

## КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ВАЛОВОГО СБОРА КАРТОФЕЛЯ НА ПРИМЕРЕ ОРЛОВСКОЙ ОБЛАСТИ

**Филимонова Ирина Алексеевна**

студент, Орловского Государственного аграрного университета им. Н. В. Парахина, РФ. г. Орел

**Яковлева Наталья Альбертовна**

канд. экон. наук, доцент, Орловский Государственный аграрный университет им. Н. В. Парахина, РФ, г. Орел

## CORRELATION ANALYSIS OF GROSS POTATO HARVEST ON THE EXAMPLE OF OREL REGION

*Irina Filimonova*

*Student, Orel state agrarian University. N.V. Parakhina, Russia, Orel*

*Natalia Yakovleva*

*Candidate of economic Sciences, associate Professor, Orel state agrarian University. N.V. Parakhina, Russia, Orel*

**Аннотация.** В данной статье мы проводим корреляционный анализ валового сбора картофеля с целью рассмотреть силу влияния посевной площади на валовой сбор картофеля.

**Abstract.** In this article, we carry out a correlation analysis of the gross potato harvest in order to consider the effect of the acreage on the gross potato harvest.

**Ключевые слова:** корреляционный анализ, валовой сбор, картофель, посевная площадь, регрессии, коэффициент эластичности, коэффициент детерминации, факторный признак, t-критерий Стьюдента.

**Keywords:** correlation analysis, gross harvest, potatoes, acreage, regression, elasticity coefficient, determination coefficient, factor feature, Student's t-criterion.

Корреляция – статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом изменения одной или нескольких из этих величин приводят к систематическому изменению другой или других величин [3]. По направлению корреляционная связь может

быть положительной ("прямой") и отрицательной ("обратной"). При положительной прямолинейной корреляции более высоким значениям одного признака соответствуют более высокие значения другого, а более низким значениям одного признака - низкие значения другого. При отрицательной корреляции соотношения обратные. При положительной корреляции коэффициент корреляции имеет положительный знак, например  $r=+0,207$ , при отрицательной корреляции - отрицательный знак, например  $r=-0,207$  [1].

Максимальное возможное абсолютное значение коэффициента корреляции  $r=1,00$ ; минимальное  $r=0,00$ .

Общая классификация корреляционных связей

- сильная, или тесная при коэффициенте корреляции  $r > 0,70$ ;
- средняя при  $0,50 < r < 0,69$ ;
- умеренная при  $0,30 < r < 0,49$ ;
- слабая при  $0,20 < r < 0,29$ ;
- очень слабая при  $r < 0,19$  [2].

Произвести расчет по итоговым значениям можно по формуле:

$$r = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2) \cdot (n \sum y^2 - (\sum y)^2)}}$$

(1)

С помощью корреляционного анализа определим влияние площади посадок картофеля, на валовый сбор по 9 годам. Для этого изобразим на графике зависимость валового сбора, от площади посадок, построим линейное уравнение регрессии, рассчитаем коэффициент корреляции и оценим его достоверность с помощью t-критерия Стьюдента.

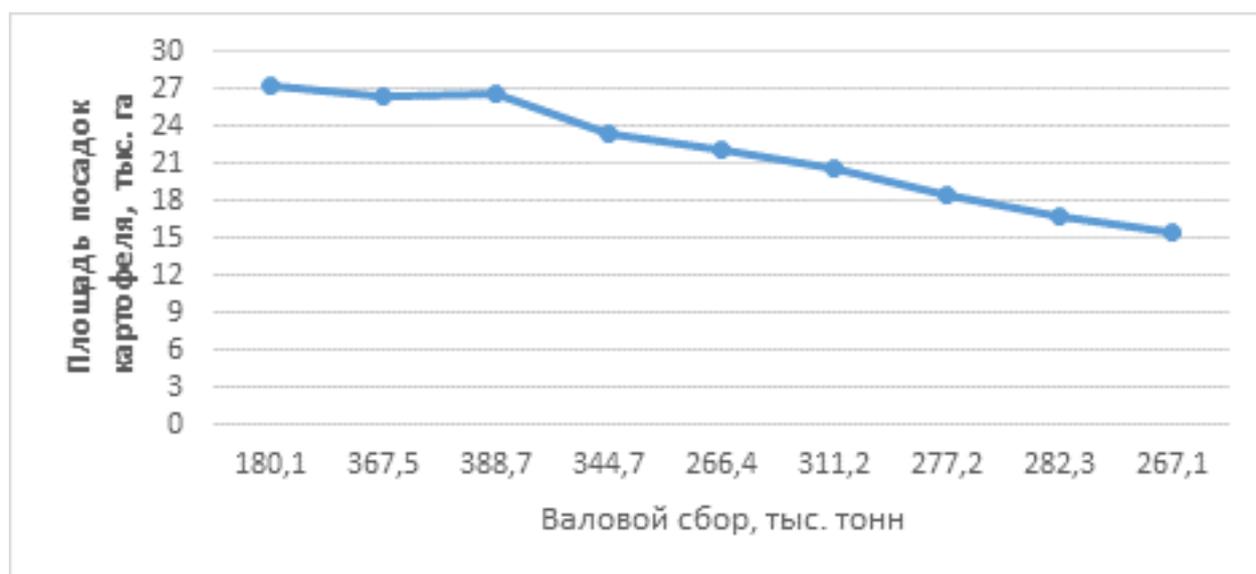
**Таблица 1.**

**Исходные данные**

Год	Валовой сбор, тыс. тонн. (x)	Площадь посадок, тыс. га. (y)
2010	180.1	27.2
2011	367.5	26.4
2012	388.7	26.6
2013	344.7	23.4
2014	266.4	22.1
2015	311.2	20.7

2016	277.2	18.5
2017	282.3	16.8
2018	267.1	15.5
Итого	2685.2	197.2

Построим график зависимости валового сбора картофеля от площади посадок картофеля (рис. 1).



**Рисунок 1. Зависимость валового сбора от площади посадок картофеля**

График показывает, что при уменьшении площади посадок картофеля, значение валового сбора картофеля в среднем понижается. В данном случае имеется линейная зависимость, которая может быть отражена уравнением регрессии:

$$\hat{y}_x = a + bx,$$

где  $\hat{y}_x$  – теоретическое значение результативного признака;

a – параметр уравнения, который не имеет экономического смысла;

b – коэффициент регрессии [5];

x – факторный признак.

**Таблица 2.**

### Данные для проведения корреляционного анализа

Год	x	y	xy	y <sup>2</sup>	x <sup>2</sup>	$\bar{y}_x$	y <sub>i</sub>	$\bar{y}_i$
2010	180.1	27.2	4898.72	739.84	32436.01	20.7	6.5	42.25
2011	367.5	26.4	9703	696.96	135056.25	24.5	1.9	3.61
2012	388.7	26.6	10339.42	707.56	151087.69	24.9	1.7	2.89
2013	344.7	23.4	8065.98	547.56	118818.09	24.0	-0.6	0.4
2014	266.4	22.1	5887.44	488.41	70968.96	22.5	-0.4	0.16
2015	311.2	20.7	6441.84	428.49	96845.44	23.4	-2.7	7.29
2016	277.2	18.5	5128.2	342.25	76839.84	22.7	-4.2	17.64
2017	282.3	16.8	4742.64	282.24	79693.29	22.8	-6	36
2018	267.1	15.5	4140.05	240.25	71342.41	22.5	-7	49
Итого	2685.2	197.2	59346.29	4473.56	833087.98	208	-	15

Для определения параметров а и b воспользуемся расчетными данными из таблицы 2 и решим систему двух уравнений, которая называется в статистике системой нормальных уравнений:

$$\sum y = an + b \sum x \quad (3)$$

$$\sum xy = a \sum x + b \sum x^2$$

$$197.2 = 9a + 2685.2b$$

$$59346.29 = 2685.2a + 833087.98b$$

Решив систему уравнений получим:

$$a = 17.14$$

$$b = 0.02$$

Таким образом, уравнение будет иметь вид:

$$\hat{y}_x = 17.14 + 0.02x$$

Коэффициент  $b > 0$ , это подтверждает, что зависимость между посевными площадями картофеля и валовым сбором картофеля прямая. При увеличении площади на 1 га, валовой сбор картофеля в среднем увеличится на 0.02 ц/га.

Воспользовавшись формулой (57)  $r = 0.24$ . Коэффициент корреляции показал, что между изучаемыми факторами существует прямая зависимость. Причем, связь между факторами слабая.

Рассчитаем коэффициент эластичности:

$$\varepsilon = b * \frac{\bar{x}}{\bar{y}} \quad (4)$$

$$\varepsilon = 0,02 * 298.4 / 21.9 = 0,27 \%$$

Коэффициент эластичности показывает на сколько процентов в среднем изменяется результативный признак при изменении факторного на 1 %. С изменением уровня посевных площадей на 1 % валовой сбор картофеля в среднем увеличится на 0,27 %.

Рассчитаем коэффициент детерминации по формуле:

$$D = r^2 * 100\% \quad (5)$$

$$D = 0,24^2 * 100\% = 6\%$$

Это означает, что зависимость очень слабая, в 6% случаев изменения посевной площади картофеля приведет к изменению валового сбора картофеля, остальные 94% вариации вызваны неучтенными в модели факторами.

Случайные ошибки параметром линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2 \div (n - 2)}{(x_i - \bar{x})^2}}; \quad m_a = \sqrt{\frac{(y_i - \hat{y}_i)^2}{(n - 2)} * \frac{\sum x^2}{n * \sum(x_i - \bar{x})^2}}; \quad m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Подставив соответствующие данные в формулы 6-8 получим:

$$m_b=0.027; m_a=8.114=8.114; m_r=0.367$$

Рассчитаем t-критерий Стьюдента для каждого показателя (a, b, r), используя соответствующие формулы:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r}{m_r}$$

$$t_b=0.74; t_a=2.112; t_r=0.65$$

По специальной таблице установим табличное значение t-критерия Стьюдента. В нашем случае  $t_{\text{табл}}=2.3646$  при уровне значимости 0.05 и числе степеней свободы  $n-2=9-2=7$ .

Итак  $t_b < t_{\text{табл}}$ ;

$t_a < t_{\text{табл}}$ ;  $t_r < t_{\text{табл}}$

Следовательно, нулевая гипотеза о не существовании коэффициента корреляции и параметров регрессии отвергается т.е.  $a$ ,  $b$  и  $r$  статистически незначимы.

Так как уравнение регрессии является статистически незначимым, расчет прогнозируемого значения валового сбора 1 т картофеля по данному уравнению не даст достоверного результата.

### Список литературы:

1. Гончарова Н.М., Анализ государственного регулирования экономического развития аграрного сектора экономики. [Текст]: Экономика и социум: сб. статей – Орел 2014. – с. 5.
2. Сидняев Н. И. Теория вероятностей и математическая статистика: учебник / Н. И. Сидняев. – М.: Юрайт, 2011. – с. 219 – (Бакалавр) – для бакалавров.
3. Яковлев В. Б. Статистика. Расчеты в Microsoft Excel: учеб. пособие / В. Б. Яковлев. – М.: КолосС, 2005. – с. 325.