

# 基于多因素农业相关性分析及预测研究

杨 鹏

四川建筑职业技术学院基础教学部, 四川 德阳

收稿日期: 2025年4月7日; 录用日期: 2025年4月28日; 发布日期: 2025年5月12日

## 摘 要

农业作为人类生存与发展的基础产业, 其发展受到多种因素的影响。本文通过选取1999年至2019年我国31个地区的农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量等因素, 建立了基于多因素农业相关性分析及预测模型。首先, 利用Z-score数据标准化法对原始数据进行标准化以获取初始分析数据; 然后, 建立了基于Pearson相关系数的农业相关性分析模型, 将农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量之间的关系分为7组极高相关关系和3组高度相关关系, 并基于该分析结果对农业发展提出合理性建议; 最后, 建立了基于多项式拟合的农业总产值预测模型, 得到农业总产值随时间变化的拟合函数为:  $f(x) = 63.9226x^2 + 1573.0684x + 8308.2916$ , 并对2020~2023年农业总产值进行了预测, 经检验其预测误差率低于3%, 表明所建立的预测模型精度高, 鲁棒性好。

## 关键词

农业, 相关性分析, Pearson相关系数, 多项式拟合, 预测

# Research on Multi-Factor Agricultural Correlation Analysis and Prediction

Peng Yang

Department of Basic Education, Sichuan College of Architectural Technology, Deyang Sichuan

Received: Apr. 7<sup>th</sup>, 2025; accepted: Apr. 28<sup>th</sup>, 2025; published: May 12<sup>th</sup>, 2025

## Abstract

Agriculture, as a fundamental industry for human survival and development, is influenced by multiple factors. This study selected agricultural data from 31 regions in China between 1999 to 2019, including total agricultural output value, grain yield, total agricultural machinery power, chemical fertilizer application amount, and rural electricity consumption. A multi-factor agricultural correlation analysis and prediction model was constructed. Firstly, the Z-score data standardization method

was applied to preprocess raw data for initial analysis. Subsequently, an agricultural correlation analysis model based on the Pearson correlation coefficient was established. This model identified 7 sets of extremely high correlations and 3 sets of highly significant correlations among the five factors. Based on these findings, practical recommendations for agricultural development were proposed. Finally, a prediction model based on polynomial fitting for agricultural output value was developed. The fitting function was determined as follows:  $f(x) = 63.9226x^2 + 1573.0684x + 8308.2916$ . Predictions for agricultural output values from 2020 to 2023 demonstrated high accuracy, with prediction error rates below 3%, confirming the model's precision and robust performance.

## Keywords

Agriculture, Correlation Analysis, Pearson Correlation Coefficient, Polynomial Fitting, Prediction

Copyright © 2025 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

农业作为国民经济的基础产业，其发展水平直接关系到国家的粮食安全和经济稳定。农业总产量作为衡量农业发展水平的重要指标，而粮食产量、农业机械总动力、化肥施用量和农村用电量作为影响农业总产值的关键因素，其研究可为相关部门制定政策提供数据支持，具有重大意义。姬翔宇老师利用 ARIMA 模型对浙江省粮食产量进行的研究，为浙江省粮食产量领域的决策制定提供了重要的参考依据，有助于确保粮食供应的稳定性和可持续性，为促进浙江省的粮食生产与管理提供了有力支持[1]；陈乾老师利用时间序列模型对浙江省农业生产总值进行的预测研究[2]，赵朝丛老师利用多元线性回归分析对农业机械化对我国农业生产总值产生的影响的研究[3]；蒙乐老师利用多元回归对农业产值进行了研究与预测[4]。但这些研究多集中于单因素的研究，而随着社会经济的不断发展，各地区的农业发展水平也得到了不同的提高，与之相应的影响农业发展水平的因素也较之前变得更为多样和复杂，因此，综合考虑多因素，即探究农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量之间的相关关系变得尤为迫切和重要。

## 2. 基于 Pearson 相关系数的多因素农业相关性分析模型

### 2.1. 数据预处理

初始分析数据即农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量等数据来源于国家统计局，为实现因素之间相关性的分析，首先按年分别对 1999 年至 2019 年之间 31 个地区的农业总产值(亿元)、粮食产量(万吨)、农业机械总动力(万千瓦)、化肥施用量(万吨)和农村用电量(亿千瓦时)求取平均值以获取初步分析数据。但由于不同因素之间单位(量级)的不同，其量纲存在较大差异，因此，为实现进一步分析，利用 Z-score 标准化法对该数据进行再处理以消除不同数据之间量纲的影响，便于后续分析和比较。

Z-score 标准化法也叫标准差标准化法，其通过公式：

$$z = \frac{X - \mu}{\sigma}$$

其中,  $X$  为个体观测值,  $\mu$  为总体数据的均值,  $\sigma$  为总体数据的标准差。

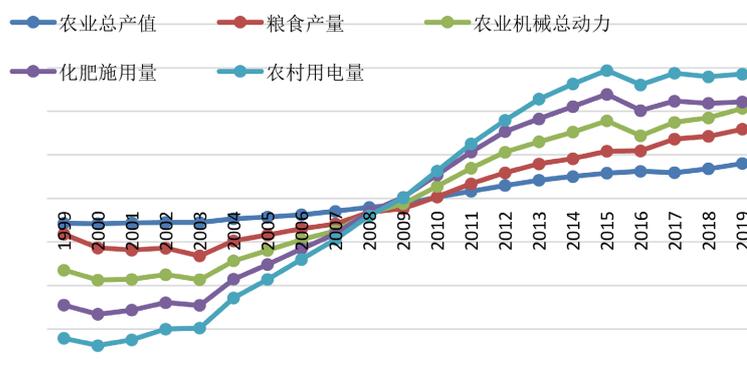
基于该标准化公式, 可得数据的 Z-score 标准化结果如表 1 所示。

**Table 1.** Z-score data standardization results

**表 1.** Z-score 数据标准化结果

年份	农业总产值	粮食产量	农业机械总动力	化肥施用量	农村用电量
1999	-1.142633422	-0.497462885	-1.650003053	-1.604687552	-1.52752957
2000	-1.154973533	-1.113059493	-1.478641469	-1.571408596	-1.432677713
2001	-1.123781798	-1.240119294	-1.349973321	-1.411996454	-1.360161937
2002	-1.098953346	-1.181225251	-1.216000071	-1.285417568	-1.213791752
2003	-1.102214396	-1.532444757	-1.096640142	-1.177855228	-1.045812171
2004	-0.929198734	-1.015866339	-0.919745123	-0.843431433	-0.854578587
2005	-0.851113519	-0.822012887	-0.707439035	-0.651186035	-0.685222124
2006	-0.748628658	-0.642786366	-0.507082058	-0.411220992	-0.486209158
2007	-0.584051493	-0.587822145	-0.30948042	-0.143533389	-0.251389614
2008	-0.404794308	-0.226744781	-0.037373604	0.051534777	-0.173557546
2009	-0.268899563	-0.198583278	0.220375634	0.297412778	-0.023974636
2010	0.066210419	0.009963711	0.477114377	0.530811117	0.177968029
2011	0.333415379	0.339429321	0.717777496	0.742370248	0.371993624
2012	0.595564653	0.584143993	0.952163883	0.942489684	0.513177555
2013	0.836813559	0.74879687	1.017642051	1.050646291	0.911234086
2014	1.010136268	0.816549663	1.219235914	1.176333777	1.039302261
2015	1.161764432	1.008553767	1.397601399	1.215411145	1.09379714
2016	1.249219434	0.939433196	0.694019677	1.158074921	1.174616277
2017	1.184207007	1.543650639	0.768727362	0.972945871	1.284074556
2018	1.363822591	1.494159634	0.845893758	0.666928043	1.220627078
2019	1.608089027	1.573446682	0.961826746	0.295778543	1.268114201

基于表 1 中的数据, 为判断农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量之间的初步关系, 通过绘制折线图如图 1 所示。



**Figure 1.** Metric trend chart  
**图 1.** 指标变化趋势图

从图 1 可看出, 总体而言, 随着时间推移, 我国的农业总产值、粮食产量、农业机械总动力、化肥施用量及农村用电量之间整体呈现线性相关的关系及逐渐递增的趋势。

因此, 为更精确地量化分析农业总产值、粮食产量、农业机械总动力、化肥施用量及农村用电量之间的线性相关关系, 可采用 Pearson 相关系数进行分析。

### 2.2. 基于 Pearson 相关系数的多因素农业相关性分析模型建立与求解

Pearson 相关系数可用于衡量自变量及因变量之间的线性关系, 主要内容为: 对于给定的两个变量  $X$  和  $Y$ , Pearson 相关系数即为它们之间协方差与各自标准差乘积的比值, 计算公式如下:

$$r = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

其中,  $r$  表示 Pearson 相关系数,  $Cov(X,Y)$  表示变量  $X$  和  $Y$  的协方差,  $\sigma(X)$  和  $\sigma(Y)$  分别表示变量  $X$  和  $Y$  的标准差。

$r$  取值范围介于  $[-1, 1]$  之间, 当  $r=1$  时, 表示两个变量之间存在完全正向线性关系; 当  $r=-1$  时, 表示两个变量之间存在完全负向线性关系; 当  $r=0$  时, 表示两个变量之间不存在线性关系, 具体而言, 当相关系数接近 1 或  $-1$  时, 说明两个变量之间线性关系较强; 当相关系数接近 0 时, 说明两个变量之间的线性关系较弱或不存在[5]。

为实现基于 Pearson 相关系数的多因素农业相关性分析模型建立与求解, 选取 Z-score 标准化后的数据作为 Pearson 相关系数分析的原始数据, 利用 SPSSPRO 将表 1 中农业总产值、粮食产量、农业机械总动力、化肥施用量及农村用电量等数据导入, 计算得到彼此间的相关系数如表 2 所示。

Table 2. Correlation coefficient table

表 2. 相关系数表

	农业总产值	粮食产量	农业机械总动力	化肥施用量	农村用电量
农业总产值	1	0.975	0.933	0.894	0.98
粮食产量	0.975	1	0.892	0.856	0.949
农业机械总动力	0.933	0.892	1	0.979	0.966
化肥施用量	0.894	0.856	0.979	1	0.948
农村用电量	0.98	0.949	0.966	0.948	1

为更直观地展现 5 种因素间的相关性情况, 绘制相关系数热力图如图 2 所示。

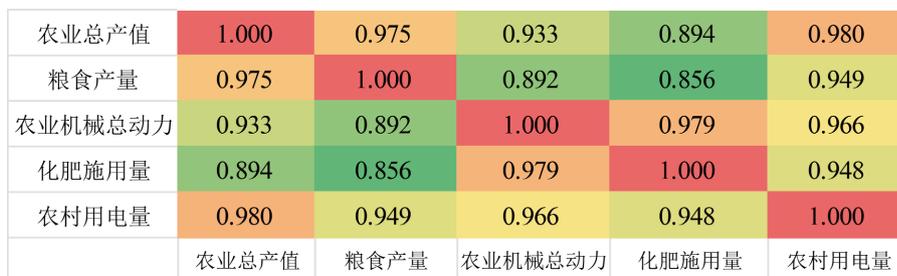


Figure 2. Correlation heatmap

图 2. 相关系数热力图

从表 2 及图 2 中, 均可看出农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量之间的相关系数都在 0.8 以上, 可见这 5 项因素之间高度正相关, 其变化趋势非常一致。进一步可将其分为极高相关[0.9~1]: 农业总产值与粮食产量, 农业总产值与农业机械总动力, 农业总产值与农村用电量, 粮食产量与农村用电量, 农业机械总动力与化肥施用量, 农业机械总动力与农村用电量, 化肥施用量与农村用电量, 共 7 组; 高度相关[0.7~0.9]: 农业总产值与化肥施用量, 粮食产量与农业机械总动力, 粮食产量与化肥施用量, 共 3 组。

由于农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量之间相关关系密切, 因此, 为实现农业发展水平的综合提升, 需要构建协同机制, 具体建议有:

1) 优化要素投入结构: 由于农业机械化, 农村用电量与农业总产值之间均为极高相关, 因此提升农业机械化与能源利用效率能有效提高农业总产值; 而化肥施用量和农业总产值为高度相关, 且化肥的施用对土壤具有一定影响, 因此需要科学调控化肥的施用。

2) 强化协同应用: 由于粮食产量与农业总产值之间为极高相关关系, 因此可以集成“四良”融合模式, 即推动良田(高标准农田)、良种(优质品种)、良机(智能农机)、良法(高产技术)协同应用, 有效提高粮食产量; 建立农业全产业链大数据平台, 实时监测机械作业、用电负荷、化肥施用等数据, 优化资源配置, 构建数据驱动决策系统。

3) 构建动态监测与调控体系: 将粮食产量与机械动力、用电量等纳入农业考核体系, 设置单位产值能耗、化肥利用率等约束性指标, 定期发布农业资源利用效率报告, 指导地方优化投入结构。

通过上述建议的系统实施, 实现协同增效, 从而推动农业从粗放型增长向集约化、智能化、可持续的高质量发展转型[6]。

### 3. 基于多项式拟合的农业总产值预测模型

#### 3.1. 基于多项式拟合的农业总产值预测模型的建立与求解

农业总产值作为衡量农业发展水平最直接的指标, 其研究对于分析农业发展水平及制定未来发展政策具有重大意义。多项式拟合广泛应用于各种数据分析、预测等领域, 可通过对某个数据集进行多项式拟合, 得到一条趋势线, 实现数据集的分析及为未来的预测提供参考[7]。

在得到 1999 年至 2019 年我国农业总产值数据的基础上, 利用 MATLAB 编程软件实现对农业总产值的多项式拟合预测。首先, 为便于程序实现, 将 1999 年定义为 1, 依次类推; 接着, 由于多项式拟合过程中,  $x$  的最高次数对于拟合效果起重要作用, 本文采用自适应的方式选择合适的拟合最高次数, 经过调试, 当拟合最高次数设置为 2 时, 拟合效果最佳, 此时得到的多项式拟合函数如下:

$$f(x) = 63.9226x^2 + 1573.0684x + 8308.2916$$

其拟合各项评价指标为: 误差平方和为 2.097e+08, 决定系数为 0.9706, 校正后的决定系数为 0.9674, 从拟合评价指标可以看出, 所实现的多项式拟合函数拟合效果良好。

基于所得到的多项式拟合函数, 预测接下来 4 年的农业总产值如表 3 所示。

**Table 3.** Polynomial regression forecasting of total agricultural output value from 2020 to 2023 (billion yuan)

**表 3.** 多项式拟合预测 2020~2023 年农业总产值(亿元)

年份	2020 年	2021 年	2022 年	2023 年
预测值	73854	78304	82881	87587
真实值	71748.23	78339.51	84438.58	87073.38
误差	2105.77	-35.51	-1557.58	513.62
误差率	2.9349%	-0.0453%	-1.8446%	0.5899%

为检验预测模型的效果，结合国家统计局公布的 2020 至 2023 年的农业总产值真实数据，从表 3 可以看出，所建立的预测模型，其预测效果误差率的绝对值均小于 3%。

为更直观地表示 1999 年至 2023 年的拟合及预测效果，特绘制拟合预测趋势图如图 3 所示。

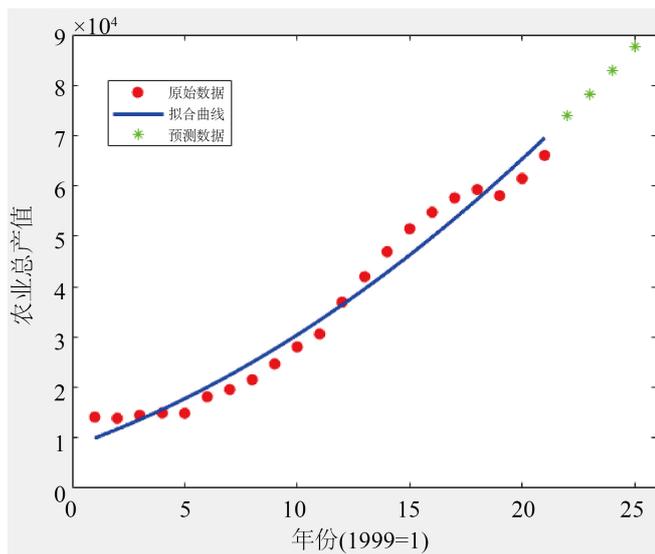


Figure 3. Fitted and forecasted trend chart of total agricultural output value from 1999 to 2023

图 3. 农业总产值 1999~2023 年拟合预测趋势图

### 3.2. 基于多项式拟合的农业总产值预测模型分析评价

所建立的预测模型，预测效果误差率的绝对值均小于 3%，经过分析，其误差产生的可能原因有以下三点：

- 1) 数据量样本不足，一定程度上降低了模型对新数据的预测能力；
- 2) 多项式拟合对阶数敏感，阶数过高易导致过拟合，表现为训练误差低但泛化误差高；
- 3) 模型假设数据符合多项式分布，但农业总产值受气候、政策等非线性因素影响时，可能导致系统性偏差，多项式拟合方式可能不能很好地拟合数据变化情况。

在未来工作中，可基于上述误差产生原因进行针对性调整以提高模型预测的准确性和实用性，如增加样本容量；增强多项式拟合阶数的调整；添加其它拟合方式(平滑拟合，高斯拟合，傅立叶拟合等)预测结果对比等。

因多项式拟合对局部波动捕捉较好，该基于多项式拟合的农业总产值预测模型可适用于以下场景：

- 1) 适用于趋势平稳，无突发干扰的短期预测；
- 2) 适用于单一因素驱动的产值预测；
- 3) 适用于数据需呈现多项式可描述的趋势(如周期性波动或平缓增长)的预测。

总体而言，所建立的基于多项式拟合的农业总产值预测模型在预测农业总产值时，预测精度高，鲁棒性强。该模型应用范围广，推广性强，可适用于单一因素驱动的短期预测类问题，但需警惕过度拟合和数据质量问题。

## 4. 结语

本文基于 1999~2019 年我国 31 个地区的农业相关数据，构建了基于多因素的农业相关性分析及预测

模型。通过 Pearson 相关系数深入分析了农业总产值、粮食产量、农业机械总动力、化肥施用量和农村用电量等对农业发展起重要影响的因素之间的关系，将其分为 7 组极高相关关系和 3 组高度相关关系，表明农业系统内部存在显著的协同作用，并据此提出农业发展需注重要素间的协同优化建议。同时，构建了基于多项式的农业总产值预测模型，对 2020~2023 年农业总产值的预测误差率低于 3%，可见模型的有效性和鲁棒性。该研究为农业政策制定提供了量化依据，兼具理论价值和实践指导意义。

## 参考文献

- [1] 姬翔宇. 基于 ARIMA 模型对浙江省粮食产量的研究[J]. 统计学与应用, 2023, 12(6): 1763-1774.
- [2] 陈乾. 基于时间序列的浙江省农业生产总值预测研究[D]: [硕士学位论文]. 杭州: 浙江农林大学, 2019.
- [3] 赵朝丛, 杨明. 农业机械化对我国农业生产总值的影响[J]. 农村实用技术, 2020(11): 82-83+86.
- [4] 蒙乐. 基于 SPSS 软件的农业产值研究与预测[J]. 应用数学进展, 2024, 13(5): 2155-2159.
- [5] 薛泽茜. 基于主成分分析法的福建省农业经济影响因素研究[J]. 现代农业科技, 2024(24): 175-177+191.
- [6] 史志乐. 用科技创新惠农兴农[N]. 光明日报, 2024-09-18(2).
- [7] 杜德鹏. 作物环境因子的分析及其预测方法研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2021.