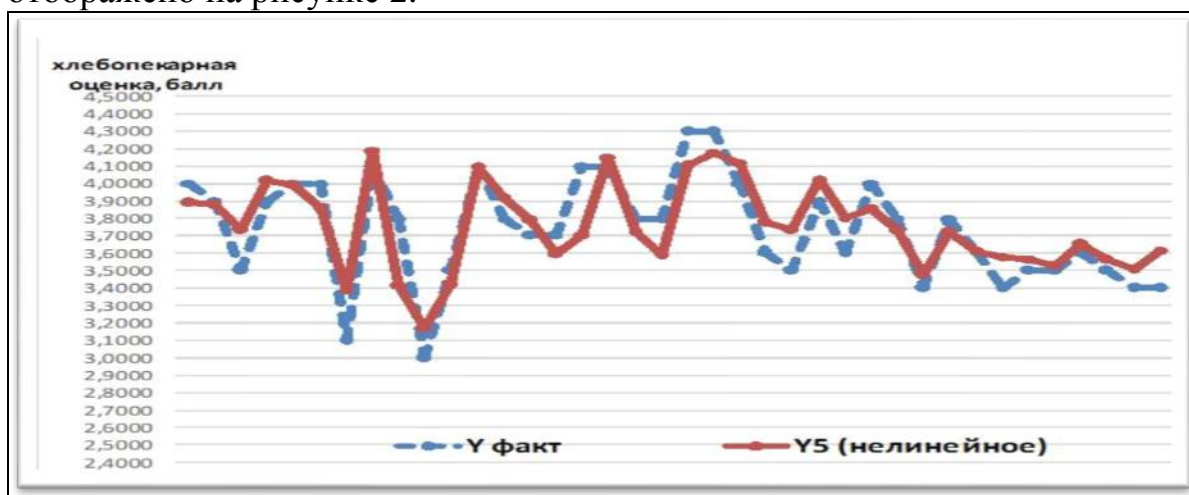


Рисунок 2 – Графики хлебопекарных оценок – фактических и рассчитанных по нелинейной регрессионной модели

Выводы. Таким образом, расчет коэффициентов регрессии и статистическая обработка полученных данных дала возможность получать регрессионные уравнения, которые адекватно описывают общую хлебопекарную оценку зерна в зависимости от выбранных исходных факторов. Сравнительная оценка моделей выявила более высокую достоверность множественной регрессионной модели, учитывающей линейное, нелинейное и взаимное влияние факторов.

Линейная модель, составленная с включением лабораторного показателя объем хлеба (см^3), может зарекомендовать себя как достоверный экспресс-метод, который определяет обобщенные хлебопекарные оценки зерна и при этом исключает субъективное мнение экспертов.

Для составления ориентировочных прогнозов хлебопекарных свойств без выпечки хлеба возможно использование нелинейных регрессионных уравнений. Разработка таких моделей имеет перспективы их дальнейшего изучения, в том числе по оптимальному подбору показателей (факторов) с учетом их сложных взаимодействий.

Литература

1. Козлов А.Ю., Шишов В.Ф. Пакет анализа MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов / Под. ред. проф. В.С. Мхитаряна.- М.: ЮНИТИ-ДАНА, 2003.- 139с.

2. Мелешкина, Е.П. Развитие системы оценки хлебопекарных свойств зерна пшеницы при его производстве и переработке. [Текст]: автореферат дисс. ... д-ра техн. Наук.: 05.18.01 / Е.П. Мелешкина – М., 2006. - 52 с.

3. Методические рекомендации по оценке качества зерна. / сост. А.А.Созинов, Н.И.Блохин, И.И.Василенко и др. М., 1977. 171 с.

4. Плеханова Л.В., Шевцова Л.Н. Моделирование хлебопекарных показателей зерна мягкой яровой пшеницы //Вестник КрасГАУ. – 2018. - № 2.