

基于回归模型的伊犁州直第一产业生产总值实证分析与预测

韩柳¹, 杨松², 王浩东¹, 刘璐璐¹

¹新疆理工学院, 新疆 阿克苏

²浙江嘉兴学院, 浙江 嘉兴

收稿日期: 2022年7月2日; 录用日期: 2022年8月1日; 发布日期: 2022年8月8日

摘要

为促进伊犁州直第一产业的又好又快发展, 本文基于线性回归的相关理论, 将伊犁州直2009~2019年第一产业生产总值作为数据基础, 利用统计软件进行分析并建立回归模型, 对伊犁州直未来三年的第一产业生产总值进行预测。结果看出伊犁州直第一产业生产总值近年呈现增长趋势, 但对第一产业方面的发展需求迫在眉睫, 进而对其进一步发展提出相关意见和建议。

关键词

第一产业, 回归方程, 预测分析

Empirical Analysis and Forecast on the GDP of Primary Industry in Yili State Based on the Theory of Linear Regression

Liu Han¹, Song Yang², Haodong Wang¹, Lulu Liu¹

¹Xinjiang Institute of Technology, Aksu Xinjiang

²Jiaxing University, Jiaxing Zhejiang

Received: Jul. 2nd, 2022; accepted: Aug. 1st, 2022; published: Aug. 8th, 2022

Abstract

In order to promote the sound and rapid development of the primary industry of Yili prefecture, based on the theory of linear regression; this paper uses the statistical software to analyze and

establish the regression model, as the base data in the GDP of primary industry in Yili state from 2009 to 2019, to predict the future three years of the first industry in GDP. The results show that the GDP of the primary industry of Yili Prefecture has shown an increasing trend in recent years, but the demand for the development of the primary industry is imminent, and put forward relevant opinions and suggestions for its further development.

Keywords

The First Industry, Regression Equation, Prediction Analysis

Copyright © 2022 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

伊犁哈萨克自治州位于祖国西北边陲，有着“三山两盆两谷”的地貌特征，土地水利矿产等资源丰富。近年来，国家出台各项惠民政策大力扶持新疆，为新疆的经济跨越式发展提供支撑，在“一带一路”背景下，伊犁州依托当地农业资源优势大力发展特色产业，在中央、自治区和自治州强有力的强农惠农政策支持下，州直农业也迎来了春天，为助力于农民增收并使边塞粮仓和河谷特色农业的优势日渐显现，对伊犁州直第一产业发展的分析研究是很有必要的。

第一产业是指农林牧渔业，其生产总值反映了第一产业发展的特征和趋势。本文以一元线性回归模型作为理论基础，对伊犁州直第一产业生产总值进行实证分析，将伊犁州直 2009~2016 年第一生产总值作为预测模型，利用统计软件进行分析及建立回归方程，然后对 2017~2019 年的第一产业生产总值进行拟合度检验，进而对伊犁州直 2020~2022 年的第一产业生产总值进行预测，最后为伊犁州直第一产业的又好又快发展提供相关意见和建议[1] [2] [3]。

2. 相关理论[4] [5] [6] [7]

一元线性回归模型实质是通过分析变量之间的相关关系，通过回归方程的方式对变量间的数量变化规律进行说明和反映，进而掌握变量与变量之间的影响程度，最终达到科学控制和预测的目的。其思路就是建立 Y 关于 X 的回归方程，并在已知 X_1, X_2, \dots, X_n 的情况下，利用回归方程来预测 Y 的平均值。

2.1. 回归方程的建立

一般的一元线性模型为： $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n$ ，其中 x_i 为自变量， y_i 是因变量， ε_i 是随机扰动项，其中优度统计量

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_0$$

得到样本回归直线方程 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, i = 1, 2, \dots, n$ 。

2.2. 回归方程的拟合度检验

回归方程的拟合优度是对样本数据点与回归线周围接近程度的衡量，是回归方程对样本观测值拟合程度的判断依据，对于简单直线回归方程来说，一般使用 R^2 对拟合优度进行检验，即判定系数。 R^2 说明了回

归方程所能解释的变差比例,其取值在 0~1 之间, R^2 越趋近于 1, 即回归方程对样本数据点的拟合度越高; 反之, R^2 越趋近于 0, 说明回归方程对样本数据点的拟合度越低, 判决系数 $R^2 = \frac{S_R}{S_T} = 1 - \frac{S_E}{S_T}$ 。其中

$$S_T = \sum_{i=1}^n (y_i - \bar{y})^2 = S_{yy} = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$$

$$S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = b^2 S_{xx} = b^2 \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right], \quad S_E = S_T - S_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2。$$

2.3. 预测估计

若回归方程经受住全部检验且具备较高的拟合度, 即可使用该回归方程预测。预测是指在给定 x_0 对相应的 y 的取值 y_0 做出估计。但点预测对估计的精确度无法掌握, 在对实际值做预测时, 需有一定的显著性水平, 确定其置信水平为 $1-\alpha$ 的置信区间, 即在 $X = x_0$ 处, Y 的置信水平为 $1-\alpha$ 的置信区间为:

$$\left[\hat{y}_0 \pm t_{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

3. 实证分析

根据伊犁州直 2009~2019 年统计年鉴[8] [9], 将伊犁州直第一产业生产总值作为数据基础, 以 2009~2016 年的第一产业生产总值为预测模型, 对 2017~2019 年的第一产业生产总值进行检验, 进而对伊犁州直 2020~2022 年的第一产业生产总值进行预测。(表 1)

Table 1. GDP of primary industry of Yili prefecture from 2009~2019 (ten thousand yuan)

表 1. 伊犁州直 2009~2019 年第一产业生产总值(万元)

年份	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
生产总值	807,800	986,517	1,155,822	1,368,794	1,616,612	1,775,638	1,879,565	1,861,631	1,678,553	1,767,902	1,974,362

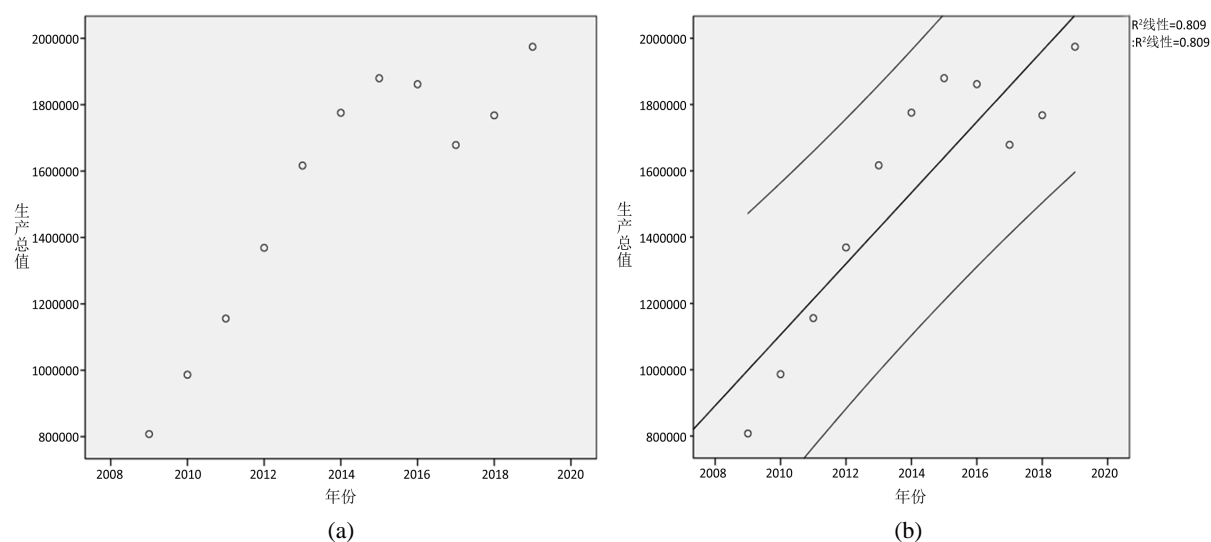


Figure1. Change chart of GDP of Yili primary industry from 2009~2019

图 1. 伊犁州直 2009~2019 年第一产业生产总值的变化图

3.1. 回归方程模型建立

用 SPSS 软件对上述 2009~2019 年数据进行处理

- 1) 见图 1, 散点图(a):
- 2) 得出拟合线, 个体的置信区间设置为 95%: 根据图 1(a)可大致判断出 X 与 Y 有线性相关关系, 如图 1(b)。
- 3) 假设年份是自变量 X, 生产总值是因变量 Y, 则得出回归直线方程

$$Y = \beta_0 + \beta_1 X$$

用 SPSS 软件继续处理上述数据, 得到相关数据如下表 2、表 3:

Table 2. Descriptive statistics

表 2. 描述性统计量

	均值	标准偏差	N
生产总值	1,533,926.91	394,830.464	11
年份	2014.00	3.317	11

Table 3. Coefficient

表 3. 系数

模型	非标准化系数		标准系数	t	Sig.	B 的 95.0%置信区间	
	B	标准误差	试用版			下限	上限
1 (常量)	-214,058,731.091	34,966,506.227		-6.122	0.000	-293,158,463.610	-134,958,998.572
1 年份	107,047.000	17,361.700	.899	6.166	0.000	67,772.107	146,321.893

- 4) 计算回归方程:

由图表中的数据得: $\beta_1 = 107047$, $\beta_0 = -214058731.091$ 。

则回归方程为 $y = -214058731.091 + 107047x$ 。

3.2. 回归方程的拟合优度的检测

对该模型做拟合度检验和 F 检验, 具体如表 4、表 5。

Table 4. The model summary

表 4. 模型汇总

模型	R	R 方	调整 R 方	标准估计的误差	更改统计量				
					R 方更改	F 更改	df1	df2	Sig. F 更改
1	0.899 ^a	0.809	0.787	182,091.042	0.809	38.016	1	9	0.000

Table 5. Anova

表 5. Anova

模型	平方和	df	均方	F	Sig.
1 回归	1,260,496,622,990.000	1	1,260,496,622,990.000	38.016	0.000 ^b
残差	298,414,329,084.909	9	33,157,147,676.101		
总计	1,558,910,952,074.909	10			

1) 计算判决系数

由以上的计算结果知,

$$S_{yy} = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 = 1558910952074.909, \quad S_T = S_{yy} = 1558910952074.909$$

$$S_R = \beta_1^2 S_{xx} = 1260496622990.000, \quad R^2 = \frac{S_R}{S_T} \approx 0.809。$$

由 $R^2 \approx 0.809$, 表明回归方程解释了总离差的 80.9%, 得出年份与第一产业生产总值之间的线性关系较强。

2) 计算估计标准误差: $S_E = 298414329084.909$, $\sigma = \sqrt{\frac{S_E}{n-2}} = 182091.042$ 。

3.3. 回归方程的显著性检验

1) 提出假设 H_0 : 线性关系不显著; H_1 : 线性关系显著。

2) 构造检验统计量。当 H_0 为真时, 检验统计量

$$F = \frac{S_R/1}{S_E/(n-2)} \sim F(1, n-2)$$

3) 给定显著性水平 $\alpha = 0.05$, 得临界点 $F_{1-0.05}(1, 9) = 5.12$, 由于

$$F = 38.016 \geq F_{1-\alpha}(1, 9) = 5.12$$

则拒绝 H_0 , 认为年份与第一产业生产总值之间具有显著的线性关系。

3.4. 利用回归方程预测

根据得出的回归方程 $y = -214058731.09 + 107047x$, 对伊犁州直 2017~2019 年第一产业生产总值进行预测并与实际数值进行比较, 发现误差率相对偏小。(表 6)

Table 6. 2017~2019 forecast estimates

表 6. 2017~2019 年预测估计

年份	预测值	实际值	误差	误差率
2017	1,855,068	1,678,553	176,515	10.52%
2018	1,962,115	1,767,902	194,213	11%
2019	2,069,162	1,974,362	94,800	4.8%

通过该模型可对伊犁州直未来三年第一产业生产总值进行预测, 2020 年该产业生产总值的点估计为:

$$y = -214058731.091 + 107047 \times 2020 \approx 2176209$$

2021 该产业生产总值的点估计为:

$$y = -214058731.091 + 107047 \times 2021 \approx 2283256$$

2022 该产业生产总值的点估计为:

$$y = -214058731.091 + 107047 \times 2022 \approx 2390303$$

根据以上分析数据计算出该产业在 2020、2021、2022 这三年的生产总值预测值分别为 2,176,209 万

元、2,283,256 万元、2,390,303 万元。

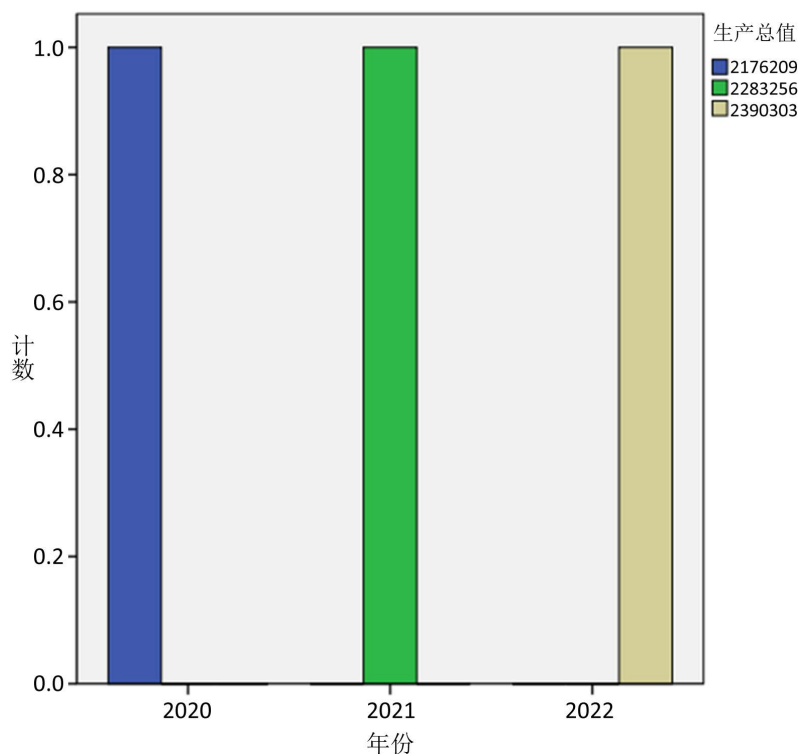


Figure 2. Histogram of GDP forecast of Yili direct primary industry in 2020~2022
图 2. 2020~2022 年伊犁州直第一产业生产总值预测柱形图

从图 2 可看出, 随着年份的推移, 伊犁州直第一产业的生产总值趋于上升趋势。

4. 相关建议及结语

本文利用线性回归的相关理论对伊犁州直 2020~2022 年第一产业生产总值进行预测, 结果看出该模型精度较高, 与实际值较为接近, 分析结果对伊犁州直第一产业发展有一定的参考价值。此外, 可看出伊犁州直第一产业生产总值近年呈现增长趋势, 对第一产业的可持续发展提出可参考的意见建议是很有必要的。

伊犁州直在第一产业上虽占有一定的优势, 但仍存在一些突出问题, 相关建议如下:

- 1) 第一产业就业人数需达标。根据产业结构的变化对第一产业就业人数进行调整, 避免不必要的人才浪费和人才缺失, 促使伊犁州直第一产业的稳定发展。
- 2) 发展主导产业、特色产业和新兴产业, 将互联网与该产业结合, 可降低交易成本、优化资源配置、提高劳动生产率, 同时通过智能化、物联化等手段实现产业跨越式发展。
- 3) 组建该产业科技服务团队, 重视该产业的人才建设, 可聘请专家在农闲时期组织农业技术人员集中学习国家农业政策以及专业农业技术。

基金项目

新疆理工学院校级人文社科项目(项目编号 SY202204)。

参考文献

- [1] 王黎明, 陈颖, 杨楠. 应用回归分析[M]. 上海: 复旦大学出版社, 2008: 14-25.

-
- [2] 邵鸿翔. 线性回归方法在数据挖掘中的应用和改进[J]. 统计与决策, 2012, 28(14): 76-80.
- [3] 虎辉, 马伟, 王岩. 突发公共卫生事件对宁夏社会经济的量化影响分析——基于一元线性回归分析[J]. 科技经济市场, 2021(9): 73-75.
- [4] 李秀兰. 一元线性回归参数的估计[J]. 山西大同大学学报(自然科学版), 2020, 36(6): 24-26.
- [5] 郭苗苗. 市场调查预测之一元线性回归分析方法[J]. 中外企业家, 2019, 35(13): 95-96.
- [6] 李久龙, 李玲, 刘瑞敏, 孙旭, 薛炜, 许腾, 杨懿. 一元线性回归分析方法在压力传感器数据拟合中的应用[J]. 计测技术, 2022, 42(2): 40-49.
- [7] 朱婉宁. 基于多元线性回归模型的安徽省地区生产总值影响因素分析[J]. 现代商业, 2022(17): 90-94.
- [8] 新疆维吾尔自治区统计局. 新疆统计年鉴[M]. 乌鲁木齐: 新疆维吾尔自治区统计局, 2009-2019.
- [9] 伊犁州统计局. 伊犁统计年鉴[M]. 伊宁: 伊犁州统计局, 2009-2019.