American Journal of Applied Science and Technology

VOLUME Vol.05 Issue 06 2025
PAGE NO. 20-26
DOI 10.37547/ajast/Volume05Issue06-05

Statistical Analysis and Forecasting of Cotton Yield Dynamics in Region

Fayziyev Axtam Asraevich

Candidate of Physical and Mathematical Sciences, Acting Professor, Tashkent University of Economics and Pedagogy, Republic of Uzbekistan

Received: 14 April 2025; **Accepted:** 10 May 2025; **Published:** 12 June 2025

Abstract: Observations of a certain phenomenon, the nature of which changes over time, give rise to an ordered sequence, which is called a time series. In the article, using the method of statistical time series analysis, statistical pattern of time series ((y_t) ̅) -average yield of cotton in 5-regions of the Republic of Uzbekistan(on materials of CSB of the Republic of Uzbekistan for 20-30 years). Point and interval estimates for the average cotton yield were constructed with a 95% guarantee, explicit types of trends were determined, and the yield in the region was predicted for subsequent years. Using statistical criteria of Durbin-Watson, it was found that the average yield of cotton in the region has an autocorrelation dependence. The used methods of processing and analysis of dynamic series after testing can be used in the research of masters and researchers.

Keywords: Discrete, dynamic, series, trend, seasonality, component, linear, smallest, hypothesis, autocorrelation, asymmetry, kurtosis.

Introduction:

In almost every field, there are phenomena that are important to study in terms of their development and changes over time. For example, one may strive to predict the future based on knowledge of the past, manage a process, or describe the key features of a data series based on a limited amount of information.

When processing time series data, the methods largely rely on techniques developed by mathematical statistics for distribution series. At present, statistics offers a wide variety of time series analysis methods, ranging from the simplest to quite complex ones ([1,2,3,4,5]).

In general, a time series yt,t∈T{y_t, t ∈ T}yt,t∈T consists of four components: trend; fluctuations around the trend; seasonal effect; and random component ([1,2,3,4,5]).

In this study, cotton yield data over the past 20–30 years in five regions of the Republic of Uzbekistan is processed and analyzed as a discrete time series. Using statistical methods of time series analysis, point and interval estimates of the average cotton yield were constructed, explicit types of trends were identified, forecasts for future yields were made, and various statistical hypotheses were tested.

The study and analysis of dynamic series have been addressed in the works of Anderson [1], Kendall [2], Tikhomirov [3], Vainu [4], Sulaimanov [5], and others.

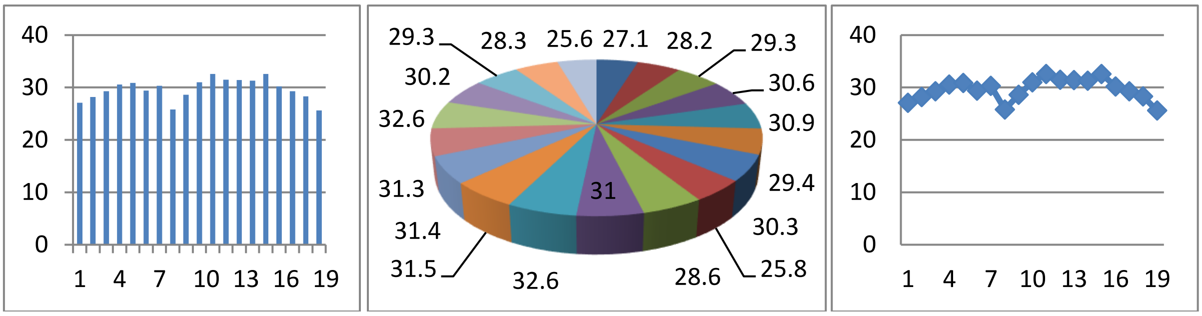
Analysis of Results and Examples.

A graphical representation of the observed data (Table 1, column 3, showing the statistical analysis of cotton yield dynamics for the Bukhara region) and the coordinate system provide a basis to preliminarily assume the hypothesis that the trend component of the process follows a linear dependence (Figure 1) of the form:

yt=a+bty_t = a + btyt=a+bt

where the unknown parameters are determined using the least squares method, i.e., based on the empirical data, by solving the following system of normal equations:

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



a) в) c)

Figure 1. Diagram of the Time Series.

Using the calculations from Table 1, we obtain:

$$\sum y_t = 564, \quad a_0 = \frac{1}{T}\sum y_t = \frac{564}{19} = 29{,}68, \quad a_1 = \frac{1}{\sum t^2}\sum y_t t = \frac{19{,}7}{570} = 0{,}035.$$

From this, the linear trend (tendency) equation for cotton yield in the region is determined as:

y(t)=0.035t+29.68(2)y(t) = 0.035t + 29.68 \tag{2}y(t)=0.035t+29.68(2)

Using statistical criteria, it was established that in this equation the null hypothesis

H0:a1=0H_0: a_1 = 0H0a1=0

is rejected, and the alternative hypothesis

H1:a1≠0H_1: a_1 \ne 0H1:a1□=0

is accepted at the significance level of α=0.05\alpha = 0.05α=0.05.

Proceeding to the calculation of the data for determining the time series trend.

Table 1

1	2	3	4	5	6	7
N	Years of Observation	y_t c/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

By substituting t=2t = 2t=2 into equation (2), the expected cotton yield in the Bukhara region for the year 2024 is found to be, on average, **29.82 centners per hectare (c/ha)**.

Results of the Statistical Analysis of Cotton Yield Dynamics in 5 Regions of the Republic of Uzbekistan (Table 2):

Table 2

№	Area	Trend part of the time series	Forecast average yields 2025 c/ha
1	Andijan	$y(t) = 0{,}30t + 29{,}47$	30,67
2	Bukhara	$y(t) = 0{,}035t + 29{,}68$	29,82
3	Samarkand	$\boldsymbol{y(t) = 0{,}11t + 23{,}82}$	24,37
4	Fergana	$y(t) = 0{,}096t + 26{,}46$	26,94
5	Khorezm	$y(t) = 0{,}091t + 25{,}54$	26
6	By Republic	$y(t) = 0{,}58t + 24{,}64$	26,96

For further statistical analysis, we calculate the finite differences (Table 2): Δy1,Δy2,…\Delta y_1, \Delta y_2, \ldotsΔy1, Δy2,…. Then, we proceed to compute the higher-order finite differences (Table 3).

Table 3

Data Calculations 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	$\Delta^2 Y_t$	$\Delta^2 Y_t^2$	$\Delta^3 Y_t$	$\Delta^3 Y_t^2$
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

According to Table 3, the coefficients of variation of the finite differences are calculated, and it is established that

V1≈V2≈V3.V_1 \approx V_2 \approx V_3.V1≈V2≈V3.

Therefore, the first-order finite differences eliminate the linear trend.

The presence of autocorrelation in the cotton yield time series is tested using the **Durbin–Watson criterion**:

d=∑t=2n(et−et−1)2∑t=1net2(3)d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{3}d=∑t=1net2∑t=2n(et−et−1)2(3)

Using formula (3), the computed value is:

dobserved=0.0026d_{\text{observed}} = 0.0026dobserved=0.0026

This is compared with the **critical value** from the table:

dcritical=1.08([5], p. 120).d_{\text{critical}} = 1.08 \quad \text{([5], p. 120)}.dcritical=1.08([5], p. 120).

Since

dobserved=0.0026<dcritical=1.08,d_{\text{observed}} = 0.0026 < d_{\text{critical}} = 1.08,dobserved=0.0026<dcritical=1.08,

it follows that the average cotton yield in the region exhibits **autocorrelation dependence**:

Yt=ρYt−1+et,Y_t = \rho Y_{t-1} + e_t,Yt=ρYt−1+et,

where

ρ=Cov(Yt,Yt+1)=E[(Yt−y¯t)(Yt+1−y¯t)].\rho = \text{Cov}(Y_t, Y_{t+1}) = \mathbb{E}[(Y_t - \bar{y}_t)(Y_{t+1} - \bar{y}_t)].ρ=Cov(Yt,Yt+1)=E[(Yt−y¯t)(Yt+1−y¯t)].

Using Table 4 and the formulas from sources [1, 2, 3, 4, 5], the **autocorrelation coefficients** RLR_LRL are determined for

L=1,2,3,4,5,L = 1, 2, 3, 4, 5,L=1,2,3,4,5,

where LLL represents the **lag**, i.e., the time shift or delay between the interrelated phenomena.

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

Data for Calculating Autocorrelation Coefficients

Table 4

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
Сумма	564	8029,92	15270,4	14390,33	13462,13	12466,36

A significant deviation of the autocorrelation coefficient RLR_LRL from zero provides grounds to assume that there is a substantial **autocorrelation dependence** in cotton yield. Consequently, the cotton yield in the Bukhara region this year depends on the yields of previous and subsequent years.

Based on sample data and using the software packages **EVM x7.2019** and **Excel**, the numerical characteristics yty_tyt for the **average cotton yield in the Bukhara region** are calculated (see **Table 5**):

Estimation of the Main Parameters of the Time Series Table 5

Sample Characteristics	Sample Estimates of Characteristics
Average cotton yield $\bar{y}_{\mathrm{T}}$ c/ha	29,68
Variance	4,15
Standard deviationотклонение σ_T	2,04
Coefficient of variation$v$ (%)	6,87 %
Skewness A_{ς}	-0,60
Kurtosis $E_{K\,\varsigma}$	-0,25
Standard error of the mean $\bar{y}_{\mathrm{T}}$, m_{y}	$m_y = \frac{\sigma_y}{\sqrt{n}} = 0{,}58$
Maximum error m'_{y}	$m'_y = t\, m_y = 2{,}06 \cdot 0{,}47 = 0{,}97$
Standard error of the standard deviation σ_T	$m_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{2{,}04}{6{,}16} = 0{,}33$
Confidence interval (95%): $\bar{y}_{\mathrm{T}} \pm t m_{\mathrm{y}}$ for cotton yield	$\bar{y}_T \pm t\, m_y = 29{,}68 \pm 0{,}97$ (28,71; 30,65) ц/га
Statistical hypothesis testing: $\mathrm{H}_0 : P(X < x) = \Phi_{\mathrm{a},\sigma}(\mathrm{x})$	95% гарантий гипотезы H_0 принимается

Results of the Statistical Analysis of Cotton Yield Dynamics in 5 Regions and in the Republic of Uzbekistan
Table 6

Parameter Estimation	Andijan	Bukhara	Samarkand	Fergana	Khorezm	Republic-wide or Across the Republic
Average cotton yield $\bar{y}_T$ c/ha	29,47	29,68	23,82	26,46	25,54	24,64
Variance	6,70	4,15	4,72	9,31	15,66	3,09
Standard deviation σ_T	2,58	2,04	2,17	3,05	3,96	1,76
Coefficient of variation v (%)	8,75	6,87	19,81	11,52	15,48	7,14
Skewness A ς	0,17	0,60	-1,81	0,66	-0,99	-1,49
Kurtosis E_K ς	0,23	0,25	2,68	0,18	2,17	1,65
Standard error of the mean $\bar{y}_T$, m_y	$m_y=\frac{\sigma_y}{\sqrt{n}}=$ 0,58	$m_y=\frac{\sigma_y}{\sqrt{n}}=$ 0,58	$m_y=\frac{\sigma_y}{\sqrt{n}}=$0, 40	$m_y=\frac{\sigma_y}{\sqrt{n}}=$ 0,58	$M_y=\frac{\sigma_y}{\sqrt{n}}=$ 0,75	$M_y=\frac{\sigma_y}{\sqrt{n}}=$ 0,44
Maximum error m'_y	$m'_y= t\ m_y =$ 2,09· 0,58 = 1,21	$m'_y=$ $t\ m_y =$ 2,06· 0,47 = 0,97	$m'_y= t\ m_y =$ 2,06· 0,40 =0,82	$m'_y=$ $t\ m_y =$ 2,06· 0,58 = 1,20	$m'_y=tm_y=$ 2,06· 0,747 = 1,54	$m'_y=tm_y=$ 2,06· 0,44 =0,94
Standard error of the standard deviation σ_T	$m_\sigma=$ $\frac{\sigma}{\sqrt{2n}}=\frac{2,58}{6,33}=$ 0,41	$m_\sigma=$ $\frac{\sigma}{\sqrt{2n}}=\frac{2,04}{6,16}=$ 0,33	$m_\sigma=\frac{\sigma}{\sqrt{2n}}=$ $\frac{2,17}{6,16}=0,35$	$m_\sigma=$ $\frac{\sigma}{\sqrt{2n}}=\frac{3,05}{7,48}=$ 0,41	$m_\sigma=$ $\frac{\sigma}{\sqrt{2n}}=\frac{3,956}{7,483}=$ 0,53	$m_\sigma=\frac{\sigma}{\sqrt{2n}}=$ $\frac{1,76}{7,483}=$ 0,24

Confidence interval (95%) for cotton yield: $\bar{y}_T \pm tm_y$	$\bar{y}_T$ ± t m_y = 29,47 ± 1,21 (28,26; 30,68) ц/га	$\bar{y}_T$ ± t m_y = 29,68 ± 0,97 (28,71; 30,65) ц/га	$\bar{y}_T$ ± t m_y = 23,82 ± 0,82 (23,00; 23,64) ц/га	$\bar{y}_T$ ± t m_y = 26,46 ± 1,20 (25,26; 27,66) ц/га	$\bar{y}_T$ ± tm_y= 25,54 ± 1,54, (24,00; 27,08) ц/га	$\bar{y}_T$ ± tm_y= 24,64 ± 0,94, (23,7; 25,6) ц/га
Statistical hypothesis test: $H_0 : P(X < x) = \Phi_{a,\sigma}(x)$	95% Guarantee hypothesis H_0 accepted	95% Guarantee hypothesis H_0 accepted	95% Guarantee hypothesis H_0 accepted	95% Guarantee hypothesis ы H_0 accepted	95% Guarantee hypothesis H_0 accepted	95% Guarantee hypothesis H_0 accepted

CONCLUSIONS

Based on the above statistical analyses of the dynamics of the average cotton yield y¯t\bar{y}_ty¯t in five regions of the Republic of Uzbekistan as a time series, with a confidence level of γ=0.95\gamma = 0.95γ=0.95, the following conclusions can be drawn:

- Point and interval statistical estimates for the average cotton yield have been constructed;
- Clear types of trends have been identified, and their linearity has been established;
- The average cotton yield has been forecasted for the coming years;
- Using statistical criteria, it has been proven that the average cotton yield in the region exhibits autocorrelation. Consequently, the cotton yield in a given year depends on the yields of previous and subsequent years.

REFERENCES

Т.Андерсон “Статистический анализ временных рядов”. Москва,“МИР”, 1976. 759 с.

М. Кендал, А. Стьюарт “Многомерный статистический анализ и временные ряды”.- Москва: “Наука”, 1976. -736 с.

Н.П.Тихомиров, Е.Ю.Дорохина “Эконометрика”.- Москва: Учебник. Изд. “Экзамен”, 2003. 512 с.

Я.Я.-Ф.Вайну “Корреляции рядов динамики”, М. “Статистика”, 1977

Б.А.Сулаймонов, А.А.Файзиев, Ж.Н. Файзиев “Тажриба маълумотларининг статистик таҳлили”. –Ташкент: Изд. ТашДАУ, 2015, 124 бет.

M.U.Achilov, A.A.Fayziev “The analysis of dynamics of fruits and berry froductivity grown in Uzbekistan” , EPRA International journal of Research and Development (IJRD). Volume: 4. Issue: 8. August 2019, 5-9 p .

А.А.Файзиев, Т. Тургунов “Статистический анализ и прогнозирование динамики урожайности хлопка в Республике Узбекистан”// Журнал.-Бюллетень Института Матема-тики. ISSN 2181-9483, http: //mib. Mathinst. Uz. Ташкент, 2020. № 1.-С.107-111.

Х.Ч.Буриев, А.А.Файзиев, А.Нишанова “Статистический анализ и прогнозирование динамики уржайности бахчавых культур”. ЖУРНАЛ, ВЕСТНИК аграрной НАУКИ УЗБЕКИСТАНА, BULLETIN OF THE AGRARIAN, SCIENCE OF UZBEKISTAN. № 1 (85), 2021. 47-52 стр.

А.А.Файзиев, О.З.Карабашов, Н.Н.Мусаева «Прогнозирование динамики урожайности хлопчатника Андижанской области». Издател. “Пробемы науки” Вестник науки и образование. Москва . Журнал N 8 (111). Апрел 2021, част - 2, 6- стр.

А.А.Файзиев, В.Вахобов “Прогнозирование динамики урожайности хлопчатника Ферганской области”, Ташкентскый институт ирригации и механизации сельского хозяйства .Журнал “Ирригация и милиорация”, № 6. 2020, 182 - 188 стр.