Nat. Volatiles & Essent. Oils, 2021; 8(5): 12109-12115

Statistical analysis and forecast of the dynamics of cotton yield in the Namangan region of the republic of Uzbekistan

Akhtam Faiziev[1], Abrorjon Turgunov[2*], Khusniddin Mamadaliev[3], Alisher Usmonov[4], Begimov Uktam[5]

[1]Department of Information technologies and mathematics, Tashkent State Agrarian University, Tashkent, Uzbekistan
[2,3,4,5]Department of Algorithmic and mathematical modeling, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan

Abstract

Observations changing over time, conducted in an ordered sequence, are called a time series. In the article, the statistical regularity of the series of dynamics is studied - the average cotton yield in the Namangan region of the Republic of Uzbekistan (according to the data provided by the Central Statistical Office of the Republic of Uzbekistan for 1991-2018); the study was conducted using the method of statistical analysis of time series. Point and interval estimates for the average cotton yield were constructed, with a 95% guarantee, explicit trends were identified, and yield in the region was predicted for subsequent years. It was found, using the statistical Durbin-Watson tests, that the average cotton yield in the region has an autocorrelation relationship.

Keywords: discrete, dynamic, series, trend, seasonality, component, linear, hypothesis, autocorrelation, asymmetry, kurtosis

Introduction

In almost every area, there are phenomena that are important to study in their development and change over time. We can, for example, strive to predict the future based on knowledge of the past, control the processes; describe the characteristic features of a series based on a limited amount of information. When processing time series, the methods are mainly based on the methods developed by mathematical statistics for distribution series. To date, statistics has a variety of methods for analyzing time series from the most elementary methods to very complex ones [1-4].

In general cases, time series $\{y_t,\ t \in T\}$ consists of four components: trend; fluctuations in relation to the trend; seasonality effect; random component [1-6].

The study of the crop productivity of agricultural processes, as a discrete time series and the forecast of their productivity on the basis of experimental data, play an important role in determining the economic efficiency of farms, deckhand farms.

In this article, the processing and analysis of cotton yields for the observation period 1991-2018 in the Namangan region of Uzbekistan was conducted, as a discrete time series. Using the methods of statistical analysis of time series, point and interval estimates for the average yield of cotton were constructed, explicit types of trends were determined and the yield was predicted for subsequent years; various statistical hypotheses were tested.

The research conducted by Anderson [1], Kendal [2], Tikhomirov [3], Sulaimanov [4] and others was devoted to the study and analysis of time series.

Analysis of results and examples

The geometric image of the observed data (Table 1, column 3), the coordinate system give a basis for the first-approximation hypothesis that the trend part of the process has a linear dependence (Fig. 1) of form

$y(t) = a_1 t + a_0$ where unknown parameters are determined by the least squares method, i.e. based on experimental data, when solving the following system of normal equations:

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

Using the calculations based on the data of Table 1, we have:

$$\sum y_t = 740{,}9 \text{ c/ha},$$
$$a_0 = \frac{1}{T} \sum y_t = \frac{740.9}{28} = 26{,}46 \text{ c/ha},$$
$$a_1 = \frac{1}{\sum t^2} \sum y_t t = \frac{175.2}{1834} = 0{,}096 \text{ c/ha}.$$

Hence, the equation of the linear trend (tendency) of cotton yield in the Namangan region is determined as [1-7]:

$$y(t) = 0{,}096t + 26{,}46 \quad (2)$$

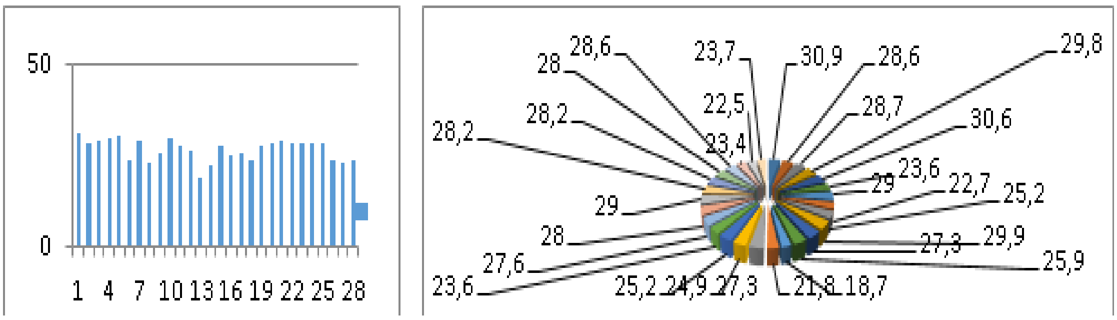
Substituting $t = 3$ into equation (2), we determine the expected cotton yield in the Namangan region in 2021, as equal, on average, to 27 kg/ha.

Table 1 Calculation of data to determine the trend of the time series

1	2	3	4	5	6	7
NN	Years of observation	y_t c/ha	t	t^2	$y_t \cdot t$	$y_t \cdot t^2$
1	1991	30,9	-13	169	-401,7	5222,1

2	1992	28,6	-12	144	-343,2	4118,4
3	1993	28,7	-11	121	-315,7	3472,7
4	1994	29,8	-10	100	-298	2980
5	1995	30,6	-9	81	-275,4	2478,6
6	1996	23,6	-8	64	-188,8	1510,4
7	1997	29	-7	49	-203	1421
8	1998	22,7	-6	36	-136,2	817,2
9	1999	25,2	-5	25	-126	630
10	2000	29,9	-4	16	-119,6	478,4
11	2001	27,3	-3	9	-81,9	245,7
12	2002	25,9	-2	4	-51,8	103,6
13	2003	18,7	-1	1	-18,7	18,7
14	2004	21,8	0	0	0	0
15	2005	27,3	1	1	27,3	27,3
16	2006	24,9	2	4	49,8	99,6
17	2007	25,2	3	9	75,6	226,8
18	2008	23,6	4	16	94,4	377,6
19	2009	27,6	5	25	138	690
20	2010	28	6	36	168	1008
21	2011	29	7	49	203	1421
22	2012	28,2	8	64	225,6	1804,8
23	2013	28,2	9	81	253,8	2284,2
24	2014	28	10	100	280	2800
25	2015	28,6	11	121	314,6	3460,6
26	2016	23,4	12	144	280,8	3369,6
27	2017	22,5	13	169	292,5	3802,5
28	2018	23,7	14	196	331,8	4645,2
	Total	740,9	14	1834	175,2	49514

Figure 1. Time series diagram.



Finite differences $\Delta Y_t = Y_{t+1} - Y_t,\ \ \Delta^2 Y_t = \Delta Y_{t+1} - \Delta Y_t,\ \ \Delta^3 Y_t = \Delta^2 Y_{t+1} - \Delta^2 Y_t$ were calculated based on the observed data (Table 2).

According to Table 2, the coefficients of variation of the differences $V_k = \frac{\sum_{t=k}^{T} (\Delta^k y_t)^2}{(T-k) C_{2k}^k}$ are calculated; and it is determined that $V_1 \approx V_2 \approx V_3$. Therefore, first-order finite differences eliminate the linear trend.

The presence of autocorrelation in the time series of cotton yield is checked using the Durbin - Watson criterion:

$$d = \sum_{t=1}^{T}(Y_{t+1} - Y_t)^2 / \sum_{t=1}^{T} Y_t^2. \quad (3)$$

Table 2 Calculating data for determining finite differences

1	2	3	4	5	6	7	8	9
Years of observation	$Y(t)$ c/ha	Y_t^2	ΔY_t	ΔY_t^2	$\Delta^2 Y_t$	$\Delta^2 Y_t^2$	$\Delta^3 Y_t$	$\Delta^3 Y_t^2$
1991	30,9	954,81						
1992	28,6	817,96	-2,3	5,29				
1993	28,7	823,69	0,1	0,01	-2,2	4,84		
1994	29,8	888,04	1,1	1,21	1,2	1,44	-1,1	1,21
1995	30,6	936,36	0,8	0,64	1,9	3,61	2	4
1996	23,6	556,96	-7	49	-6,2	38,44	-5,1	26,01
1997	29	841	5,4	29,16	-1,6	2,56	-0,8	0,64
1998	22,7	515,29	-6,3	39,69	-0,9	0,81	-7,9	62,41
1999	25,2	635,04	2,5	6,25	-3,8	14,44	1,6	2,56
2000	29,9	894,01	4,7	22,09	7,2	51,84	0,9	0,81
2001	27,3	745,29	-2,6	6,76	2,1	4,41	4,6	21,16
2002	25,9	670,81	-1,4	1,96	-4	16	0,7	0,49
2003	18,7	349,69	-7,2	51,84	-8,6	73,96	-11,2	125,44
2004	21,8	475,24	3,1	9,61	-4,1	16,81	-5,5	30,25
2005	27,3	745,29	5,5	30,25	8,6	73,96	1,4	1,96
2006	24,9	620,01	-2,4	5,76	3,1	9,61	6,2	38,44
2007	25,2	635,04	0,3	0,09	-2,1	4,41	3,4	11,56
2008	23,6	556,96	-1,6	2,56	-1,3	1,69	-3,7	13,69
2009	27,6	761,76	4	16	2,4	5,76	2,7	7,29
2010	28	784	0,4	0,16	4,4	19,36	2,8	7,84
2011	29	841	1	1	1,4	1,96	5,4	29,16
2012	28,2	795,24	-0,8	0,64	0,2	0,04	0,6	0,36
2013	28,2	795,24	0	0	-0,8	0,64	0,2	0,04
2014	28	784	-0,2	0,04	-0,2	0,04	-1	1
2015	28,6	817,96	0,6	0,36	0,4	0,16	0,4	0,16
2016	23,4	547,56	-5,2	27,04	-4,6	21,16	-4,8	23,04
2017	22,5	506,25	-0,9	0,81	-6,1	37,21	-5,5	30,25
2018	23,7	561,69	1,2	1,44	0,3	0,09	-4,9	24,01
Total	740,9	19856,19	-7,2	51,84	-13,3	405,25	-18,6	463,78

$d_{наб} = 0,0026$ is calculated by the formula (3) and compared with the table value $d_{крит} = 1,08$ [4]. Since $d_{наб} = 0,0026 < d_{крит} = 1,08$, the average cotton yield in the region has an autocorrelation relation

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \text{ where } \rho = Cov(Y_t, Y_{t+1}) = M[(Y_t - \bar{y}_t)(Y_{t+1} - \bar{y}_t)].$$

Table 3 To the calculation of data to determine the indices of autocorrelation relation

1	2	3	4	5	6	7
T	Y_t	$Y_t \cdot Y_{t+1}$	$Y_t \cdot Y_{t+2}$	$Y_t \cdot Y_{t+3}$	$Y_t \cdot Y_{t+4}$	$Y_t \cdot Y_{t+5}$
1991	30,9					

1992	28,6	883,74				
1993	28,7	820,82	886,83			
1994	29,8	855,26	852,28	920,82		
1995	30,6	911,88	878,22	875,16	945,54	
1996	23,6	722,16	703,28	677,32	674,96	729,24
1997	29	684,4	887,4	864,2	832,3	829,4
1998	22,7	658,3	535,72	694,62	676,46	651,49
1999	25,2	572,04	730,8	594,72	771,12	750,96
2000	29,9	753,48	678,73	867,1	705,64	914,94
2001	27,3	816,27	687,96	619,71	791,7	644,28
2002	25,9	707,07	774,41	652,68	587,93	751,1
2003	18,7	484,33	510,51	559,13	471,24	424,49
2004	21,8	407,66	564,62	595,14	651,82	549,36
2005	27,3	595,14	510,51	707,07	745,29	816,27
2006	24,9	679,77	542,82	465,63	644,91	679,77
2007	25,2	627,48	687,96	549,36	471,24	652,68
2008	23,6	594,72	587,64	644,28	514,48	441,32
2009	27,6	651,36	695,52	687,24	753,48	601,68
2010	28	772,8	660,8	705,6	697,2	764,4
2011	29	812	800,4	684,4	730,8	722,1
2012	28,2	817,8	789,6	778,32	665,52	710,64
2013	28,2	795,24	817,8	789,6	778,32	665,52
2014	28	789,6	789,6	812	784	772,8
2015	28,6	800,8	806,52	806,52	829,4	800,8
2016	23,4	669,24	655,2	659,88	659,88	678,6
2017	22,5	526,5	643,5	630	634,5	634,5
2018	23,7	533,25	554,58	677,82	663,6	668,34
Total	740,9	18943,11	18233,21	17518,32	16681,33	15854,68

Using the data given in Table 3 and formulas from the literature sources [1-4], we determine the values of the autocorrelation coefficients R_L *for* $L = 1, 2, 3, 4, 6$ (where

$L_{лаг}$ is the time shift, i.e., the time lag between one phenomenon and another associated with it):

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

The difference of R_L from zero gives reason to believe that there is a significant autocorrelation relationship between the cotton yields. Based on the sample data, using the x7.2019 program package and Excel [5-9], numerical characteristics y_t were calculated for the average cotton yield of the Namangan region (Table 4):

Table 4 Assessment of the main parameters of the time series

Sample characteristics	Estimates of sample characteristics
Average cotton yield $\bar{y}_T$ c / ha	26,46
Dispersion	9,31
Root mean square deviation	3,05
The coefficient of variation v(%)	11,52%
Asymmetry $A\varsigma$	-0,66
Kurtosis $E_{K\varsigma}$	-0,18
Error of mean value of $\bar{y}_T, m_y$	$m_y = \frac{\sigma_y}{\sqrt{n}} = 0,58$
Marginal error m'_y	$m'_y = tm_y = 2,06 \cdot 0,58 = 1,20$
Standard deviation error σ_T	$m_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{3,05}{7,48} = 0,41$
Interval estimate (95%) $\bar{y}_T \pm tm_y$ for cotton yield	$\bar{y}_T \pm tm_y = 26,46 \pm 1,20$ (25,26; 27,66) c/ha
Statistical Hypothesis Testing $H_0 : P(X < x) - \Phi_{a,\sigma}(x)$	95% guarantee of hypothesis H_0 accepte

Conclusions

Based on the above statistical analyses, and defined dynamics $\bar{y}_t$ of the cotton yield in the Namangan region as a time series with reliability $\gamma = 0,95$, the following conclusions could be drawn:

1) point and interval statistical estimates were constructed for the average cotton yield (25.26; 27.66) c/ha;

2) the explicit types of the trend were determined and its linearity was established;

3) the crop yield for subsequent years was predicted and various statistical hypotheses were tested;

4) it was established by the Durbin-Watson criterion that there exists a significant autocorrelation of the cotton yield. Consequently, the cotton yield in the Namangan region this year depends on the yield of previous years.

REFERENCES

T. Anderson, “Statistical analysis of time series,” Moscow: MIR, 1976, 759 p.

M. Kendal, A. Stewart “Multivariate statistical analysis and time series,” Moscow: Nauka, 1976, 736 p.

N.P. Tikhomirov, E.Yu. Dorokhina, “Econometrics,” Moscow: Textbook. “Exam”, 2003, 512 p.

B.A. Sulaimonov, A.A. Faiziev, Zh.N. Fayziev, “Tajriba malumotlarining statisticians told,” Tashkent: TashDAU, 2015, 124 p.

M.U. Achilov, A.A. Fayziev, “The analysis of dynamics of fruits and berry productivity grown in Uzbekistan,” EPRA Int. J. of Research and Development (IJRD), 4(8), 2019, p. 5-9.

A.A. Faiziev, T. Turgunov, “Statistical analysis and forecasting of the dynamics of cotton yield in the Republic of Uzbekistan,” J. Bul. of the Ins. of Math., 1, 2020, p.107-111.

B.N. Ishniyazov, A.A. Fayziev, D.Sh. Atabaeva, “Forecasting yields using dynamic range,” Elect. J. Cap. of Sci., 6 (23), 2020, p. 1-7.

Zh.N. Fayziev, AA Faiziev, AA Farkhodov, “Statistical analysis and forecasting of the dynamics of grape yield,”. J. Bul. of Agrar. Sci.of Uzb., 5 (83), 2020. p. 112-116.

A.A. Faiziev, O.Z. Karabashov, N.N. Musaeva, “Forecasting the dynamics of cotton yield in the Andijan region,” Prob. of Sci. Bul. of Sci. and Edu., 8 (111), 2021, p. 1-6.

Abdurakhmonov IY, Buriev ZT, Saha S, Jenkins JN, Abdukarimov A, Pepper AE. Phytochrome RNAi enhances major fibre quality and agronomic traits of the cotton Gossypium hirsutum L. Nature Communications. 2014 Jan 16;5(1):1-0.