Statistical analysis of vegetable productivity dynamics of Uzbekistan

Akhtamjon Fayziev[1,*], *Kudrat* Ruzmetov[1], *Nurilla* Noraliev[1] and *Tolibjon* Turgunov[1]

[1]Tashkent State Agrarian University, 2, University street, Tashkent, 100140, Uzbekistan

Abstract. In all fields, there are a lot of random events at a given time. In particular, the process of growing agricultural crops, which is repeated over a certain period, that is, seasonally, is the basis for our analysis as a discrete $\{Y_t, t \in T\}$ random dynamic series. This research contributes to the existing literature on agricultural productivity by focusing on the unique context of Uzbekistan. By utilizing a rigorous statistical approach and considering multiple influencing factors, we aim to provide evidence-based recommendations to policymakers and stakeholders for enhancing vegetable productivity and ensuring sustainable agricultural development. In the article, the average yield of $\{Y_t, t \in T\}$ vegetables grown in open and closed lands of the Republic of Uzbekistan in 2006-2020 was statistically analyzed using modern methods of mathematical statistics suitable for stable dynamic series, and point and interval statistical estimates were made for vegetable yield with a 95% guarantee. The trend part, which characterizes the main direction of vegetable cultivation, is determined, and the yield obtained from vegetables in the coming years is predicted. The fact that this process has an autocorrelation relationship was determined with the help of statistical criteria, and its important characteristics and laws were studied.
Keywords. Discrete, dynamic, statistical, trend, autocorrelation.

1 Introduction

Vegetable productivity is a crucial aspect of agricultural production, contributing to food security and economic growth in many countries [1-5]. Uzbekistan, located in Central Asia, has a rich agricultural heritage and is known for its diverse and abundant vegetable production. However, understanding the dynamics of vegetable productivity and its associated factors is essential for sustainable agricultural development and effective policymaking [6, 7].

Statistical analysis plays a vital role in comprehending the complex relationships between various factors influencing vegetable productivity. By employing rigorous statistical methods, researchers can identify patterns, trends, and potential drivers of productivity changes, enabling policymakers and stakeholders to make informed decisions to enhance agricultural practices and optimize resource allocation [8, 9].

* Corresponding author: fayziev47@mail.ru

This manuscript presents a comprehensive statistical analysis of vegetable productivity dynamics in Uzbekistan. The study aims to investigate the trends, patterns, and key determinants of vegetable productivity over a specific time period. By examining historical data and applying advanced statistical techniques, we seek to identify the factors that have contributed to changes in vegetable productivity, both at the national and regional levels.

It is known that the main components of the time series $\{Y_t, t \in T\}$ can be: 1) a trend showing the main direction, 2) a part that oscillates around the trend, 3) a seasonal part, 4) a random part [10-15]. Moving average, finite differences, least squares, autocorrelation, statistical hypothesis testing of various dynamic series, and other analysis methods are used to solve these problems [16, 17]. Because, in discrete dynamical series, the observation results can be correlated in general.

There are many scientific works and literature devoted to the study and analysis of dynamic series, including Anderson [1], Kendal [2], Tikhomirov [3], Sulaymanov [4], Fayziev [5] and others.

In this article, the average yield of vegetables grown in open and closed lands of the Republic of Uzbekistan in 2006-2020 was statistically analyzed as a stable discrete dynamic series, and guaranteed theoretical and practical conclusions were drawn.

2 Materials and methods

We statistically analyze the 15-year average yield of vegetables grown in Uzbekistan in 2006-2020 as a discrete stable dynamic series.

In order to determine the main direction (trend part) of the dynamics of the yield $\{Y_t, t \in T\}$ of vegetables grown in open and closed lands of Uzbekistan in 2006-2020, we geometrically interpret the collected statistical data in the Cartesian coordinate system.

The unknown parameters involved in the determined linear model are determined by solving the following system of normal equations by the method of least squares using statistical data:

$$\begin{cases} a_0 T + a_1 \sum t = \sum y_t \\ a_0 \sum t + a_1 \sum t^2 = \sum y_t t \end{cases} \quad (1)$$

The solution of equation (1) is as follows:

$$a_0 = \frac{1}{\mathrm{T}} \sum y_t, \quad a_1 = \frac{1}{\sum t^2} \sum y_t t. \quad (2)$$

We study the dynamics of the average yield of vegetables $\{Y_t, t \in T\}$ grown in the open and closed lands of Uzbekistan in 2006-2020, using the given statistical data, the unknown (2) parameters participating in the trend part, and the least squares method.

3 Results and discussion

In order to determine the main direction (trend part) of the dynamics of vegetable productivity $\{Y_t, t \in T\}$, we compile the following Table 1 according to statistical data:

Table 1. Main direction (trend part) of the dynamics of vegetable productivity.

#	**Observations, *t***	**Y_t, q/ha**	**t**	**t^2**	**y_tt**	**$y_t t^2$**
1	2006	223	-7	49	-1561	10927
2	2007	228.4	-6	36	-1370.4	8222.4
3	2008	246.4	-5	25	-1232	6160
4	2009	253.6	-4	16	-1014.4	4057.6
5	2010	252.5	-3	9	-757.5	2272.5
6	2011	263.7	-2	4	-527.4	1054.8
7	2012	265.6	-1	1	-265.6	265.6
8	2013	270.9	0	0	0	0
9	2014	268.3	1	1	268.3	268.3
10	2015	271	2	2	542	542
11	2016	271.1	3	9	813.3	2439.9
12	2017	253.6	4	16	1014.4	4057.6
13	2018	237.1	5	25	1185.5	5927.5
14	2019	235.6	6	36	1413.6	8481.6
15	2020	234.4	7	49	1640.8	11485.6
	Σ	3775.2	0	278	149.6	66162.4

Using the calculations in the table, we determine the unknown parameters involved in the linear relationship (Equation 2) using Table 1:

$$\sum y_t = 3775{,}2\,,\ \ a_0 = \frac{1}{\mathrm{T}}\sum y_t = \frac{3775{,}2}{15} = 251{,}68,\quad a_1 = \frac{149{,}6}{278} = 0{,}54$$

Based on the statistical analysis of the vegetable yield grown in Uzbekistan in the years 2006-2020 as a discrete, stationary dynamic $\{Y_t, t \in T\}$ series, we have the following equation of the trend part, which represents the connection of its main direction with t-years:

$$y(t) = 0{,}54t + 251{,}68 \qquad (3)$$

In order to determine how well the selected trend (3) model was selected using statistical criteria, this hypothesis H_0 was checked in the $\alpha = 0{,}05$ alternative $H_0 :\ a_1 = 0$ condition with H_1 value level, and $H_1 :\ a_1 \neq 0$ was determined to be correct. So, with a 95% guarantee, the $y = a_0 + a_1 t$ linear model is correctly chosen.

By putting the value *t* = 3 into Equation 3, i.e., the trend model of vegetable cultivation dynamics, we determine with a 95% guarantee that the point statistical average value of vegetable productivity in Uzbekistan in the 2023 economic year will be approximately 253.30 q/ha.

The autocorrelation coefficient is important in studying dynamic series properties. To answer this question, we calculate the following quantities by the finite difference method, delaying the time by L years (L-lag) (Table 2):

$$\Delta Y_t = Y_{t+1} - Y_t\ ,\quad \Delta^2 Y_t = \Delta Y_{t+1} - \Delta Y_t,\quad \Delta^3 Y_t = \Delta^2 Y_{t+1} - \Delta^2 Y_t$$

Table 2. Calculation table of finite differences.

t	Y_t, q/ha	Y_t^2	ΔY_t	ΔY_t^2	$\Delta^2 Y_t$	$\Delta^2 Y_t^2$	$\Delta^3 Y_t$	$\Delta^3 Y_t^2$
2006	223	49729	-	-	-	-	-	-
2007	228.4	52166.56	5.4	29.16	-	-	-	-
2008	246.4	60712.96	18	324	23.4	547.56	-	-
2009	253.6	64312.96	7.2	51.84	25.2	635.04	1.8	3.24
2010	252.5	63756.25	-1.1	1.21	6.1	37.21	-19.1	364.81
2011	263.7	69537.69	11.2	125.44	10.1	102.01	4	16
2012	265.6	70543.36	1.9	3.61	13.1	171.61	3	9
2013	270.9	73386.81	5.3	28.09	7.2	51.84	-5.9	34.81
2014	268.3	71984.89	-2.6	6.76	2.7	7.29	-4.5	20.25
2015	271	73441	2.7	7.29	0.1	0.01	-2.6	6.76
2016	271.1	73495.21	0.1	0.01	2.8	7.84	2.7	7.29
2017	253.6	64312.96	-17.5	306.25	-17.4	302.76	-20.2	408.04
2018	237.1	56216.41	-16.5	272.25	-34	1156	-16.6	275.56
2019	235.6	55507.36	-1.5	2.25	-18	324	16	256
2020	234.4	54943.36	-1.2	1.44	-2.7	7.29	15.3	234.09
Σ	3775.2	954046.78	11.4	1159.6	18.6	3350.46	-26.1	1635.85

According to Table 2, when we calculate the values of the following quantities, we see that the values of the coefficients of variation of the finite differences are close to each other (Due to space limitations, we will not give full calculations, $V_1 \approx V_2 \approx V_3$).

$$V_k = \frac{\sum_{i=k}^{T} (\Delta^2 y_k)^2}{(T-k)C_{2k}^k} \quad k = 1,2,3$$

This confirmation also confirms that the trend part of the studied dynamic series has a linear relationship.

Using statistical data to calculate autocorrelation coefficients, we make the following Table 3:

Table 3. Calculation of autocorrelation coefficients

#	Y_t	$Y_t \cdot Y_{t+1}$	$Y_t \cdot Y_{t+2}$	$Y_t \cdot Y_{t+3}$	$Y_t \cdot Y_{t+4}$	$Y_t \cdot Y_{t+5}$
1	223	-	-	-	-	-
2	228.4	50933.2	-	-	-	-
3	246.4	56277.76	54947.2	-	-	-
4	253.6	62487.04	57922.24	56552.8	-	-
5	252.5	64034	62216	57671	56307.5	-
6	263.7	66584.25	66874.32	64975.68	60229.08	58805.1
7	265.6	70038.72	67064	67356.16	65443.84	60663.04
8	270.9	71951.04	71436.33	68402.25	68700.24	66749.76
9	268.3	72682.47	71260.48	70750.71	67745.75	68040.88
10	271	72709.3	73413.9	71977.6	71462.7	68427.5
11	271.1	73468.1	72736.13	73440.99	72004.16	71489.07
12	253.6	68750.96	68725.6	68040.88	68700.24	67356.16
13	237.1	60128.56	64277.81	64254.1	63613.93	64230.39
14	235.6	55860.76	59748.16	63871.16	63847.6	63211.48
15	234.4	55224.64	55576.24	59443.84	63545.84	63522.4
Σ	3775.2	901130.8	846198.41	786737.17	721600.88	652495.78

Based on Table 3, we calculate the autocorrelation coefficients with the following formula:

$$R_L = \frac{\sum_{i=1}^{N-L} Y_i Y_{i+L} - \frac{\sum_{i=1}^{N-L} Y_t \sum_{i=L+1}^{N} Y_t}{N-L}}{\sqrt{\left[\sum_{i=1}^{N-L} Y_t^2 - \frac{\left(\sum_{i=1}^{N-L} Y_t\right)^2}{N-L}\right]\left[\sum_{i=L+1}^{N} Y_t^2 - \frac{\left(\sum_{i=L+1}^{N} Y_t\right)^2}{N-L}\right]}} \quad (4)$$

Calculations show that all R_2 , ..., R_5 values (L=1,2,,3…) of R_L are non-zero. Based on this statement, it can be assumed that the studied dynamic series has an autocorrelation relationship:

$$y_t = \rho \cdot y_{t-1} + \varepsilon_t,$$

where, $\rho = Cov(y_t, y_{t+1}) = M[(y_t - \overline{y_t})(y_{t+1} - \overline{y_t})]$.

When we test this hypothesis with the Darbin-Watson criterion using the above calculations:

$$d_{kuz} = \sum_{t=1}^{T-1}(Y_{t+1} - Y_t)^2 / \sum_{t=1}^{T} Y_t^{\,2}. \quad (5)$$

$d_{kuz} = 0{,}05$- values are found in a special table, $d_{krit} = 1.08$ we see that $d_{kuz} < d_{krit}$ is smaller than the critical value. Therefore, with 95% confidence, average vegetable yield has an autocorrelated $y_t = \rho y_{t-1} + \varepsilon,\quad \rho = Cov(y_t, y_{t+1})$, relationship. That is, the yield of vegetables obtained in this agricultural year depends on the yields of the previous year and the next year.

The statistical hypothesis H_0 that the annual average yield of vegetables grown in Uzbekistan has a normal distribution:

$$H_0 : P(X < x) = F_{a,\sigma}(x)$$
$$H_1 : P(X < x) \neq F_{a,\sigma}(x) \quad (6)$$

H_1 in the alternative condition, as a result of testing with Jacques-Berra, Shapiro-Wilcoxon, Pearson and parametric criteria, it was found to have a normal distribution with a 95% guarantee (Table 4). In addition, we can see the correctness of this conclusion by checking with the help of the following parametric criterion:

$$|A_s| < 1{,}5\sigma_1 , \quad \left|E_k + \frac{\sigma}{T+1}\right| < 1{,}5\sigma_2 \quad (7)$$

Here, skewness, kurtosis, and corrected variance are calculated using the following formulas:

$$A_s = \frac{m_3}{S_T^3}, \quad E_R = \frac{m_4}{S_T^4} - 3, \quad S_T^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \overline{x_i})^2 n_i \quad (8)$$

On the basis of these confirmations, it is possible to construct an interval statistical estimate of the annual average yield of vegetables with a value level of $\alpha = 0{,}05$ using the following formula:

$$\overline{Y_{T+\iota}} - t(T-2;\alpha)\bar{\sigma}_y \leq a_0 + a_1(T+i) \leq \overline{Y_{T+1}} + t(T-2;\alpha)\bar{\sigma}_y \quad (9)$$

where, the value of $t(T-2;\alpha)$ is found from the Student distribution table.

In fact, the average yield $\overline{y(t)}$ of agricultural crops, especially vegetables, has a normal distribution based on the central limit theorem of the theory of probability, as the land is grown in the same conditions, with the same thickness, in large areas.

Based on the formula (9), the interval statistical estimate for Uzbekistan vegetables with a 95% guarantee is as follows:

(243.36; 260.00) q/ha

According to the statistical data, using the program "Excel" and and 7x, we have the following statistical estimates for the important selection characteristics of the average vegetable yield (Table 4).

Table 4. Important selection characteristics of the average vegetable yield.

Selection characteristics of vegetable productivity	Statistical evaluation of selection characteristics of vegetable productivity
Average productivity, $\bar{y}_T$ q/ha	251.68
Selective variance, D_T	278.89
Sample mean squared deviation, σ_T	16.70
Variation, v (%)	6.64%
Skewness, A_ς	-0.32

Kurtosis, $E_{K_{\varsigma}}$	-1.33
Selective mean error, m_y	$m_y = \frac{\sigma_y}{\sqrt{n}} = 3.87$
Error, m'_y	$m'_y = t\ m_y = 2.15 \cdot 3.87 = 8.32$
Mean standard deviation error, m_σ	$m_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{16.70}{5.48} = 3.05$
Interval statistical estimation (95%), $\bar{y}_T \pm t\ m_y$	$\bar{y}_T \pm t\ m_y = 251.68 \pm 8.32$ (243.36; 260.00) q/ha
Hypothesis, $H_0: P(\bar{y}_T < x) = F_{a,\sigma}(x)$	H_0 is accepted with a 95% confidence

4 Conclusions

The following conclusions can be drawn based on the statistical analysis of the yield of vegetables grown in open and closed fields in Uzbekistan in 2006-2020 as $\{Y_t, t \in T\}$, discrete dynamic series:

1) The yield of vegetables grown in open and closed fields in Uzbekistan is a complex random process that depends on the quality of seeds, soil fertility, agrotechnical measures, weather and many other factors;

2) The trend part $y(t) = 0{,}54t + 251{,}68$, which characterizes the main direction of the vegetable growing process, has a linear connection, from which there is a 95% guarantee that the vegetable yield in Uzbekistan will be 253.30 q/ha at t = 3, in the 2023 economic year;

3) Since the average yield of vegetables in Uzbekistan $\overline{y(t)}$ is a random variable with a normal distribution, the interval statistical estimate for it with a 95% guarantee is in the range of 243.36; 260.00 q/ha;

4) The productivity of vegetables grown in Uzbekistan has an autocorrelation relationship, that is, the productivity of this year 2023 depends on the productivity of previous years and the following years, $y_t = \rho y_{t-1} + \varepsilon_t$;

5) In general, the annual yield of vegetables grown in Uzbekistan is a non-stable dynamic series.

References

1. T. Anderson “Statistical analysis of time series”. - Moscow: "MIR" Press, 1996. - 759 p.
2. 2. M. Kendal, A. Stuart “Multivariate statistical analysis and time series. - Moscow: “Nauka” Press, 1996. -736 p.
3. N.P.Tikhomirov "Econometrics". - Moscow: "Ekzamen" Press, 2013. - 512 p.
4. B.A. Sulaymanov, A.A. Fayziev, J.N. Fayziev "Statistical analysis of experimental data". - Tashkent: TSAU Press, 2015. - 124 p.
5. A.A. Fayziev "Mathematical statistics", Study guide. "Ilm-ziyo-zakovat", p. 218, Tashkent - 2022.
6. B. Abdalimov, A.A. Fayziev "Mathematics", textbook. "Ilm-ziyo-zakovat", p. 518, Tashkent - 2022.
7. M.U.Achilov, A.A.Fayziev. “The analysis of dynamics of fruits and berry froductivity grown in Uzbekistan”, EPRA Internatsional journal of Research and Deve lopment (IJRD.Indiya).Volum: 4. Issue: 8. August 2019, Pp.5-9.
8. B. Abdalimov, A.A. Fayziev "Analysis of the economic indicators of the farm using mathematical modeling" // Journal "Uzbekistan Agrarian Science Bulletin". - Tashkent, 2019. - No. 2 (76), - p. 164–167.

9. A. A. Fayziev, O. Z. Karabashov, N. N. Musaeva "Forecasting the dynamics of the yield of cotton in the Andijan region." "Problems of science," Vestnik nauki i obrazovanie, Moscow. Journal N 8 (111). April 2021, part 2, 6-p.
10. V. Vahabov, A.A. Fayziev "Statistical analysis and forecasting of cotton yield dynamics Bukhara region", Tashkent state transport university. 1 st International Scientifik Conference "Modern Materials Science: Topical Issues, Achievements and Innovations (ISCMMSTIAL-2022)" (Tashkent, Mart 4-5, 2022). 5 p.
11. A.A.Fayziev, A.Turgunov, X.Mamadaliev, S. Nasridinov "Statistikal analysis and forecasting of potato yield dynamics in the republic of Uzbekistan". Tashkent university of information technologies named after muhammad al-khwarizmi. Icisct 2022. International conference on information science end communications technologies application sc, trends and opportunities. Tashkent, 28-30 September, 2022, 5 –p.
12. A ATukhtaboev, F Turaev, B A Khudayarov, E Esanov, and K Ruzmetov. Vibrations of a viscoelastic dam-plate of a hudro-technical structure under seismic load // IOP Conf. Series:Earth and Environmental Science 614(2020)012051. 2020, 1-7 https://doi.org/10.1088/1755-1315/614/1/012051
13. A.G. Sherov, B.A. Khudayarov, K.Sh. Ruzmetov and J. Aliyarov. Numerical investigation of the effects angles of attack on the flutter of a viscoelastic plate // Advances in Aircraft and Spacecraft Science 2020, 215-228 https://doi.org/10.12989/aas.2020.7.3.215
14. B. A. Khudayarov, K. Sh. Ruzmetov, F. Zh. Turaev, V.V. Vaxobov, M.A. Hidoyatova, S.S. Mirzaev and R. Abdikarimov. Numerical modeling of nonlinear vibrations of viscoelastic shallow shells // Engineering Solid Mechanics homepage 2020, 199-205 p. www.GrowingScience.com/esm
15. N Kurbonov and S Aminov. Computer modeling of filtration processes with piston extrusion // J. Phys.: Conf. Ser. 1441 012147. 2020. 1-9 p. https://iopscience.iop.org/article/10.1088/1742-6596/1441/1/012147
16. Gulomov, B.A. Khudayarov, K.S. Ruzmetov and F.Z. Turaev "Quadratic forms related to the voronoi's domain faces of the secondperfect form in seven variables" Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications & Algorithms 28 (2021) 15-23 Copyright c 2021 Watam Press
17. F.Z. Turaev O.K. Gulomov, B.A. Khudayarov and K.S. Ruzmetov "Quadratic Forms Related to the Voronoi'S Domain Faces of the Second Perfect Form in Seven Variables" Dynamics of Continuous, Discrete and Impulsive Systems Series B:Applications & Algorithms том 28 № (28)2021 p.15-21.