

基于ARIMA模型对陕西省苹果产量的分析与预测

贺成婷, 张 晶*

昌吉学院, 数学与数据科学学院, 新疆 昌吉

收稿日期: 2024年7月2日; 录用日期: 2024年7月23日; 发布日期: 2024年8月22日

摘 要

苹果含有碳水化合物、维生素以及微量元素多种营养物质, 有利于儿童的生长发育, 增强记忆力, 可以说营养价值在众多水果中堪称完美。在我国的苹果种植区内, 陕西省是适合生产的最佳地区, 并且随着时间的增长, 陕西省苹果产量呈现出逐渐上升的状态, 其产量位居首位。以统计学、经济学为基础, 采用ARIMA模型对陕西省苹果产量进行了分析预测研究。对陕西统计年鉴40年间陕西省苹果产量统计数据做统计分析, 依据求和自回归移动平均模型理论, 运用Eviews软件通过平稳检验、白噪声检验、模型识别、参数估计、模型的检验与优化等一系列的过程, 最终建立ARIMA(2,1,1)模型。利用建立的模型对陕西省近3年的苹果产量进行了预测比较, 对模型进行评估, 其平均相对误差在6.61%, 可以较好的反映陕西省苹果产量的发展趋势。使用ARIMA模型分析苹果产量等时间序列数据, 可以利用其优势来预测未来的产量趋势, 这对于农业生产规划和市场分析具有重要的实际应用价值。

关键词

苹果产量, 时间序列分析, 预测模型

Analysis and Prediction of Apple Yield in Shaanxi Province Based on ARIMA Model

Chengting He, Jing Zhang*

School of Mathematics and Data Science, Changji University, Changji Xinjiang

Received: Jul. 2nd, 2024; accepted: Jul. 23rd, 2024; published: Aug. 22nd, 2024

Abstract

Apples contain carbohydrates, vitamins and trace elements that are beneficial for children's growth

*通讯作者。

文章引用: 贺成婷, 张晶. 基于 ARIMA 模型对陕西省苹果产量的分析与预测[J]. 现代管理, 2024, 14(8): 1887-1893.
DOI: 10.12677/mm.2024.148219

and memory; in other words, their nutritional value is perfect among many fruits. In China's apple planting area, Shaanxi Province is suitable for production of the best area, and with the growth of time, Shaanxi Province apple output shows a state of gradual rise. Its output ranks the first. Based on statistics and economics, ARIMA model was used to analyze and forecast the apple yield in Shaanxi Province. The statistical data of apple yield in Shaanxi Province in the past 40 years of Shaanxi Statistical Yearbook were analyzed. Based on the theory of autoregressive moving average model, the ARIMA(2,1,1) model was finally established by using Eviews through a series of processes, such as stationary test, white noise test, model recognition, parameter estimation, model test and optimization. The established model is used to forecast and compare the apple yield of Shaanxi Province in the past three years. The average relative error of the model is 6.61%, which can better reflect the development trend of apple yield in Shaanxi Province and make short-term forecast. On the basis of the above, some suggestions on improving apple yield in Shaanxi province are put forward.

Keywords

Apple Yield, Time Series Analysis, Forecast Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

俗话说：一天一苹果，疾病远离我。苹果含有丰富的营养物质，含有维生素以及微量元素，有利于儿童的生长发育，增强记忆力，同样也是美容佳品，可以说营养价值在众多水果中堪称完美。我国苹果产量位于水果总产量首位，其栽培面积可达全球栽培面积的一半，苹果产量占全球总产量的 30%，在我国农产品产业中有着鲜明的国际竞争力。水果能够提供人类所需要的多种营养元素，水果对于民众的身体健康是十分重要的[1]。所以，对苹果产量的预测可以使苹果达到一个供求平衡的状态，为种植苹果的果农提供一定的理论条件。

国内对农产品及其水果产量的预测研究连绵不绝，其中以 ARIMA 模型对产量的预测最为广泛，赵嘉宝，陈杰，安霞，孙占海，张学东基于吐鲁番市葡萄产业的现状，利用 ARIMA(3,2,1)模型对 2018~2010 年的葡萄产量进行了预测，预测区间为喇叭状，预测结果较为精确[2]；阴明哲，李一帆，芦铃元，邢家轩，崔永福采用 ARIMA(3,1,2)模型对河北省 2018~2020 年的花生产量进行了预测并对花生产业进行了比较细密的分析[3]；陈鼎玉，万坚，程瀚锋基于 1948~2018 年的粮食数据，利用 ARIMA 模型对我国粮食产量进行了预测，并对粮食增长速度进行了预测[4]；李娟丽，许英选取 1990~2015 年的棉花产量数据进行产量预测研究，利用 Eviews 软件建立 ARIMA 模型进一步预测未来棉花产量预测[5]；高蕾利用以 1949~2013 年的粮食产量为数据进行分析，建立 ARIMA 模型，对安徽省粮食产量做预测，为有关部门制定政策提供了理论支撑[6]；刘立国，姜健，贾媛媛对锦州市 2000~2012 年的粮食数据做了统计分析，建立 ARIMA 模型，利用该模型对粮食产量进行预测，精准的预测了锦州市未来 3 年的粮食产量[7]。本文采用 ARIMA 模型分析陕西省苹果产量，其优势在于能够处理非平稳的时间序列数据，通过差分处理将非平稳时间序列转换为平稳时间序列，从而进行有效建模。此外，ARIMA 模型还能够捕捉序列的自回归和移动平均效应，通过选择适当的参数来提高预测准确性。这种模型的灵活性较强，可以通过调整参数适应各种不同类型的时间序列，如季节性、趋势性等。因此，使用 ARIMA 模型分析苹果产量等时间序列

数据，可以有效地预测未来的产量趋势，这对于农业生产和市场分析具有重要的指导意义。

2. 模型介绍

ARIMA 模型在 70 年代被提出，是时间序列中用来推测未来值一种数学方法。其基本思想：将长时间形成的数据视为随机数据，利用模型来描述未来预测值的数学方法。平稳的数据才能利用该模型来预测。如果数据没有表现出稳定的状态，是不能进行模型预测的，对数据做处理使得稳定，才能进行建模预测。

ARIMA(p,d,q)模型有以下结构：

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{1}$$

(1)式中， $\nabla^d = (1-B)^d$ ； $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ ，为平稳可逆 ARMA(p,q)模型的自回归系数多项式， $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ，为平稳可逆 ARMA(p,q)模型的移动平均系数多项式。

ARIMA 模型可以分为下面三个结构：

- 当 $d = 0$ 时，ARIMA(p,d,q)模型实际为 ARMA(p,q)模型；
- 当 $p = 0$ 时，ARIMA(0,d,q)可以简记为 IMA(d,q)模型；
- 当 $q = 0$ 时，ARIMA(p,d,0)可以简记为 ARI(p,d)模型。

3. ARIMA 模型的建立与分析

选取 1979~2018 年陕西省苹果产量数据为研究变量，记为序列 X_t 。对序列建立相应的 ARIMA 模型进行分析预测，利用 Eviews 软件进行相应的操作。

3.1. 序列的平稳性处理

该数据来源于陕西省统计年鉴，首先绘制出陕西省 1979~2018 年苹果产量的时序图(见图 1)。

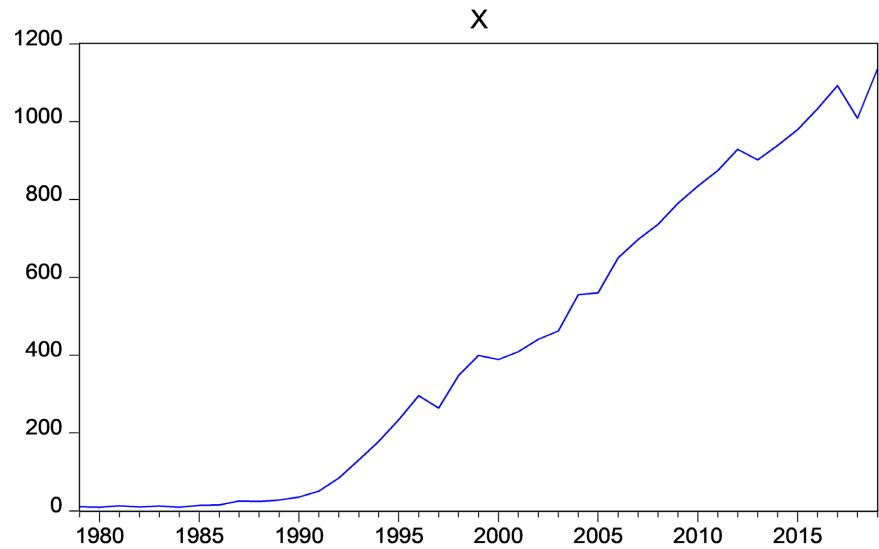


Figure 1. Time series diagram of apple production in Shanxi Province
图 1. 陕西省苹果产量时序图

序列 X_t 的时序图是有逐渐上升的趋势(见图 1), 不是平稳时间序列, 对序列 X_t 做平稳处理。

3.1.1. 平稳性处理

由于序列 X_t 呈现指数上升的趋势, 对序列 X_t 取对数, 并检验其平稳性(见图 2)。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.078579	0.2540
Test critical values: 1% level	-3.610453	
5% level	-2.938987	
10% level	-2.607932	

Figure 2. ADF test after logarithmic transformation
图 2. 取对数后的 ADF 检验

T 统计量的值为-2.078579(见图 2), 对应的概率 P 为 0.2540, 在 0.1 水平下接受原假设, 认为经过对数后的序列认为非平稳序列, 对该序列做一阶差分并检验其平稳性(见图 3)。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.510600	0.0000
Test critical values: 1% level	-3.610453	
5% level	-2.938987	
10% level	-2.607932	

Figure 3. ADF test after logarithmic difference
图 3. 对数差分后的 ADF 检验

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.043	-0.043	0.0798	0.778
		2 0.553	0.552	13.577	0.001
		3 -0.114	-0.114	14.165	0.003
		4 0.329	0.036	19.203	0.001
		5 -0.053	0.083	19.338	0.002
		6 0.251	0.064	22.442	0.001
		7 -0.097	-0.106	22.925	0.002
		8 0.011	-0.223	22.932	0.003
		9 -0.073	0.064	23.220	0.006
		10 -0.204	-0.253	25.547	0.004
		11 -0.028	-0.009	25.593	0.007
		12 -0.140	0.144	26.763	0.008
		13 0.008	-0.005	26.767	0.013
		14 -0.137	-0.026	27.981	0.014
		15 -0.007	-0.017	27.984	0.022
		16 -0.139	0.034	29.333	0.022

Figure 4. Autocorrelated and partial autocorrelated graphs
图 4. 自相关图和偏自相关图

对经过对数差分后的序列记为序列 Y_t 。对序列 Y_t 做 ADF 检验，结果如图 3 所示， T 统计量的值为 -6.510600 ，该值小于在 0.01、0.05、0.1 水平下的值，且相对应的概率 P 值为 0.0000，所以拒绝原假设，则序列 Y_t 为平稳的时间序列。

3.1.2. 白噪声检验

要建立 ARIMA 模型，首先要判断序列是否为白噪声序列。由图 3 可以看出，经过对数并做一阶差分后的序列 Y_t 是平稳的时间序列，对该序列进行纯随机性检验(见图 4)。

由图 4 可看出，序列 Y_t 的自相关函数相对应的 Q 统计量的 P 值基本在 0.05 以下，拒绝原假设，所以序列 Y_t 为平稳非白噪声序列，符合建模条件，可选用序列做模型拟合。

3.2. 模型估计与检验

3.2.1. 模型估计

由图 4 可知，序列 Y_t 的自相关系数在滞后 2 阶落在两倍标准差之外，偏相关系数拖尾。根据 ACF 和 PACF 的特性选择建立模型，建立 ARIMA(2,1,1)模型(见图 5)。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.120965	0.043410	2.786586	0.0087
AR(1)	-0.334988	0.171560	-1.952597	0.0591
AR(2)	0.434571	0.139352	3.118512	0.0037
MA(1)	0.718810	0.144770	4.965182	0.0000
R-squared	0.436947	Mean dependent var		0.118265
Adjusted R-squared	0.387266	S.D. dependent var		0.179570
S.E. of regression	0.140563	Akaike info criterion		-0.987027
Sum squared resid	0.671767	Schwarz criterion		-0.814649
Log likelihood	22.75351	Hannan-Quinn criter.		-0.925696
F-statistic	8.795035	Durbin-Watson stat		2.142861
Prob(F-statistic)	0.000187			
Inverted AR Roots	.51	-.85		
Inverted MA Roots	-.72			

Figure 5. Parameter estimation diagram of model ARIMA(2,1,1)
图 5. 模型 ARIMA(2,1,1)的参数估计图

由图 5 所示，参数有效性检验的 P 值小于 0.1，说明模型各项系数在 10%的水平下是显著的，通过检验，可以对模型进行下一步分析。

3.2.2. 模型检验

检验模型的拟合效果，即对模型的残差序列进行纯随机性检验(见图 6)，如果残差序列为白噪声序列的话，说明序列的信息已被充分利用，建立的模型是有效的，此时才能利用该模型进行下一步预测。

由图 6 可以看出，自偏相关函数相对应的 Q 统计量的概率值大于 0.05，即残差序列为白噪声序列，接受残差序列不存在序列相关性的原假设，通过白噪声检验，表明建立的模型是有效的。

3.3. 模型预测结果分析

利用建立的疏系数模型 ARIMA(2,1,1)，预测陕西省苹果的产量，以 Eviews 分析软件中的 forecast 函数进行操作(见表 1)。

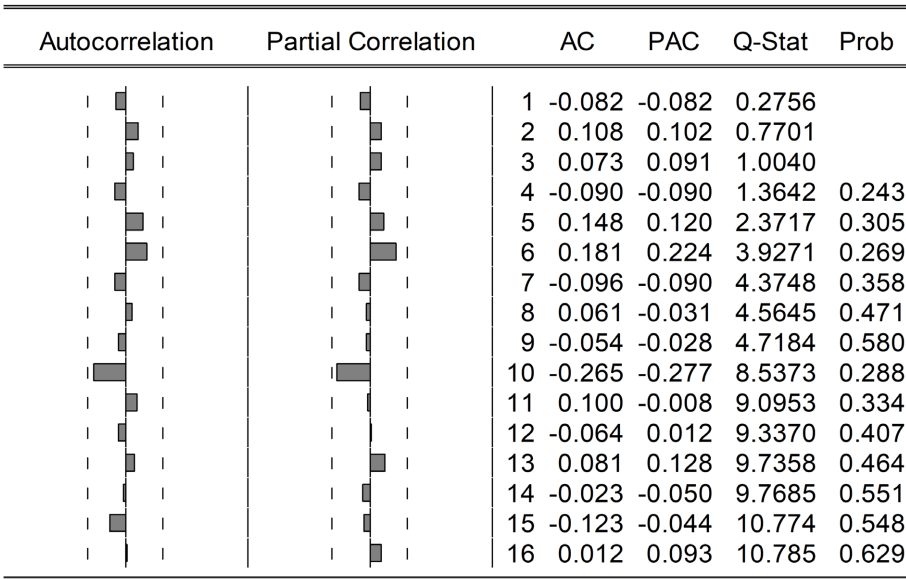


Figure 6. Residual test of model ARIMA(2,1,1)
图 6. 模型 ARIMA(2,1,1)的残差检验

Table 1. Comparison results of predicted and actual values
表 1. 预测值与真实值的比较结果

年份	实际产量(万 t)	预测产量(万 t)	相对误差(%)	平均绝对误差(%)
2019	1135.58	1043.02	-8.15%	
2020	1185.21	1249.59	5.43%	6.61%
2021	1242.46	1320.46	6.27%	

由表 1 可知, 2019 年、2020 年、2021 年的预测产量与实际产量有一定的误差, 其相对误差在 2020 年是最小的, 平均相对误差在 6.61%, 可以用来预测短期内的苹果产量。由于苹果产量受外界因素影响, 而时间序列模型是不考虑影响因素的条件下进行的预测, 所以该模型仅用于参考。利用该模型对陕西省 2022 和 2023 年两年的苹果产量进行预测, 其结果为: 1472.86 万 t、1662.73 万 t。

4. 结论

通过对陕西省苹果产量的分析可知, 陕西省苹果产量存在上升的趋势, 在 2018 年受外界情况的影响有所下降, 且陕西省苹果种植主要以延安市、渭南市以及咸阳市三个地区为主。本文运用统计学与经济学的理论方法作为基础, 结合陕西省苹果产业的发展实际情况, 运用时间序列方法 ARIMA 模型对近几年陕西省苹果产量做预测研究, 得到结论如下:

依据陕西统计年鉴中 1979~2018 年苹果产量数据, 利用 Eviews 软件对数据平稳分析处理, 取对数差分使数据成为平稳序列, 然后根据 ACF 和 PACF 进行模型识别, 通过模型检验与优化, 最终建立 ARIMA(2,1,1)模型, 预测出陕西省 2022 年的苹果产量为 1472.86 万 t, 2023 年的苹果产量为 1662.73 万 t。

本文构建 ARIMA 模型对陕西省苹果产量做预测, 预测平均相对误差为 6.61%, 可以较好的反应陕西省苹果产量的趋势, 但是该模型是在不考虑苹果产量的影响因素的情况下对未来趋势做预测, 由于苹果的产量受各种外界情况的影响, 例如, 如果出现气候灾害的情况下, 会导致苹果产量下降, 此时预测误

差会较大, 所以该模型仅用来进行短期预测, 如果模型被用来预测苹果的长期产量, 就需要进一步加强对模型的研究, 提高模型预测的精度, 使得预测更为精准, 预测更接近实际值。

利用 ARIMA 模型对陕西省苹果产量做分析并实行预测仅为苹果产业的发展及果农提供一定量的参考。除此之外, 利用组合模型对陕西省苹果产量的研究, 以及对苹果产量的影响因素研究, 也是日后研究的一个点。

参考文献

- [1] 任嘉颖. 山西运城苹果产量预测及影响因素分析[D]: [硕士学位论文]. 晋中: 山西农业大学, 2017.
- [2] 赵嘉宝, 陈杰, 安霞, 孙占海, 张学东. 基于 ARIMA 模型的吐鲁番市葡萄产量预测分析[J]. 江苏科技信息, 2019, 36(31): 34-39.
- [3] 阴明哲, 李一帆, 芦铃元, 邢家轩, 崔永福. 基于 ARIMA 的河北省花生产量预测分析[J]. 中国集体经济, 2021(3): 15-16.
- [4] 陈鼎玉, 万坚, 程瀚锋. 基于 ARIMA 模型的我国粮食产量预测[J]. 营销界, 2019(13): 95-96.
- [5] 李娟丽, 许英. 基于 ARIMA 模型的新疆棉花产量的预测分析[J]. 教育现代化, 2018. 5(17): 149-150+155.
- [6] 高蕾. 基于 ARIMA 模型的安徽省粮食产量预测研究[J]. 合肥学院学报(社会科学版), 2015, 32(5): 85-87.
- [7] 刘立国, 姜健, 贾媛媛. 基于 ARIMA 模型的锦州市粮食产量预测研究[J]. 辽宁工业大学学报(社会科学版), 2014, 16(3): 18-19+34.