Statistical analysis and forecasting of cotton yield dynamics in Bukhara region

V. Vakhabov[1*], and *A. A.* Fayziev[2]

[1]"Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Tashkent, Uzbekistan
[2]Tashkent State Agrarian University, Tashkent, Uzbekistan

Abstract. Observations on some phenomena, the nature of which changes in time, are ordered sequences, which is called the time series. In the article, by the method of statistical analysis of time series, the statistical regularity of the series of dynamics of the average yield of cotton in the Bukhara region, the Republic of Uzbekistan (based on the materials of the CSO of the Republic of Uzbekistan for 2001-2019) was studied. Point and interval estimates for the average cotton yield were built with a 95% guarantee, explicit types of trends were determined, and yields in the region were predicted for subsequent years. With the help of Durbin-Watson statistical criteria, it was found that the average cotton yield in the region has an autocorrelation dependence.

1 Introduction

In almost every field there are phenomena that are important to study in their development and change over time. One can, for example, seek to predict the future on the basis of knowledge of the past, to control a process, to describe the characteristic features of a series on the basis of a limited amount of information. When processing time series, they rely largely on the developed mathematical statistical methods for distribution series. To date, statistics has a variety of methods for analyzing time series.

In general, the time series consists of four components: trend; fluctuations relative to the trend; seasonality effect; random component. The following works are devoted to the study and analysis of dynamic series: Anderson, Kendal, Brillinger, Chetyrkin, Vain, Sulaimanov $\{y_t,\ t \in T\}$ [1-7] other.

The study of the yield of agricultural processes, as a discrete dynamic series and forecasting their yield based on experimental data, play an important role in determining the economic efficiency of farming and dekhkan farms.

And so, in this work, the processing and analysis of the cotton yield for the observation period of 2001-2019 in Bukhara was carried out coy areas Uzbekistan, as a time series.

*Corresponding author: v.vaxobov2019@gmail.com

https://doi.org/10.1051/e3sconf/202340103051

2 Analysis

The geometric image of the observed data (table 1, column 3), the coordinate system give grounds in the first approximation to assume the hypothesis that the trend part of the process has a linear dependence (Fig. 1) of the form$y(t) = a_1 t + a_0$ where unknown parameters are determined by the least squares method based on experimental data, solving the system of normal equations:

$$\begin{cases} a_0 T + a_1 \sum t = \sum y_t \\ a_0 \sum t + a_1 \sum t^2 = \sum y_1 t \end{cases} (1)$$

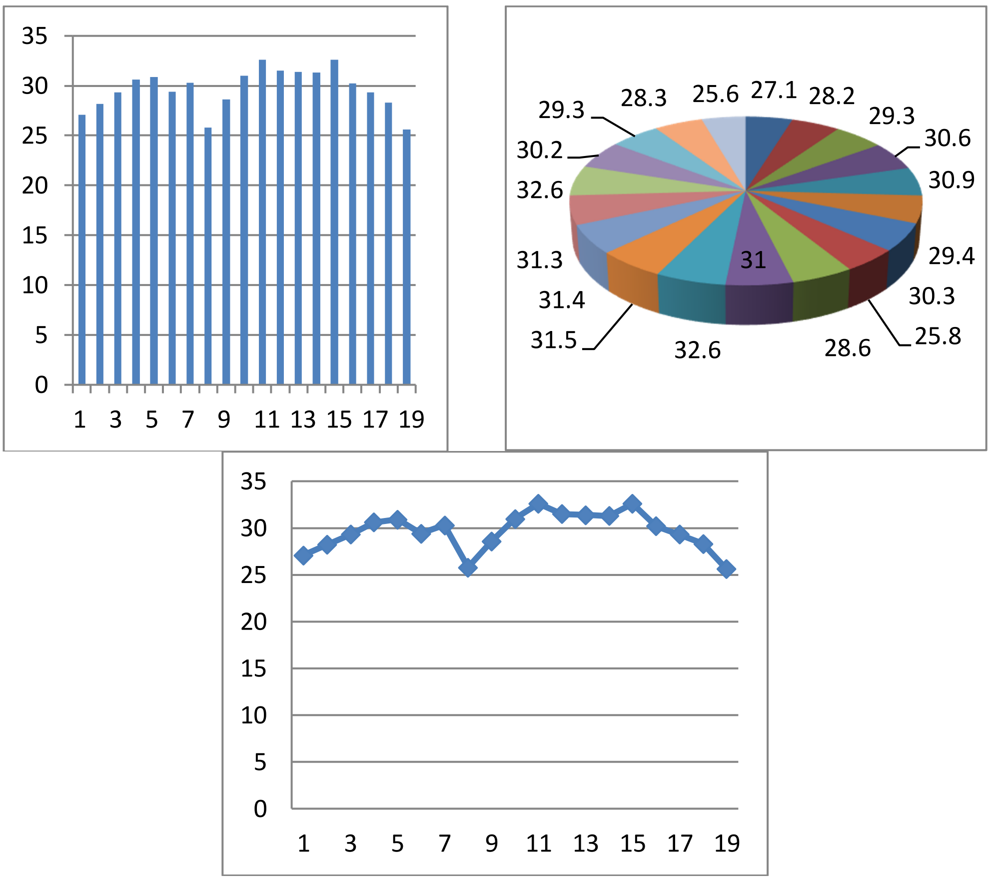

Fig. 1. Time series diagram

Table 1. The calculation of data to determine the trend of the time series

1	2	3	4	5	6	7
Np/n	Years of observations	y_tc/ha	t	t^2	$Y_t \cdot t$	$Y_t \cdot t^2$
1	2001	27.1	-9	81	-243.9	2195.1
2	2002	28.2	-8	64	-225.6	1804.8
3	2003	29.3	-7	49	-205.1	1435.7
4	2004	30.6	-6	36	-183.6	1101.6
5	2005	30.9	-5	25	-154.5	772.5
6	2006	29.4	-4	16	-117.6	470.4
7	2007	30.3	-3	9	-90.9	272.7
8	2008	25.8	-2	4	-51.6	103.2
9	2009	28.6	-1	1	-28.6	28.6
10	2010	31	0	0	0	0
11	2011	32.6	1	1	32.6	32.6
12	2012	31.5	2	4	63	126
13	2013	31.4	3	9	94.2	282.6
14	2014	31.3	4	16	125.2	500.8
15	2015	32.6	5	25	163	815
16	2016	30.2	6	36	181.2	1087.2
17	2017	29.3	7	49	205.1	1435.7
18	2018	28.3	8	64	226.4	1811.2
19	2019	25.6	9	81	230.4	2073.6
	Total	564	0	570	19.7	16349.3

Using the calculations in Table 1, we have:

$$\sum y_t = 564\,, \quad a_0 = \frac{1}{T}\sum y_t = \frac{564}{19} = 29.68\,, .a_1 = \frac{1}{\sum t^2}\sum y_t t = \frac{19{,}7}{570} = 0.035$$

From here, the equation of the linear trend (trend) of the cotton yield of the area is found:

$$y(t) = 0.035t + 29.68 \qquad (2)$$

With the help of statistical criteria, it was established that in the equation $y(t) = a_1 t + a_0$ the main hypothesis is rejected and an alternative hypothesis with the same level of significance is accepted. $H_0: \ a_1 = 0 \ \ H_1: \ a_1 \neq 0 \ \ \alpha = 0.05$
Substituting the value t = 2 into equation (2), we find the expected cotton yields in the Bukhara region in 2021, on average c/ha.$29,75$
Based on the observed data, we calculated $\Delta Y_t = Y_{t+1} - Y_t\,, \ \Delta^2 Y_t = \Delta Y_{t+1} - \Delta Y_{t'}, \ \Delta^3 Y_t = \Delta^2 Y_{t+1} - \Delta^2 Y_t$end differences (Table 2).

Table 2. To the calculation of data for determining finite differences

Years of observations embankment	$Y\_(t)$ c/ha	Y_t^2	ΔY_t	ΔY_t^2	$\Delta^2 Y_t$	$\Delta^2 Y_t^2$	$\Delta^3 Y_t$	$\Delta^3 Y_t^2$
1	2	3	4	5	6	7	8	9
2001	27.1	734.41						
2002	28.2	795.24	1.1	1.21				
2003	29.3	858.49	1.1	1.21	2.2	4.84		
2004	30.6	936.36	1.3	1.69	2.4	5.76	0.2	0.04
2005	30.9	954.81	0.3	0.09	1.6	2.56	-0.8	0.64
2006	29.4	864.36	-1.5	2.25	-1.2	1.44	-2.8	7.84
2007	30.3	918.09	0.9	0.81	-0.6	0.36	0.6	0.36
2008	25.8	665.64	-4.5	20.25	-3.6	12.96	-3	9
2009	28.6	817.96	2.8	7.84	-1.7	2.89	1.9	3.61
2010	31	961	2.4	5.76	5.2	27.04	6.9	47.61
2011	32.6	1062.76	1.6	2.56	4	16	-1.2	1.44
2012	31.5	992.25	-1.1	1.21	0.5	0.25	-3.5	12.25
2013	31.4	985.96	-0.1	0.01	-1.2	1.44	-1.7	2.89
2014	31.3	979.69	-0.1	0.01	-0.2	0.04	1	1
2015	32.6	1062.76	1.3	1.69	1.2	1.44	1.4	1.96
2016	30.2	912.04	-2.4	5.76	-1.1	1.21	-2.3	5.29
2017	29.3	858.49	-0.9	0.81	-3.3	10.89	-2.2	4.84
2018	28.3	800.89	-1	1	-1.9	3.61	1.4	1.96
2019	25.6	655.36	-2.7	7.29	-3.7	13.69	-1.8	3.24
Total	564	16816.56	-1.5	61.45	-1.4	106.42	-5.9	103.97

Table 2 calculates$v_k = \frac{\sum_{t=k}^{T}(\Delta^k Y_t)^2}{(T-t)C_{2k}^k}$coefficients of variation of differences and found that . Therefore, first-order finite differences eliminate the linear trend.$V_1 \approx V_2 \approx V_3$
The presence of autocorrelation in the series of cotton yield dynamics is checked using the Durbin-Watson criterion:

$$d = \frac{\sum_{t=1}^{T-1}(Y_{t+1}-Y_t)^2}{\sum_{t=1}^{T-1} Y_t^2} \qquad (3)$$

Calculated by formula (3) is compared = 0.0026 therefore, the average cotton yield in the region has an autocorrelation dependence

$$d_{observations} = 0.0026 \text{ c } d_{krit} = 1.08$$
$$d_{observations} < d_{krit} = 1.08$$

table value ([6], page 120). Since

$$Y_t = \rho Y_{t-1} + \varepsilon t, \rho = \mathrm{Cov}(\mathrm{Yt}, \mathrm{Yt}+1) = \mathrm{M}[(Y_t - \bar{y}_t)(Y_{t+1} - \bar{y}_t)]$$

Table 3. The calculation of data to determine indicators (4) of autocorrelation

1	2	3	4	5	6	7
T	Y_t	$Y_t \cdot Y_{t+1}$	$Y_t \cdot Y_{t+2}$	$Y_t \cdot Y_{t+3}$	$Y_t \cdot Y_{t+4}$	$Y_t \cdot Y_{t+5}$
2001	27.1					
2002	28.2	764.22				
2003	29.3	826.26	794.03			
2004	30.6	896.58	862.92	829.26		
2005	30.9	945.54	905.37	871.38	837.39	
2006	29.4	908.46	899.64	861.42	829.08	796.74
2007	30.3	890.82	936.27	927.18	887.79	854.46
2008	25.8	781.74	758.52	797.22	789.48	755.94
2009	28.6	737.88	866.58	840.84	883.74	875.16
2010	31	886.6	799.8	939.3	911.4	957.9
2011	32.6	1010.6	932.36	841.08	987.78	958.44
2012	31.5	1026.9	976.5	900.9	812.7	954.45
2013	31.4	989.1	1023.64	973.4	898.04	810.12
2014	31.3	982.82	985.95	1020.38	970.3	895.18
2015	32.6	1020.38	1023.64	1026.9	1062.76	1010.6
2016	30.2	984.52	945.26	948.28	951.3	984.52
2017	29.3	884.86	955.18	917.09	920.02	922.95
2018	28.3	829.19	854.66	922.58	885.79	888.62
2019	25.6	724.48	750.08	773.12	834.56	801.28
Total	564	8029.92	15270.4	14390.33	13462.13	12466.36

Using Table 3, the formulas from the literature determine the values of the autocorrelation coefficients (where: time shift, i.e. the time interval of one phenomenon lagging behind the other associated with it):$[1,2,3,4,5,6,7] R_L at \quad L = 1,2,3,4,5 L_{Lag}$

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

The difference of the value from zero gives reason to believe that there is a significant autocorrelation between the yield of cotton. Consequently, the yield of cotton in the Bukhara region this year depends on the yield of past years R_L [8-15].
Based on sample data, using the x7.2019 program package and Excel computer, the numerical characteristics of the average cotton yield -$y_t$in the Bukhara region are calculated (Table 4):

Table 4. Estimation of the main parameters of the dynamic series

Selected characteristics	Estimates of sample characteristics
Average cotton yield $\bar{y}_T cen/h$	29.68
Dispersion	4.15
Standard deviation σ_T	2.04
The coefficient of variation (%)v	6.87%
Asymmetry A_ς	-0.60
Excess E_{K_ς}	-0.25
Error of the mean, $\bar{y}_T m_y$	$m_y = \frac{\sigma_y}{\sqrt{n}} = 0.58$

Continuation of table № 2.

Marginal error m'_y	$m'_y = t \cdot m_y = 2.06 \cdot 0.47 = 0.97$
Standard deviation error σ_T	$m\sigma = \frac{\sigma}{\sqrt{2n}} \frac{2.04}{6.16} = 0.33$
Interval evaluation (95%) $\bar{y}_T \pm tm_y$ for cotton yield	$\bar{y}_T \pm tm_y = 29.68 \pm 0.97, (28.71; 30.65)$c/ha
Statistical hypothesis testing $H_0: P(X < x) = \Phi_{a,\sigma}(x)$	95% guarantee of the hypothesis is accepted H_0

3 Conclusions

Based on the above statistical analyzes, the dynamics of cotton yield in the Bukhara region as a time series with reliability: point and interval statistical estimates for sample characteristics (28.71; 30.65) c/ha; explicit types of the trend are determined and its linearity is established; using the Durbin-Watson criterion, it was established that autocorrelations in the considered series of dynamics have linear tendencies.$\bar{y_t} - \gamma = 0.95 y(t) = 0.035t + 29.68$

References

1. Kendall M., and Stuart A. Multivariate statistical analysis and time series. M.: Nauka, pp.65-68. (1976).
2. Lewis C. D. Methods of forecasting economic indicators. M.: FIS, (1986).
3. Ya.Ya.-F.Vaynu. Correlations of series of dynamics, M.: Statistics, (1977).
4. Ayvazyan S. A. Statistical study of dependencies. M.: Metallurgy, (1968).
5. Achilov M.U., Fayziev A.A. The analysis of dynamics of fruits and berry productivity grown in Uzbekistan, EPRA International journal of Research and Development. Vol.4(8). (2019).
6. Fayziev A. A., Vakhobov V. Forecasting the dynamics of cotton yields in the Fergana region. Tashkent Institute of Irrigation and Agricultural Mechanization. Journal Irrigation and Melioration. Vol. 6. pp.182 – 188. (2020).
7. Abdullayev A., Hidoyatova M. Exact method to solve finite difference equations of linear heat transfer problems. AIP Conference Proceedings, Vol. 2402, p.070021. (2021).
8. Abdullayev A., Kholturayev K., Safarbayeva N. Exact method to solve of linear heat transfer problems. In E3S Web of Conferences Vol. 264, p.02059 (2021).
9. Abdullayev A., Zhuvanov K., Ruzmetov K. A generalized solution of a modified Cauchy problem of class R_2 for a hyperbolic equation of the second kind. Journal of Physics: Conference Series, 1889(2), 022121 (2021).
10. Yuldashev T. K., Islomov B. I., Abdullaev A. A. On Solvability of a Poincare-Tricomi Type Problem for an Elliptic–Hyperbolic Equation of the Second Kind. Lobachevskii Journal of Mathematics, Vol. 42(3), pp. 663–675. (2021).
11. Abdullayev A. A., Ergashev T. G. Poincare-tricomi problem for the equation of a mixed elliptico-hyperbolic type of second kind. Vestnik Tomskogo Gosudarstvennogo Universiteta, Matematikai Mekhanika, Vol. 65, pp. 5–21. (2020).

12. Abdullaev A., Hidoyatova M. Innovative distance learning technologies. Journal of Critical Reviews, Vol. 7(11), pp. 337–339. (2020).
13. Ruzmetov K., Faiziev A., Murodov S., and Kurbonbekova, O. Statistical analysis and forecasting of cotton yield dynamics in Kashkadarya region of Republic of Uzbekistan. In E3S Web of Conferences, Vol. 389, p. 03080. (2023).
14. Rakhmanov S., Turgunov T. T., Kusharov Z. K., and Mengnorov A. A. Econometric methods for solving problems of analysis and forecasting dynamics of yield of agricultural crops. In IOP Conference Series: Earth and Environmental Science, Vol. 614, No. 1, p. 012165. (2020).
15. Papageorgiou E. I., Markinos A. T., and Gemtos T. A. Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. Applied Soft Computing, Vol. 11(4), pp.3643-3657. (2011).