

Хозяйственно полезные признаки картофеля, позволяющие прогнозировать его урожайность

© 2025. Е. П. Шанина, Д. А. Оберюхтин✉

ФГБНУ «Уральский федеральный аграрный научно-исследовательский центр
Уральского отделения Российской академии наук», г. Екатеринбург,
Российская Федерация

Научные исследования по селекции картофеля направлены на повышение урожайности и устойчивости сортов к болезням и вредителям, улучшение качества продукции. На основе полученных результатов разрабатываются статистические модели, которые позволяют не только создать систему прогнозирования исследуемого параметра, но и объяснить вклад отдельных признаков, формирующих этот параметр. Цель исследования – определить оптимальную модель прогноза урожайности картофеля на основе анализа хозяйственно полезных признаков (средний вес клубня в кусте; количество клубней в кусте; содержание в клубнях крахмала, протеина, общих и редуцирующих сахаров, витамина С и нитратов) 100 сортов картофеля урожая 2024 г. (Свердловская область). В настоящей работе составлена модель линейной регрессии урожайности картофеля (т/га) с тремя предикторами – средний вес клубня (г), количество клубней (шт.) и натуральный логарифм содержания нитратов в клубне (мг/кг). Полученная модель смогла предсказать 77,9 % дисперсии данных ($R^2 = 0,785$, $R^2_{adj} = 0,779$, $p < 0,001$), а степень соответствия модели оказалась идеальной, поскольку значение средней абсолютной ошибки MAPE < 10 %. Проверка предположений модели регрессии проводили по алгоритму LINE: линейность, независимость, нормальность, однородность. Этот алгоритм доказал адекватность полученной модели. Благодаря методу линейной регрессии показано, что при увеличении среднего веса товарного клубня на 1 г и количества клубней в кусте на 1 шт. урожайность картофеля возрастает соответственно на 0,314 и 2,386 т/га. Однако при повышении концентрации нитратов на каждую единицу натурального логарифма урожайность снижается на 3,63 т/га. В то время как вес клубня и количество клубней имели функциональную связь с урожайностью, содержание нитратов имело лишь косвенное отношение. По литературным данным, концентрация нитратов в клубнях картофеля в конце уборки урожая представляет «итоговую оценку» эффективности использования азота конкретным сортом в течение всего вегетативного периода. Полученной моделью обусловлены критерии отбора при селекции высокопродуктивных сортов картофеля: отбор генотипов с большим количеством и весом клубней, низким содержанием остаточных нитратов в клубнях при стандартном режиме азотного питания.

Ключевые слова: *Solanum tuberosum* L., урожай картофеля, нитраты, климатические условия, модель линейной регрессии, статистическая модель, агрономия

Благодарности: работа выполнена при поддержке Минобрнауки РФ в рамках Государственного задания ФГБНУ «Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук» (тема № 0532-2021-0008).

Авторы благодарят рецензентов за их вклад в экспертную оценку этой работы.

Конфликт интересов: авторы заявили об отсутствии конфликта интересов.

Для цитирования: Шанина Е. П., Оберюхтин Д. А. Хозяйственно полезные признаки картофеля, позволяющие прогнозировать его урожайность. Аграрная наука Евро-Северо-Востока. 2025;26(6):1226–1240.

DOI: <https://doi.org/10.30766/2072-9081.2025.26.6.1226-1240>

Поступила: 29.04.2025

Принята к публикации: 16.10.2025

Опубликована онлайн: 26.12.2025

Agronomic traits of potato enabling yield prediction

© 2025. Elena P. Shanina, Denis A. Oberiukhtin✉

Ural Federal Agrarian Scientific Research Centre, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russian Federation

Scientific research on potato breeding is aimed at increasing crop yields and the resistance of cultivars to diseases and pests, as well as improving product quality. Based on the results obtained, statistical models are developed that allow not only to create a system for forecasting the parameter under study, but also to explain the contribution of individual traits that form this parameter. The aim of the study is to determine the optimal model for forecasting potato yield based on the analysis of agronomic traits (average tuber weight per a potato plant; the number of tubers per a plant; content of starch, protein, total and reducing sugars, vitamin C and nitrates in tubers) of 100 potato cultivars harvested in 2024 (Sverdlovsk Region). In this work, a linear regression model of potato yield (t/ha) was developed with three predictors: average tuber weight (g), the number of tubers (units) and natural logarithm of nitrate content in tubers (mg/kg). The resulting model was able to predict 77.9 % of the data dispersion ($R^2 = 0.785$, $R^2_{adj} = 0.779$, $p < 0.001$), by this the degree of model fit was ideal, since the mean absolute error MAPE was less than 10%. The regression model assumptions were tested using the LINE algorithm: linearity, independence,

normality, homogeneity. This algorithm proved the adequacy of the obtained model. Using the linear regression method, it was shown that with an increase in the average weight of marketable tubers by 1 g and the number of tubers by 1 unit, potato yield increased by 0.314 and 2.386 t/ha, respectively. However, with an increase in nitrate concentration per unit of natural logarithm, the yield decreased by 3.63 t/ha. While tuber weight and tuber number had a functional relationship with the yield, nitrate content had only an indirect relationship. According to the literature, the nitrate concentration in potato tubers at the end of the harvest represents a «final assessment» of the efficiency of nitrogen use by a particular cultivar throughout the growing season. The model obtained determines the selection criteria for breeding high-yielding potato cultivars: selection of genotypes with a large number and weight of tubers, low residual nitrate content in tubers under standard nitrogen nutrition conditions.

Keywords: *Solanum tuberosum* L., potato yield, nitrates, climatic conditions, linear regression model, statistical model, agronomy

Acknowledgements: the research was carried out under the support of the Ministry of Science and Higher Education of the Russian Federation within the state assignment of Ural Federal Agrarian Scientific Research Centre, Ural Branch of the Russian Academy of Sciences (theme No. 0532-2021-0008).

The authors thank the reviewers for their contribution to the peer review of this work.

Conflict of interest: the authors stated that there was no conflict of interest.

For citation: Shanina E. P., Oberiukhtin D. A. Agronomic traits of potato enabling yield prediction. *Agrarnaya nauka Evro-Severo-Vostoka* = Agricultural Science Euro-North-East. 2025;26(6):1226–1240. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2025.26.6.1226-1240>

Received: 29.04.2025

Accepted for publication: 16.10.2025

Published online: 26.12.2025

Картофель – это один из основных источников питания для большого количества людей во всем мире [1]. Точное прогнозирование урожайности картофеля позволяет рационально использовать оборудование и методы возделывания, обеспечивая тем самым продовольственную безопасность регионов страны. Достижение целей управляемого возделывания культурных растений лежит через цифровизацию – новый этап развития сельского хозяйства [2]. Существующие подходы моделирования условий урожайности основаны на гео-климатических данных: температура воздуха, атмосферные осадки, суммарная инсоляция, кислотность и минеральный состав почв [3, 4, 5], которые обрабатываются с помощью методов множественной линейной регрессии [3, 5], машинного обучения [4] и нейронных сетей [5, 6].

Однако при выращивании картофеля селекционеры отслеживают не только агроклиматические условия, но и хозяйственно полезные признаки выводимых культур – скороспелость, урожайность, товарность клубней, количество и масса клубней на куст [7], содержание белка, сахаров, витамина С, нитратов [8]. Эти данные могут быть использованы для анализа уровня получаемой урожайности картофеля и корректировки программ селекции. Безусловно, сложные статистические методы машинного обучения и нейронных сетей позволяют с высокой точностью прогнозировать интересующий параметр [9], но объяснить полученные результаты с помощью современных знаний физиологии

и биохимии сельскохозяйственных растений бывает затруднительно. Поэтому использование более простых методов исследования – методов линейной регрессии позволяет не только достоверно прогнозировать интересующие величины, но и понимать логику полученных результатов.

Цель исследования – определить оптимальную модель прогноза урожайности на основе анализа хозяйственно полезных признаков сортов картофеля.

Научная новизна – установлено, что концентрация нитратов в клубнях картофеля после сбора урожая является надежным предиктором урожайности вместе со средним весом клубня и количеством клубней на куст.

Материал и методы. Результаты исследований получены на основе полевых экспериментов 2024 г. в Уральском научно-исследовательском институте сельского хозяйства – филиале ФГБНУ «Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук».

Исследования проводили на дерново-среднеподзолистой почве со следующей характеристикой: удельная масса – 2,1–2,6 г/см³; объёмная масса – 1,1–1,3 г/см³; гумус – 4,86 %; рН_{сол.} – 5,29; кислотность гидролитическая – 11,5 мг-экв/100 г; сумма поглощённых оснований – 17,9 мг-экв/100 г; азот общий – 0,20 %; азот легкогидролизуемый – 130 мг/кг; фосфор подвижный – 262 мг/кг и калий обменный – 146 мг/кг почвы.

Опыты заложены в соответствии с общепринятыми методиками^{1, 2} (по 20 растений каждого сорта в трехкратной повторности). Количество сортов раннеспелой группы – 35, среднеранней – 46, среднеспелой – 19 (всего 100 сортов). Опыт размещен в действующем трёхпольном севообороте, предшественник картофеля – сидеральный пар (клевер 1 г. п.).

Посадку картофеля проводили при физиологической спелости почвы – 15–17 мая; уборку – 27–31 августа. Внесение удобрений – под весеннюю культивацию в дозе N₈₀K₇₀P₇₀. Средства защиты: Титус (50 г/га), Лазурит (200 мл/га), Эспера (150 мл/га), Ширлан (0,4 л/га), обработка растений картофеля во время вегетации.

Урожай и его структуру при уборке учитывали взвешиванием, крахмал в лабораторных условиях – по удельному весу³. В аналитической лаборатории проведена оценка биохимических показателей: содержание азота по Кьельдалю⁴, сахаров – эбулиостатически⁵, аскорбиновая кислота (витамин С) – по Мурри⁶, нитраты – ионометрическим методом⁷.

Метеоусловия, включающие температуру воздуха и осадки, фиксировали по декадам каждого месяца с мая по сентябрь. Из полученных данных строили комбинированные диаграммы в Microsoft Excel 2024.

Статистический анализ проводили с помощью программного обеспечения Jamovi⁸. В качестве мер центральной тенденции использовали величины медианы (median) и среднего (mean). Для описания разброса данных определяли минимальные (min) и максимальные (max) значения, стандартное отклонение (SD). В качестве показателя асимметрии данных использовали коэффициент асимметрии (As). Однородность данных определяли по коэффициенту

вариации (CV), который рассчитывали по следующей формуле:

$$CV = \frac{SD}{mean}.$$

Распределение данных оценивали с помощью критерия Шапиро-Уилка (Shapiro-Wilk test). В случае, если полученное значение p (p-value) было больше порогового 0,05, то считали, что данные распределены нормально; в противном случае – распределение данных отличное от нормального. Для визуализации показателей строили гистограмму распределения и график квантиль-квантиль (QQ-plot). Формирование колоколообразной формы гистограммы распределения и положение точек вдоль одной прямой на QQ-plot дополнительно подтверждают гипотезу о нормальном распределении данных.

С целью определения экстремальных значений в данных пользовались правилом Тьюке (правило «ящик с усами»)⁹. Для этого у каждого показателя рассчитывали квартиль 1 (Q₁) и квартиль 3 (Q₃), а также разницу между двумя квартилями (межквартильный размах, IQR). По формуле Тьюке экстремальными значениями считаются все значения, выходящие из диапазона:

$$[Q_1 - 1,5 * IQR; IQR * 1,5 + Q_3].$$

Поиск взаимосвязей между исследуемыми признаками проводили с помощью корреляционного анализа. Для этого строили диаграммы рассеяния и рассчитывали коэффициент корреляции Пирсона. Качественную оценку силы связи коэффициента корреляции определяли по шкале Чеддока: от 0 до 0,3 – очень слабая; от 0,3 до 0,5 – слабая; от 0,5 до 0,7 – средняя; от 0,7 до 0,9 – сильная; от 0,9 до 1,0 – очень сильная.

¹Методика государственного сортоиспытания сельскохозяйственных культур: В 7 вып. Госкомиссия по сортоиспытанию с.-х. культур при МСХ СССР. Вып. 4: Картофель, овощные и бахчевые культуры. Разраб. акад. ВАСХНИЛ Д. Д. Брежнев, агр., кандидаты с.-х. наук В. А. Бакулина, Н. К. Давидич и др. М.: Колос, 1975. 182 с.

²Методические указания по поддержанию и изучению мировой коллекции картофеля. Сост. акад. С. М. Букасов, д-ра с.-х. наук А. Г. Зыкина, А. Я. Камераз и др. Всесоюз. акад. с.-х. наук им. В. И. Ленина, ВНИИ растениеводства им. Н. И. Вавилова. Л.: ВИР, 1976. 27 с.

³ГОСТ 7194-81. Картофель свежий. Правила приемки и методы определения качества. М.: Стандартинформ, 2010. 13 с. URL: <https://files.stroyinf.ru/Data2/1/4294822/4294822617.pdf>

⁴ГОСТ 13496.4-2019. Корма, комбикорма, комбикормовое сырье. Методы определения содержания азота и сырого протеина. М.: Стандартинформ, 2019. 20 с. URL: <https://files.stroyinf.ru/Data2/1/4293727/4293727957.pdf>

⁵ГОСТ 8756.13-87. Продукты переработки плодов и овощей. Методы определения сахаров. М.: Стандартинформ, 2010. 11 с. URL: <https://files.stroyinf.ru/Data2/1/4294821/4294821427.pdf>

⁶ГОСТ 24556-89. Продукты переработки плодов и овощей. Методы определения витамина С. М.: ИПК Изд-во стандартов, 2003. 11 с. URL: <https://files.stroyinf.ru/Data2/1/4294829/4294829760.pdf>

⁷ГОСТ 29270-95. Продукты переработки плодов и овощей. Методы определения нитратов. М.: Стандартинформ, 2010. 15 с. URL: <https://files.stroyinf.ru/Data2/1/4294825/4294825356.pdf>

⁸The jamovi project. Jamovi. (Version 2.6 – 2024). [Computer Software]. Retrieved from: <https://www.jamovi.org>

⁹Гайдышев И. П. Моделирование стохастических и детерминированных систем: руководство пользователя программы AtteStat. Курган, 2015. С. 435–436.

Знак коэффициента корреляции говорит о направлении связи: положительный указывает на прямо пропорциональную зависимость, отрицательный – на обратно пропорциональную. Корреляцию считали достоверной при значении $p < 0,05$.

Данные, имеющие достоверную корреляцию, включали в регрессионный анализ. В модели линейной регрессии методом наименьших квадратов для каждого предиктора рассчитывали вес, стандартную ошибку среднего (SE), 95%-й доверительный интервал, t -значение статистики, p -значение. Предиктор считали достоверным при значении $p < 0,05$. Для получившейся модели определяли коэффициент детерминации R^2 и скорректированный коэффициент детерминации R^2_{adj} , показывающие какой процент дисперсии данных способна объяснить текущая модель.

Проверку предположений модели регрессии проводили по алгоритму LINE: linearity (линейность), independence (независимость), normality (нормальность), equality or homogeneity (качество или однородность). Проверку на нормальность остатков проводили с помощью построения графика QQ-plot и расчёта статистики Колмогорова-Смирнова. Дополнительно проводили проверку на коллинеарность по расчёту коэффициента увеличения дисперсии VIF (variance inflation factor). В биологических исследованиях пороговое значение VIF установлено на уровне 2.

Для расчёта ошибки аппроксимации использовали формулу средней абсолютной ошибки (Mean absolute percentage error, MAPE), выраженной в %:

$$MAPE = 100 \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|,$$

где A_t – фактическое значение, F_t – прогнозируемое значение, n – количество наблюдений.

По версии¹⁰, если значение MAPE менее 10 %, то степень соответствия модели идеальна, диапазон от 10 до 20 % указывает на хорошее соответствие, от 20 до 50 % – на приемлемый уровень, выше 50 % – означает низкую точность модели и не допускается к практическому использованию.

Результаты и их обсуждение. Период вегетации картофеля в 2024 г. характеризовался резкими колебаниями температуры воздуха и неравномерным распределением осадков (рис. 1, 2). В мае температурные показатели воздуха находились на уровне среднесезонных данных, но при избыточном увлажнении, что привело к поздней посадке картофеля – с 29 мая. В июне – июле осадков было достаточно для формирования необходимого количества клубней, но длительное переувлажнение способствовало появлению фитофтороза – с 27 июля и далее наблюдали эпифитотию данного заболевания, к началу августа ботва восприимчивых сортов была полностью уничтожена.

Описательная статистика исходных данных по сортам картофеля, включенным в исследование, приведена в таблице 1.

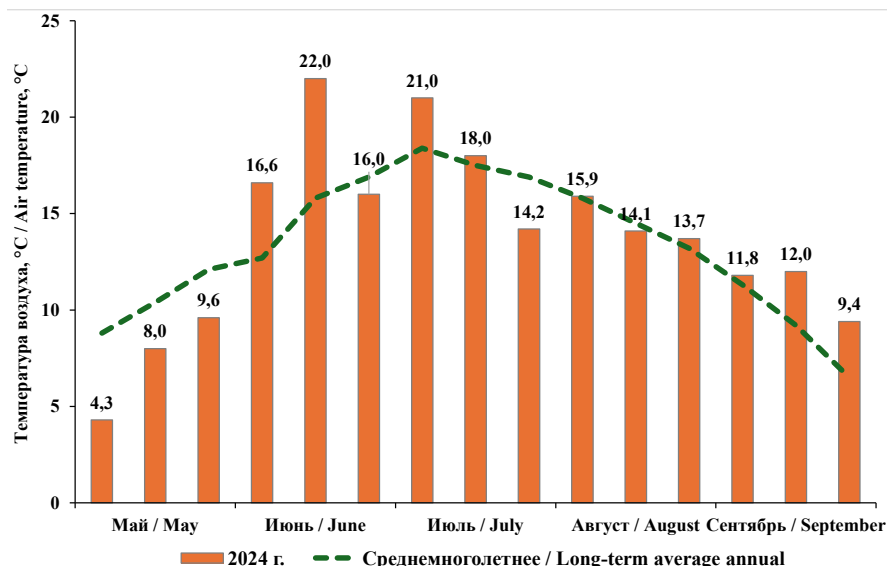


Рис. 1. Средние значения температуры воздуха по декадам в период вегетации картофеля (г. Екатеринбург, 2024 г.) /

Fig 1. Average air temperature values by decades during the potato growing season (Yekaterinburg, 2024)

¹⁰Montaño Moreno J. J., Palmer Pol A., Sesé Abad A., Cajal Blasco B. Using the R-MAPE index as a resistant measure of forecast accuracy. Psicothema. 2013;25(4):500–506. DOI: <https://doi.org/10.7334/psicothema2013.23>

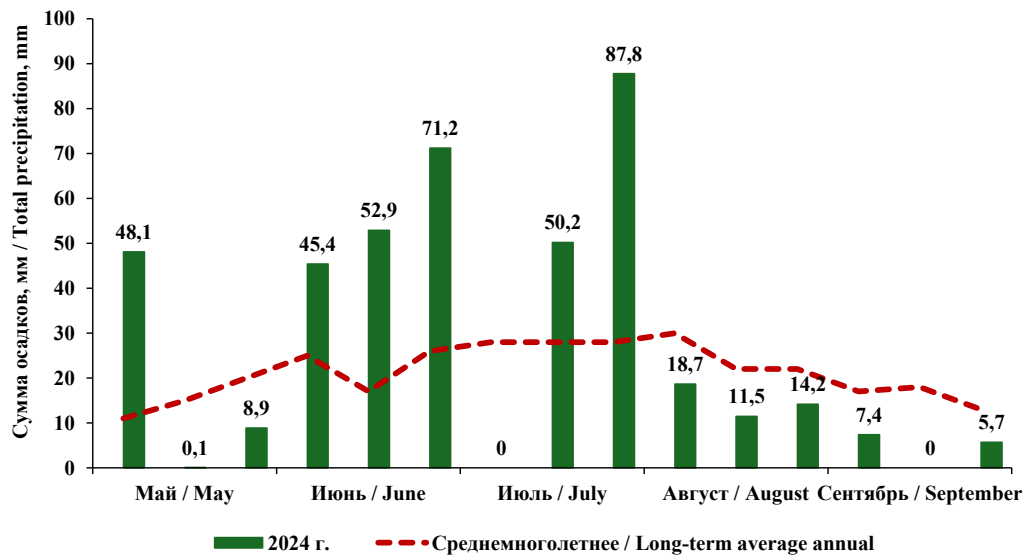


Рис. 2. Сумма осадков по декадам в период вегетации картофеля (г. Екатеринбург, 2024 г.) /
Fig. 2. Total precipitation by decades during the potato growing season (Yekaterinburg, 2024)

Таблица 1 – Описательная статистика хозяйственно полезных признаков сортов картофеля, выращенных в 2024 г. (n = 100) /

Table 1 – Descriptive statistics of agronomic traits of potato cultivars grown in 2024 (n = 100)

Параметр / Parameter	Min	Max	Median	Mean	SD	CV	As	Shapiro-Wilk test, p-value
Урожайность, т/га / Yield, t/ha	16,6	64,1	34,9	35,6	8,46	0,24	0,417	0,366
Средний вес клубня в кусте, г / Average tuber weight per a plant, g	52,0	144,0	90,0	94,5	20,4	0,22	0,284	0,291
Количество клубней в кусте, шт. / Tuber quantity per a plant, units	4,0	17,8	10,3	10,6	2,47	0,23	0,425	0,085
Крахмал, % / Starch, %	9,40	17,00	12,7	12,7	1,82	0,14	0,242	0,055
Протеин, % / Protein, %	1,06	2,94	1,88	1,92	0,421	0,22	0,512	0,019
Общих сахаров, % / Total sugar, %	0,13	0,83	0,32	0,363	0,15	0,41	1,17	< 0,001
Редуцирующих сахаров, % / Reducing sugar, %	0,00	0,39	0,140	0,144	0,0753	0,52	0,0574	0,007
Витамин С, мг% / Vitamin C, mg%	11,4	61,6	15,6	17,9	8,72	0,49	3,61	< 0,001
Нитраты, мг/кг / Nitrates, mg/kg	35,1	189,0	88,3	91,6	32,3	0,35	0,625	0,016

Примечания: Min – минимальное значение в выборке, Max – максимальное значение в выборке, Median – медиана, Mean – среднее значение, SD – стандартное отклонение, CV– коэффициент вариации (SD/Mean), As – асимметрия, Shapiro-Wilk test – тест Шапиро-Уилка /

Notes: Min – minimum value in the sample, Max – maximum value in the sample, Mean – average value, SD – standard deviation, CV – coefficient of variation (SD/Mean), As – asymmetry

Из данных таблицы 1 видно, что исходные значения урожайности, среднего веса клубня, количества клубней и содержания крахмала имеют распределение данных, близкое к нормальному (Shapiro-Wilk test $p>0,05$; median \approx mean), в то время как содержание протеина, общих сахаров, редуцирующих сахаров, витамина С, нитратов имеют распределение, отличное от нормального (Shapiro-Wilk test $p<0,05$). Также исходные данные урожайности, среднего веса клубня, количества клубней в кусте,

крахмала, протеина свидетельствуют об однородности ($CV<0,3$), тогда как данные по содержанию общих и редуцирующих сахаров, витамина С, нитратов не выполняли этого требования ($CV>0,3$). Наконец, данные об урожайности, среднем весе клубня, количестве клубней, содержании крахмала, редуцирующих сахаров отличались умеренной правосторонней асимметрией ($0<As<0,5$), данные по протеину, общим сахарам, витамину С, нитратам – существенной правосторонней асимметрией ($As>0,5$).

Для обнаружения данных, сильно отклоняющихся от общего массива наблюдений, рассчитаны Q_1 , Q_3 , IQR, а также определены границы допустимых значений: нижняя (Low. Ref.) и верхняя (Upp. Ref.). Значения вне

полученного диапазона считаются экстремальными. Благодаря такому подходу из дальнейшего анализа были исключены 19 наблюдений. Результаты проведенных расчётов приведены в таблице 2.

Таблица 2 – Расчёт интервала для исключения выбросов /
Table 2 – Calculation of the interval to exclude outliers

Параметр / Parameter	Q_1	Q_3	IQR	Low. Ref.	Upp. Ref.
Урожайность, т/га / Yield, t/ha	29,9	41,0	11,1	13,25	57,65
Средний вес клубня в кусте, г / Average tuber weight per a plant, g	81,8	109,0	27,5	40,55	150,25
Количество клубней в кусте, шт. / Tuber quantity per a plant, units	8,88	12,00	3,15	4,16	16,73
Крахмал, % / Starch, %	11,20	14,00	2,78	7,03	18,17
Протеин, % / Protein, %	1,63	2,19	0,56	0,79	3,03
Общих сахаров, % / Total sugar, %	0,268	0,420	0,152	0,04	0,648
Редуцирующих сахаров, % / Reducing sugar, %	0,100	0,180	0,08	-0,02	0,3
Витамин С, мг% / Vitamin C, mg%	14,3	17,9	3,65	8,83	23,38
Нитраты, мг/кг / Nitrates, mg/kg	68,1	110,0	41,4	6,0	172,1

Примечания: Q_1 – квартиль 1; Q_3 – квартиль 3; IQR – межквартильный размах; Low. Ref. – минимальное значение, ниже которого наблюдаемое значение считается экстремальным; Upp. Ref. – максимальное значение, выше которого наблюдаемое значение считается экстремальным /

Notes: Q_1 – quartile 1, Q_3 – quartile 3, IQR – Interquartile range, Low. Ref. – the minimum value below which an observed value is considered extreme; Upp. Ref. – the maximum value above which an observed value is considered extreme

После исключения наблюдений с экстремальными значениями изучаемые признаки

были проанализированы с помощью построения гистограмм распределения и QQ-plot (рис. 3–11).

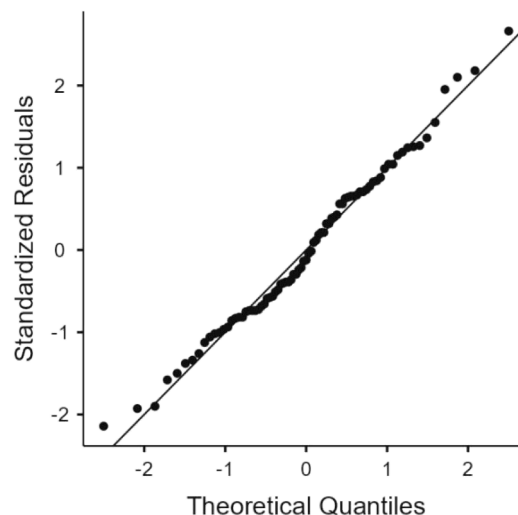
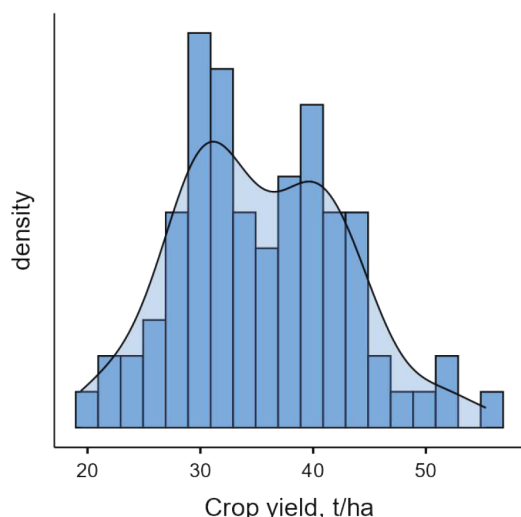


Рис. 3. Гистограмма распределения и QQ-plot для набора данных по урожайности картофеля после исключения выбросов, т/га /

Fig. 3. Distribution histogram and QQ-plot for the potato yield dataset after excluding outliers, t/ha

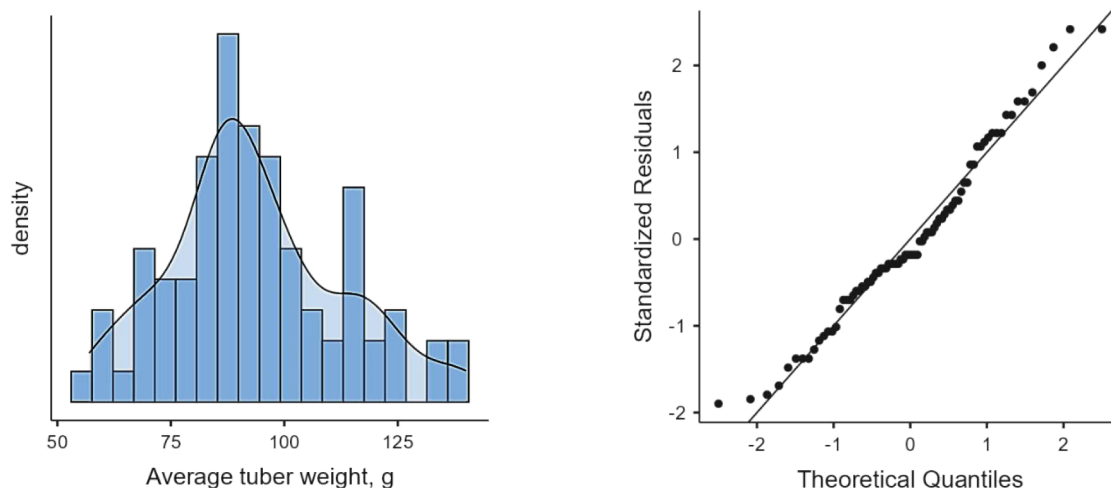


Рис. 4. Гистограмма распределения и QQ-plot для набора данных по среднему весу товарного клубня после исключения выбросов, г /

Fig. 4. Distribution histogram and QQ-plot for the average potato tuber weight dataset after excluding outliers, g

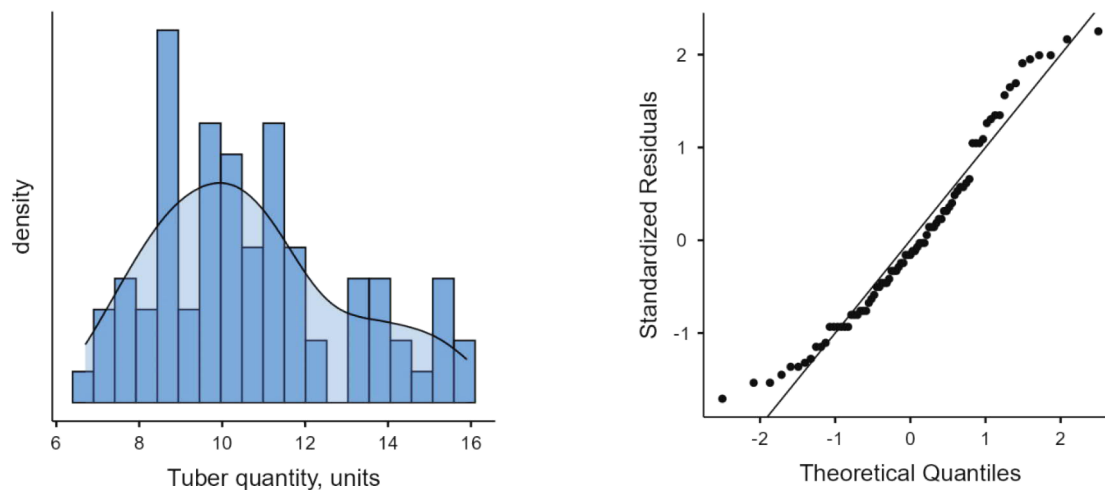


Рис. 5. Гистограмма распределения и QQ-plot для набора данных по количеству клубней в кусте после исключения выбросов, шт. /

Fig. 5. Distribution histogram and QQ-plot for the potato tuber quantity in a bush dataset after excluding outliers, units

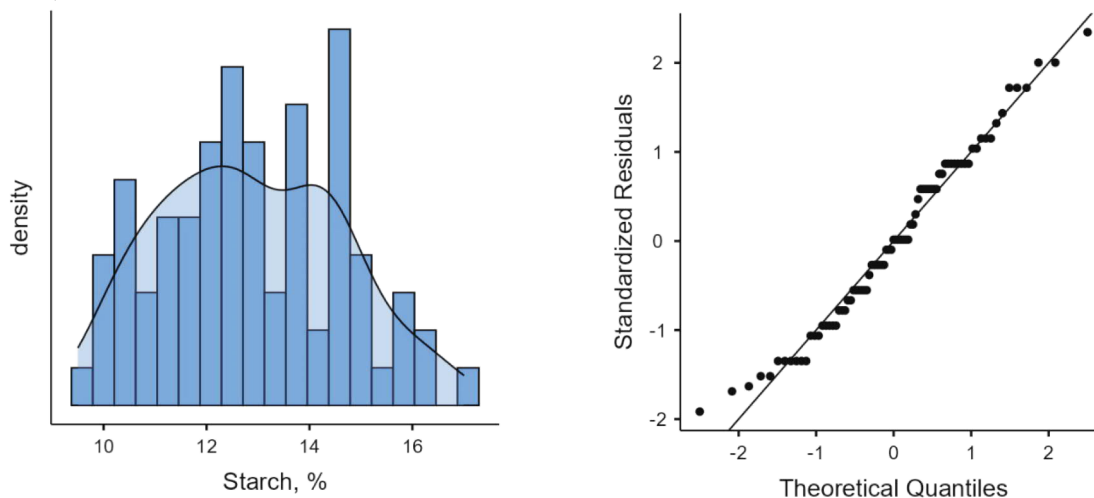


Рис. 6. Гистограмма распределения и QQ-plot для набора данных по содержанию крахмала в клубнях картофеля после исключения выбросов, % /

Fig. 6. Distribution histogram and QQ-plot for the potato tuber starch dataset after excluding outliers, %

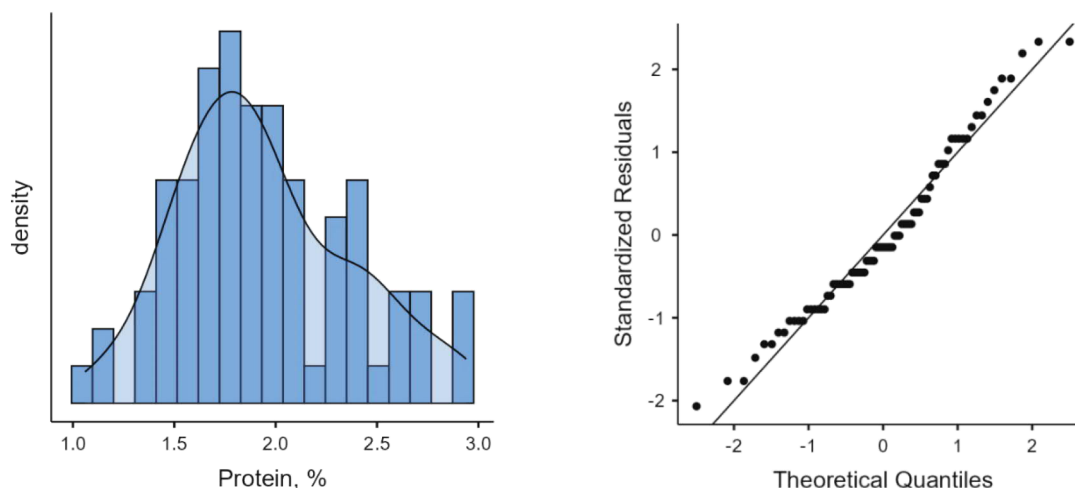


Рис. 7. Гистограмма распределения и QQ-plot для набора данных по содержанию протеина в клубнях картофеля после исключения выбросов, % /

Fig. 7. Distribution histogram and QQ-plot for the potato tuber protein dataset after excluding outliers, %

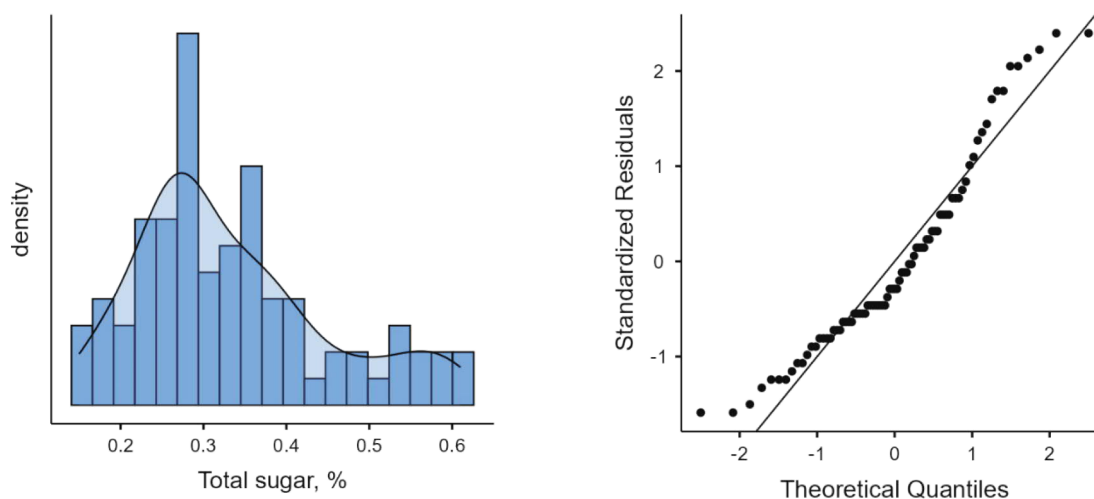


Рис. 8. Гистограмма распределения и QQ-plot для набора данных по содержанию общих сахаров в клубнях картофеля после исключения выбросов, % /

Fig. 8. Distribution histogram and QQ-plot for the potato tuber total sugar dataset after excluding outliers, %

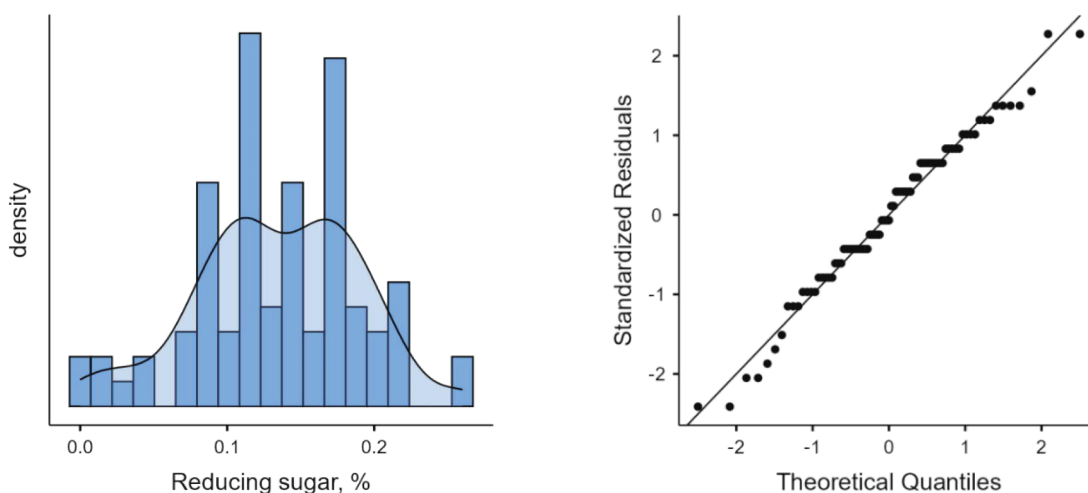


Рис. 9. Гистограмма распределения и QQ-plot для набора данных по содержанию редуцирующих сахаров в клубнях картофеля после исключения выбросов, % /

Fig. 9. Distribution histogram and QQ-plot for the potato tuber reducing sugar dataset after excluding outliers, %

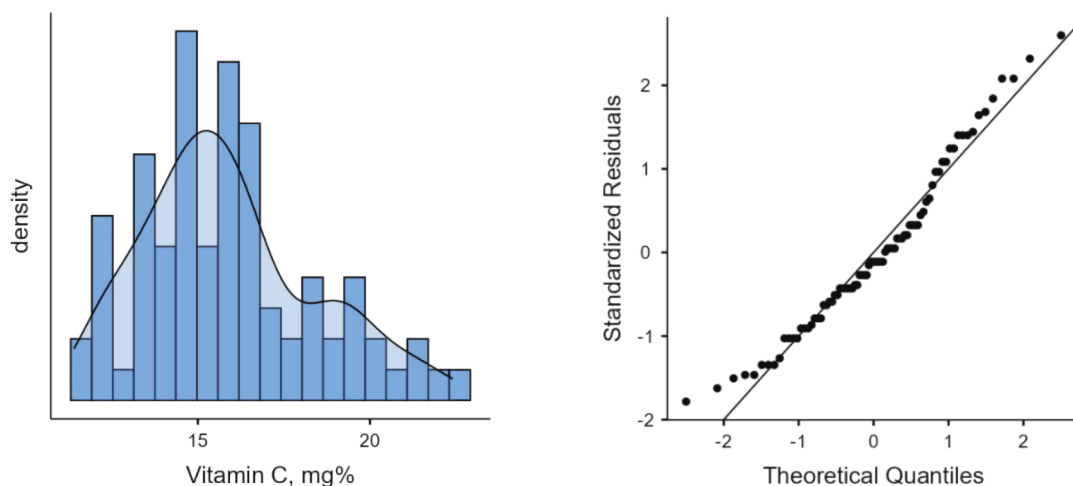


Рис. 10. Гистограмма распределения и QQ-plot для набора данных по содержанию витамина С в клубнях картофеля после исключения выбросов, мг% /

Fig. 10. Distribution histogram and QQ-plot for the potato tuber vitamin C dataset after excluding outliers, mg%

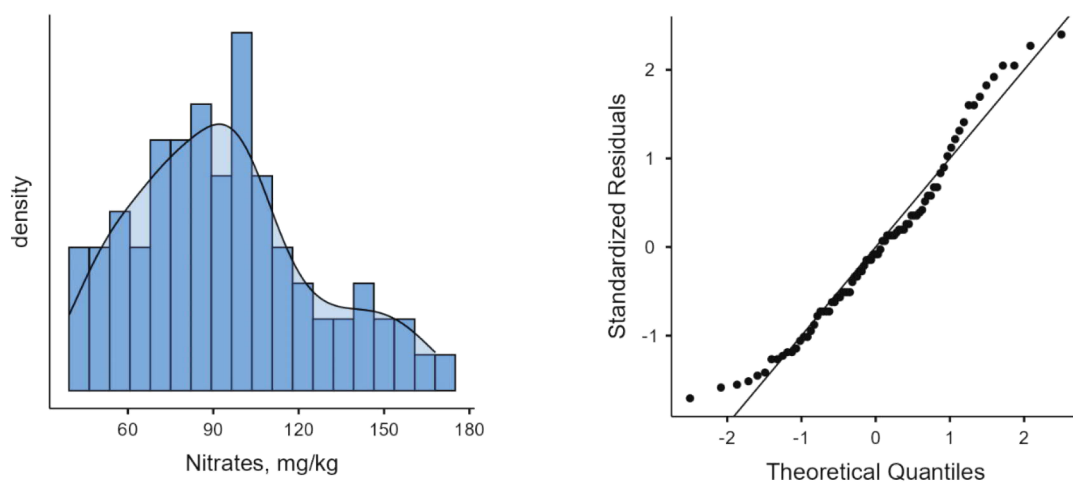


Рис. 11. Гистограмма распределения и QQ-plot для набора данных по содержанию нитратов в клубнях картофеля после исключения выбросов, мг/кг /

Fig. 11. Distribution histogram and QQ-plot for the potato tuber nitrates (mg/kg) dataset after excluding outliers, mg/kg

Из рисунков видно, что форма распределения данных на гистограмме, близкая к нормальной (колоколообразной), была практически у всех признаков, за исключением урожайности (рис. 3), у которой наблюдалось расщепление, в результате чего форма графика стала двугорбой. На графиках QQ-plot у исследуемых признаков большинство точек лежит на одной прямой. Все это подтверждает допущение о распределении данных, близком к нормальному, несмотря на то что проведенный тест Шапиро-Уилка не подтвердил этого допущения в отношении протеина, общих сахаров, редуцирующих сахаров, витамина С, нитратов. Предполагается, что тест Шапиро-Уилка не смог подтвердить нормальность данных ввиду асимметрии (табл. 1).

Корреляционный анализ по методу Пирсона показал, что урожайность прямо коррелирует

со средним весом клубня ($r = 0,570$, $p < 0,001$), количеством клубней ($r = 0,450$, $p < 0,001$) и обратно коррелирует с содержанием в клубнях нитратов ($r = -0,280$, $p < 0,05$). Результаты отражены на рисунке 12, диаграмма рассеяния – на рисунке 13. С другими признаками корреляции были слабыми и незначимыми.

После того, как были отобраны признаки, с которыми может быть взаимосвязана урожайность, приступили к построению статистической модели с помощью линейной регрессии. Значения урожайности, среднего веса товарного клубня и количества клубней брали в исходных величинах, а из значений нитратов извлекали натуральный логарифм (Ln). Результаты вычисленных коэффициентов линейной регрессии приведены в таблице 3.

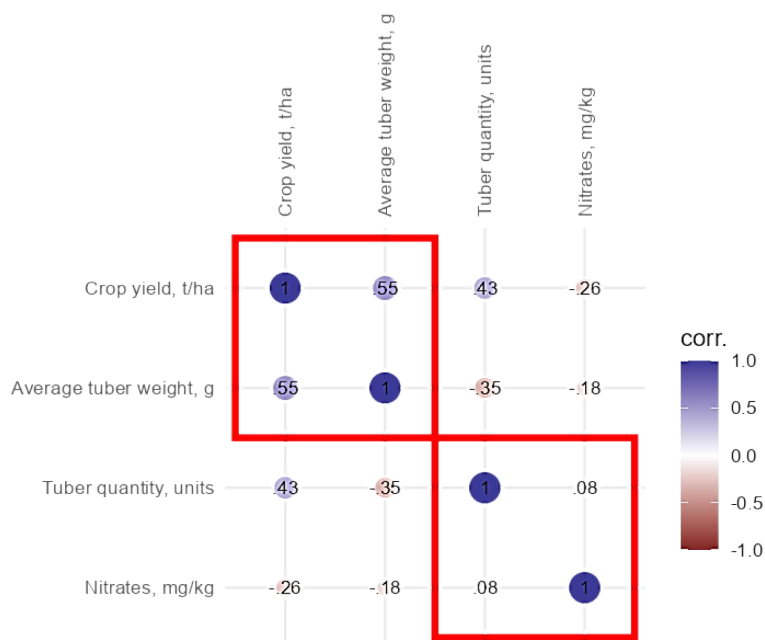


Рис. 12. Корреляционная матрица зависимости урожайности картофеля от среднего веса клубня, количества клубней в кусте и содержания в них нитратов /

Fig. 12. Correlation matrix of potato crop yield dependence on average tuber weight, tuber quantity and nitrates content

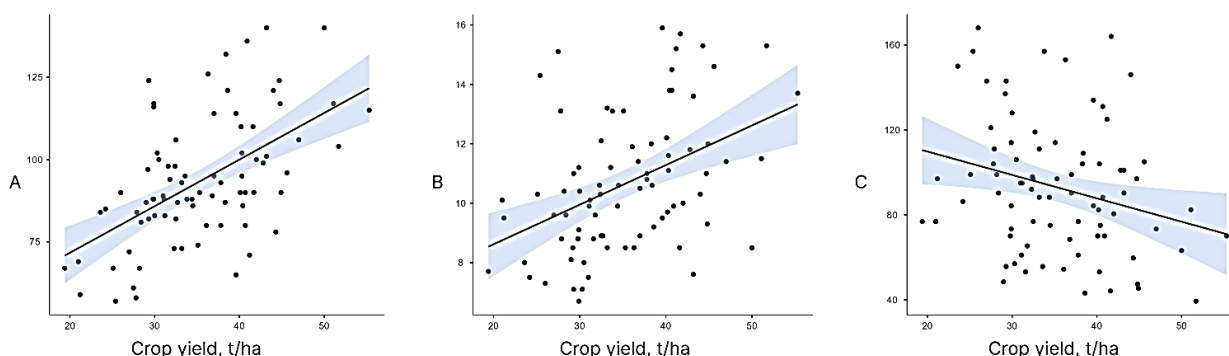


Рис. 13. Диаграммы рассеяния значений урожайности картофеля и среднего веса клубня (А), количества клубней (В), содержания нитратов в клубнях (С) /

Fig. 13. Scatterplots of values of potato yield and average tuber weight (A), tuber quantity (B), nitrates content (C)

Таблица 3 – Коэффициенты линейной регрессии в модели прогнозирования урожайности картофеля /
Table 3 – Linear regression coefficients in the potato yield forecasting model

Предиктор / Predictor	Вес (β) / Estimate (β)	SE	95%-ный ДИ веса / 95% CI of estimate		t	p-value
			нижний / lower	верхний / upper		
Константа / Intercept	-3,236	6,3870	-15,914	9,442	-0,507	0,614
Средний вес товарного клубня / Average tuber weight	0,314	0,0213	0,272	0,356	14,770	<0,001
Количество клубней в кусте / Tuber quantity	2,386	0,1708	2,047	2,725	13,967	<0,001
Ln (нитраты) / Natural log of nitrates	-3,630	1,1426	-5,898	-1,362	-3,177	0,002

Примечания: SE – стандартная ошибка; t – значение t-статистики; ДИ – доверительный интервал /
Notes: SE – standard error, t – t-value of test; CI – confidence interval

Иначе говоря, модель линейной регрессии можно представить следующим образом:

$$Y = 0.314 * M + 2.386 * N - 3.63 * \ln(\text{NO}_3) - 3.236,$$

где Y – урожайность, т/га; M – масса среднего товарного клубня, г; N – количество клубней, шт.; NO_3 – нитраты, мг/кг.

Модель показывает, что при увеличении среднего веса товарного клубня на 1 г урожайность возрастает на 0,314 т/га, при увеличении количества клубней в кусте на 1 шт. урожайность возрастает на 2,386 т/га, а при увеличении концентрации нитратов на каждую единицу натурального логарифма урожайность снижается на 3,63 т/га.

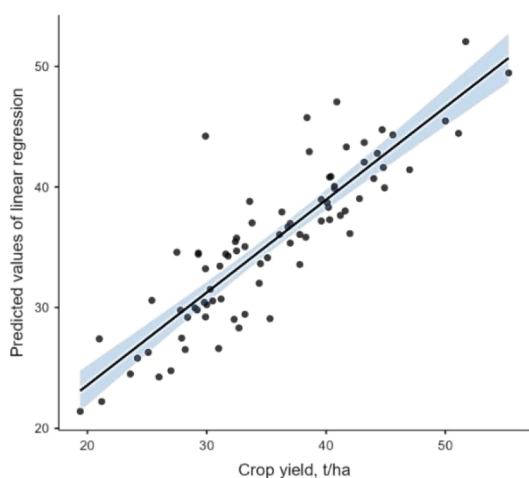


Рис. 14. Диаграмма рассеяния фактических значений урожайности (по оси абсцисс) и прогнозируемых значений линейной регрессии (по оси ординат) /

Fig. 14. Scatterplot of actual values of crop yield (X-axis) and predicted values of linear regression (Y-axis)

Проверку на нормальность остатков провели с помощью построения графика QQ-plot и расчёта статистики Колмогорова-Смирнова. Из рисунка 16 видно, что остатки на большом участке протяженности QQ-plot укладываются в прямую линию, за исключением самых низких и самых высоких значений, а гистограмма имеет куполообразную форму. Статистика Колмогорова-Смирнова также указывает на допущение о нормальном распределении остатков ($K\text{-Stest} = 0,0933$, $p = 0,348$).

Для диагностики коллинеарности рассчитан показатель VIF (табл. 4). Наши предикторы не превышают порогового значения VIF (2,0), что свидетельствует об отсутствии коллинеарности среди предикторов.

Для расчёта ошибки аппроксимации использована формула средней абсолютной ошибки в процентах MAPE. Для нашей модели

Полученная модель способна значимо предсказать 77,9 % дисперсии данных ($R^2 = 0,785$, $R^2_{\text{adj}} = 0,779$, $p < 0,001$).

Для проверки линейности и гомогенности полученной модели построили диаграмму рассеяния фактических значений урожайности и прогнозируемых значений линейной регрессии (рис. 14), которая свидетельствует о наличии линейной взаимосвязи. Дополнительно построили диаграмму рассеяния прогнозируемых значений линейной регрессии и остатков линейной регрессии (разница между предсказываемым значением и реальным) (рис. 15), показывающую, что линия регрессии близка к прямой – «средний остаток» для всех прогнозируемых значений более или менее одинаков (гомогенен).

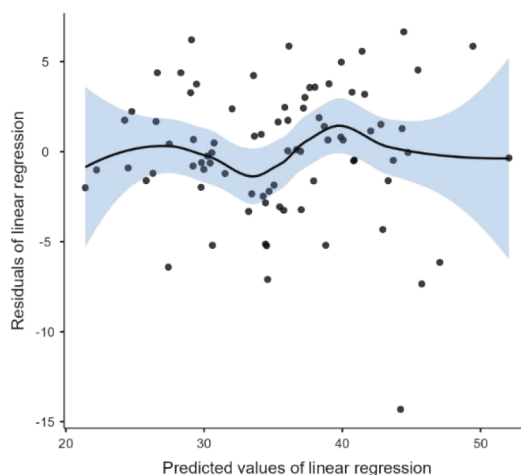


Рис. 15. Диаграмма рассеяния прогнозируемых значений линейной регрессии (по оси абсцисс) и остатков линейной регрессии (по оси ординат) /

Fig. 15. Scatterplot of predicted values of linear regression (X-axis) and residuals of linear regression (Y-axis)

линейной регрессии значение MAPE составило 8,55 %, что говорит о том, что степень соответствия модели идеальна, поскольку $\text{MAPE} < 10\%$. Эти результаты доказывают хорошую подгонку модели к реальному набору данных и, следовательно, открывают широкие возможности для применения.

Таблица 4 – Проверка коллинеарности /
Table 4 – Collinearity check

Предиктор / Predictor	VIF*
Средний вес товарного клубня / Average tuber weight	1,17
Количество клубней / Tuber quantity	1,12
Ln (нитраты) / Natural log of nitrates	1,05

* VIF – коэффициент увеличения дисперсии /
Variance inflation factor

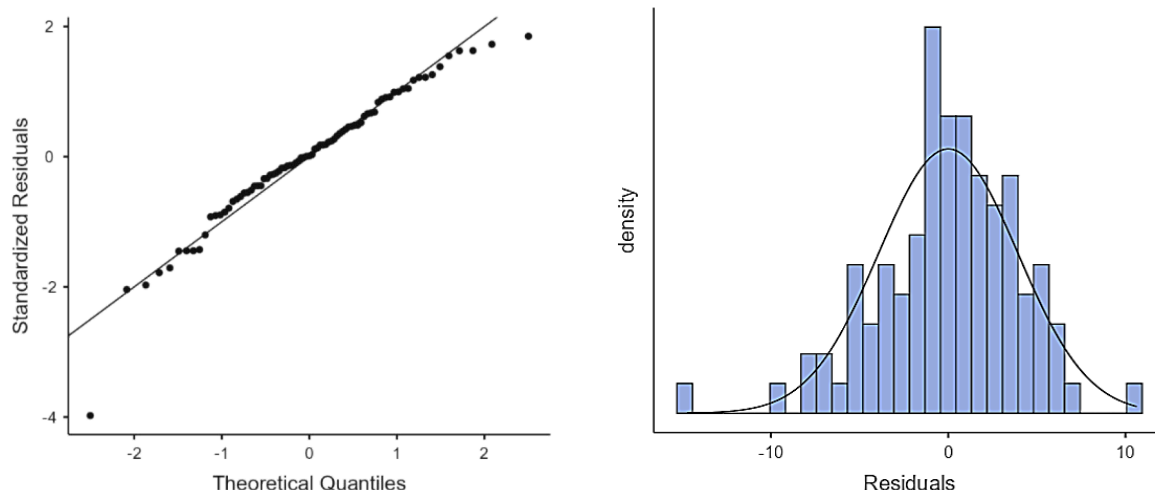


Рис. 16. График QQ-plot и гистограмма для стандартизованных остатков регрессии /
Fig. 16. QQ-plot and histogram plots for standardized regression residuals

В настоящем исследовании была разработана надежная модель множественной линейной регрессии, которая объясняет значительную часть дисперсии урожайности картофеля ($R^2 = 0,785$; $R^2_{adj} = 0,779$, $p < 0,001$) для сотни различных сортов в установленных агроклиматических условиях. Предикторы модели (средний вес клубня, количество клубней на растение и натуральный логарифм концентрации нитратов в клубне) дают важную информацию о физиологических факторах, определяющих урожайность и эффективность использования питательных веществ (табл. 3). Положительные и статистически достоверные коэффициенты для среднего веса клубня ($\beta = 0,314$; $p < 0,001$) и количества клубней на растение ($\beta = 2,386$; $p < 0,001$) относятся к хорошо известным и фундаментальным компонентам урожайности [10]. При этом, анализируя вес (β) двух предикторов среди сотни генетически разнообразных сортов, становится понятным, что количество клубней на растение – более мощный компонент урожайности, чем рост самих клубней (средний вес клубня). Это факт указывает на то, что агрономические усилия предприятий должны быть направлены на оптимизацию мероприятий по закладке клубней, что может дать большую отдачу урожая по сравнению с теми методами, которые сосредоточены исключительно на увеличении размера отдельных клубней.

Установленная в настоящем исследовании сильная отрицательная зависимость между натуральным логарифмом концентрации нитратов в клубнях при сборе урожая и общей урожайностью картофеля ($\beta = -3,630$; $p = 0,002$)

указывает на то, что после учета основных компонентов урожайности сорта с более высоким содержанием нитратов в клубнях имели более низкую урожайность. Высокая концентрация нитратов в органах хранения питательных веществ в конце вегетационного периода является хорошо известным показателем метаболической неэффективности [11]. Она представляет собой запас азота, который был поглощён растением, но не был эффективно ассимилирован в белки и нуклеиновые кислоты, необходимые для накопления биомассы и формирования урожая [11, 12]. Это исследование согласуется с принципом, согласно которому избыточное азотистое питание в конце сезона приводит к уменьшению урожая и росту зеленой части растения [13]. Наша модель эффективно количественно оценивает урожайность, связанную с этой неэффективной утилизацией азота.

Сложившиеся агроклиматические условия в текущем исследовании – плодородная почва с высоким содержанием гумуса (4,86 %) и легкогидролизуемого азота (130 мг/кг) с добавлением удобрения $N_{80}K_{70}P_{70}$ – вероятно, создали среду с достаточным содержанием азота, а для некоторых генотипов – избыточным. Эти условия идеально подходят для выявления генетических различий в эффективности использования азота (Nitrogen Use Efficiency) [14]. Результаты исследования дополняют работу К. Френковяк с соавт. (K. Frąckowiak et al.), которые продемонстрировали, что питательный статус растений в начале клубнеобразования является мощным инструментом для прогнозирования потенциальной урожайности [15]. Наши результаты представляют окончательную оценку после

сбора урожая, предполагая, что даже если азота в начале сезона достаточно для создания высокого потенциала продуктивности, этот потенциал не реализуется у сортов, которые не могут эффективно перерабатывать доступный азот в течение всего периода набора массы. Кроме того, текущая модель подтверждает обоснованность использования инструментов мониторинга азота в течение сезона, предложенных в работе [16]. Отсутствие в полученной модели данных крахмала, протеина, сахаров и витамина С не говорит об отсутствии их роли в формировании урожайности, а указывает на сложные нелинейные взаимосвязи, которые невозможно отразить в модели линейной регрессии. Более сложные методы исследования (машинное обучение, деревья решений, ансамблевые методы или нейронные сети) могут разрешить эти противоречия.

Ограничения исследования. Критическим фактором, который не был включен в регрессионную модель, может быть распределение сортов по различным классам зрелости. Он неразрывно связан с продолжительностью вегетационного периода и, следовательно, как с потенциальной урожайностью, так и с общей потребностью в азоте. Вероятно, что равномерное внесение азота было сверхоптимальным для многих раннеспелых сортов, которые имеют более короткий вегетационный период и более низкую общую потребность в азоте, что располагает их к накоплению более высоких остаточных концентраций нитратов. Так, отрицательная корреляция между нитратами и урожайностью, наблюдаемая в настоящей модели,

может быть особенно выраженной в когорте раннеспелых сортов. Помимо этого, коэффициенты модели относятся к условиям испытания в течение одного года на одном участке и требуют проверки в различных условиях. В будущих исследованиях следует уделить приоритетное внимание повторному анализу этого набора данных с использованием анализа ковариации (ANCOVA) с включением класса зрелости в качестве категориального фактора. Такой анализ позволит количественно оценить влияние класса зрелости на урожайность и, что более важно, определить существует ли значимое взаимодействие между классом зрелости и другими предикторами, в частности концентрацией нитратов в клубнях.

Заключение. Исследование демонстрирует, что урожайность картофеля, в основном, зависит от количества и размера клубней. Концентрация нитратов в клубнях при сборе урожая служит надежным отрицательным предиктором и мощным показателем эффективности использования азота. Полученная модель предоставляет новый мощный критерий отбора сортов: помимо выявления сортов с большим количеством и весом клубней селекционные программы должны активно отбирать генотипы, демонстрирующие низкое содержание остаточных нитратов в клубнях при стандартном режиме азотного питания. Это обеспечивает прямой, количественно измеримый показатель превосходной эффективности использования азота, который является ключевой характеристикой для разработки более продуктивных и экологически устойчивых сортов.

Список литературы

1. Devaux A., Goffart J. P., Kromann P., Andrade-Piedra J., Polar V., Hareau G. The potato of the future: opportunities and challenges in sustainable agri-food systems. *Potato Research*. 2021;64(4):681–720. DOI: <https://doi.org/10.1007/s11540-021-09501-4>
2. Амирова Е. Ф., Гаврильева Н. К., Григорьев А. В., Соргутов И. В. Цифровизация в сельском хозяйстве: проблемы внедрения. *Siberian Journal of Life Sciences and Agriculture*. 2021;13(6):144–155. DOI: <https://doi.org/10.12731/2658-6649-2021-13-6-144-155> EDN: KZNVKC
3. Камышенко Г. А. Климатическая составляющая в расчете урожайности картофеля. Природные ресурсы. 2021;(2):14–21. Режим доступа: <https://elibrary.ru/item.asp?id=47503785> EDN: ABQQIV
4. Kuradusenge M., Hitimana E., Hanyurwimfura D., Rukundo P., Mtonga K., Mukasine A. et al. Crop yield prediction using machine learning models: case of irish potato and maize. *Agriculture*. 2023;13(1):225. DOI: <https://doi.org/10.3390/agriculture13010225>
5. Piekutowska M., Niedbała G., Piskier T., Lenartowicz T., Pilarski K., Wojciechowski T. et al. The application of multiple linear regression and artificial neural network models for yield prediction of very early potato cultivars before harvest. *Agronomy*. 2021;11(5):885. DOI: <https://doi.org/10.3390/agronomy11050885>
6. Воробьев Н. И., Лысов А. К., Корнилов Т. В., Хютти А. В. Вычислительная нейросеть для обработки светоотражательных спектров растений и дистанционного фитосанитарного мониторинга картофеля. *Аграрная наука Евро-Северо-Востока*. 2024;25(2):283–292. DOI: <https://doi.org/10.30766/2072-9081.2024.25.2.283-292> EDN: KUSISD

7. Шанина Е. П., Клюкина Е. М., Беляева Н. В., Стафеева М. А., Келик Л. А., Ахметханов В. Ф. Испытание сортов картофеля различного целевого использования. Картофель и овощи. 2024;(1):39–43. DOI: <https://doi.org/10.25630/PAV.2024.92.35.003> EDN: VKRFOX
8. Шанина Е. П., Клюкина Е. М., Стафеева М. А., Беляева Н. В., Гончар О. Н. Сравнительный анализ сортов картофеля коллекционного питомника в зависимости от географического происхождения. Достижения науки и техники АПК. 2020;34(6):75–78. DOI: <https://doi.org/10.24411/0235-2451-2020-10614> EDN: FYDLHG
9. Cravero A., Pardo S., Sepúlveda S., Muñoz L. Challenges to use machine learning in agricultural big data: a systematic literature review. Agronomy. 2022;12(3):748. DOI: <https://doi.org/10.3390/agronomy12030748>
10. Лыскова И. В., Пермяков П. В., Кратюк Е. И. Результаты изучения коллекционных сортов и новых селекционных номеров картофеля по хозяйственно полезным признакам. Аграрная наука Евро-Северо-Востока. 2024;25(6):1019–1027. DOI: <https://doi.org/10.30766/2072-9081.2024.25.6.1019-1027> EDN: UUWSBI
11. Wszelaczyńska E., Pobereźny J., Keutgen A. J., Keutgen N., Gościńska K., Milczarek D. et al. Antinutritional nitrogen compounds content in potato (*Solanum tuberosum* L.) tubers depending on the genotype and production system. Agronomy. 2022;12(10):2415. DOI: <https://doi.org/10.3390/agronomy12102415>
12. Ding K., Shan Y., Wang L., Zhang Y., Tian G. Transcriptomics combined with physiological analysis and metabolomics revealed the response of potato tuber formation to nitrogen. BMC Plant Biology. 2024;24(1):1109. DOI: <https://doi.org/10.1186/s12870-024-05758-2>
13. Muleta N. D., Aga M. C. Role of nitrogen on potato production: a review. Journal of Plant Sciences. 2019;7(2):36–42. DOI: <https://doi.org/10.11648/j.jps.20190702.11>
14. Zebarth B. J., Tai G., Tarn R., de Jong H., Milburn P. H. Nitrogen use efficiency characteristics of commercial potato cultivars. Canadian Journal of Plant Science. 2004;84(2):589–598. DOI: <https://doi.org/10.4141/P03-050>
15. Frąckowiak K., Potarzycki J., Grzebisz W., Szczepaniak W. Potato nutritional status at the onset of tuberisation – a yield prediction tool. Plant, Soil and Environment. 2020;66(2):86–92. DOI: <https://doi.org/10.17221/533/2019-PSE>
16. do Carmo Milagres C., Fontes P. C. R., da Silveira M. V., Moreira M. A., de Carvalho Lopes I. P. Índices de nitrogênio e modelo para prognosticar a produção de tubérculos de batata. Revista Ceres. 2018;(65):261–270. DOI: <https://doi.org/10.1590/0034-737X201865030006>

References

1. Devaux A., Goffart J. P., Kromann P., Andrade-Piedra J., Polar V., Hareau G. The potato of the future: opportunities and challenges in sustainable agri-food systems. Potato Research. 2021;64(4):681–720. DOI: <https://doi.org/10.1007/s11540-021-09501-4>
2. Amirova E. F., Gavril'yeva N. K., Grigor'ev A. V., Sorgutov I. V. Digitalization in agriculture: problems of implementation. Siberian Journal of Life Sciences and Agriculture. 2021;13(6):144–155. (In Russ.). DOI: <https://doi.org/10.12731/2658-6649-2021-13-6-144-155>
3. Kamishenko G. A. Climatic component in the calculation potato yields. *Prirodnye resursi* = Natural Resources. 2021;(2):14–21. (In Belarus). URL: <https://elibrary.ru/item.asp?id=47503785>
4. Kuradusenge M., Hitimana E., Hanyurwimfura D., Rukundo P., Mtonga K., Mukasine A. et al. Crop yield prediction using machine learning models: case of Irish potato and maize. Agriculture. 2023;13(1):225. DOI: <https://doi.org/10.3390/agriculture13010225>
5. Piekutowska M., Niedbała G., Piskier T., Lenartowicz T., Pilarski K., Wojciechowski T. et al. The application of multiple linear regression and artificial neural network models for yield prediction of very early potato cultivars before harvest. Agronomy. 2021;11(5):885. DOI: <https://doi.org/10.3390/agronomy11050885>
6. Vorobyov N. I., Lysov A. K., Kornilov T. V., Hyutti A. V. Computational neural network for processing light-reflective spectra of plants and remote phytosanitary monitoring of potatoes. *Agrarnaya nauka Evro-Severo-Vostoka* = Agricultural Science Euro-North-East. 2024;25(2):283–292. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2024.25.2.283-292>
7. Shanina E. P., Klyukina E. M., Belyaeva N. V., Stafeyeva M. A., Kelik L. A., Akhmetkhanov V. F. Testing of potato varieties of different intended uses. *Kartofel' i ovoshchi* = Potato and Vegetables. 2024;(1):39–43. (In Russ.). DOI: <https://doi.org/10.25630/PAV.2024.92.35.003>
8. Shanina E. P., Klyukina E. M., Stafeyeva M. A., Belyaeva N. V., Gonchar O. N. The comparative geographical analysis of potato varieties from a collection nursery. *Dostizheniya nauki i tekhniki APK* = Achievements of Science and Technology of AICis. 2020;34(6):75–78. (In Russ.). DOI: <https://doi.org/10.24411/0235-2451-2020-10614>
9. Cravero A., Pardo S., Sepúlveda S., Muñoz L. Challenges to use machine learning in agricultural big data: a systematic literature review. Agronomy. 2022;12(3):748. DOI: <https://doi.org/10.3390/agronomy12030748>
10. Lyskova I. V., Permyakov P. V., Kratyuk E. I. The results of study of collection cultivars and potato new breeding numbers according to agronomic traits. *Agrarnaya nauka Evro-Severo-Vostoka* = Agricultural Science Euro-North-East. 2024;25(6):1019–1027. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2024.25.6.1019-1027>

11. Wszelaczyńska E., Pobereżny J., Keutgen A. J., Keutgen N., Gościńska K., Milczarek D. et al. Antinutritional nitrogen compounds content in potato (*Solanum tuberosum* L.) tubers depending on the genotype and production system. *Agronomy*. 2022;12(10):2415. DOI: <https://doi.org/10.3390/agronomy12102415>
12. Ding K., Shan Y., Wang L., Zhang Y., Tian G. Transcriptomics combined with physiological analysis and metabolomics revealed the response of potato tuber formation to nitrogen. *BMC Plant Biology*. 2024;24(1):1109. DOI: <https://doi.org/10.1186/s12870-024-05758-2>
13. Muleta N. D., Aga M. C. Role of nitrogen on potato production: a review. *Journal of Plant Sciences*. 2019;7(2):36–42. DOI: <https://doi.org/10.11648/j.jps.20190702.11>
14. Zebbarth B. J., Tai G., Tarn R., de Jong H., Milburn P. H. Nitrogen use efficiency characteristics of commercial potato cultivars. *Canadian Journal of Plant Science*. 2004;84(2):589–598. DOI: <https://doi.org/10.4141/P03-050>
15. Frąckowiak K., Potarzycki J., Grzebisz W., Szczepaniak W. Potato nutritional status at the onset of tuberisation – a yield prediction tool. *Plant, Soil and Environment*. 2020;66(2):86–92. DOI: <https://doi.org/10.17221/533/2019-PSE>
16. do Carmo Milagres C., Fontes P. C. R., da Silveira M. V., Moreira M. A., de Carvalho Lopes I. P. Índices de nitrogênio e modelo para prognosticar a produção de tubérculos de batata. *Revista Ceres*. 2018;(65):261–270. DOI: <https://doi.org/10.1590/0034-737X201865030006>

Сведения об авторах

Шанина Елена Петровна, доктор с.-х. наук, доцент, главный научный сотрудник, Уральский научно-исследовательский институт сельского хозяйства – филиал ФГБНУ «Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук», ул. Главная, д. 21, пос. Исток, г. Екатеринбург, Российская Федерация, 620061, e-mail: uralniishoz@list.ru,
ORCID: <https://orcid.org/0009-0000-5818-3813>

✉ **Оберюхтин Денис Андреевич**, младший научный сотрудник, Уральский научно-исследовательский институт сельского хозяйства – филиал ФГБНУ «Уральский федеральный аграрный научно-исследовательский центр Уральского отделения Российской академии наук», ул. Главная, д.21, пос. Исток, г. Екатеринбург, Российская Федерация, 620061, e-mail: uralniishoz@list.ru, ORCID: <https://orcid.org/0009-0006-1497-5740>, e-mail: oberuhtindenis@gmail.com

Information about the authors

Elena P. Shanina, DSc in Agriculture, associate professor, chief researcher, Ural Research Institute of Agriculture – branch of the Ural Federal Agrarian Scientific Research Centre, Ural Branch of the Russian Academy of Sciences, Glavnaya Str., 21, village Istok, Yekaterinburg, Russian Federation, 620061, e-mail: uralniishoz@list.ru,
ORCID: <https://orcid.org/0009-0000-5818-3813>

✉ **Denis A. Oberiukhtin**, junior researcher, Ural Research Institute of Agriculture – branch of the Ural Federal Agrarian Scientific Research Centre, Ural Branch of the Russian Academy of Sciences, Glavnaya Str., 21, village Istok, Yekaterinburg, Russian Federation, 620061, e-mail: uralniishoz@list.ru,
ORCID: <https://orcid.org/0009-0006-1497-5740>, e-mail: oberuhtindenis@gmail.com

✉ – Для контактов / Corresponding author