ISSN: 1813-4890

Quantitative Forecast of Land Based on Regression and SPSS

Bi He[1], Chuanzhi Geng[2]

[1]School of Civil Engineering, Shandong Jiaotong University, Jinan 250357, China;

[2] Beijing Sunway Technology Co., Ltd., Beijing 102206, China.

Abstract

Reasonable and orderly use of land resources is an important foundation for sustainable development. In order to use land in a reasonable and orderly manner, it is necessary to make quantitative predictions of land demand. In response to this demand, this article proposed to establish a forecast model based on historical data and regression methods. Firstly, three factors closely related to land use were selected as independent variables, and industrial land quantity was used as the dependent variable to construct a regression equation, after conducting corresponding accuracy tests on the equation, the future land use was quantitatively predicted by using the equation. The research results can provide effective references for the rational use of land.

Keywords

Quantitative forecast, land, regression, SPSS.

1. Introduction

Land is one of the most important natural resources for human living and economic development. According to different uses, government departments classify land into different types such as agricultural land, industrial land, commercial land, residential land, etc. Due to the limited availability of land resources, in order to achieve sustainable development, the government has strict regulations on the supply and use of land. In the process of land use management, predicting the quantity of land resources is the foundation for land management work. For quantitative prediction, it is an ideal method to establish regression equations using regression methods on the premise of rich historical data. This article is based on relevant historical data in Shandong Province to construct a regression equation for industrial land use, in order to predict the future industrial land use in Shandong Province.

2. Method and Data

2.1. Data Collection

Machine learning aims to find patterns behind things through historical data, and linear models are a method of describing the laws behind things. At the same time, most problems can also be represented by linear models Linear regression analysis is a statistical analysis method that determines the quantitative relationship between two or more variables that are interdependent. Essentially, this dependency relationship between variables is a type of linear correlation, which is the theoretical basis of linear regression models

For the industrial land use, through relevant research, we selected GDP (x1), fixed assets investment (x2) and industrial output value above designated size (x3) of Shandong Province from 2001 to 2015 as independent variables to predict the industrial land use (y) of Shandong Province. According to the statistical data published by the government over the years, we sorted out the relevant sample data, as shown in Table 1.

ISSN: 1813-4890

Table 1: Historical data on industrial land use and related indicators in Shandong Province

Year	Industrial land quantity(Km^2)	GDP(Billion CNY)	Fixed assets investment(Billion CNY)	Industrial output value above designated size(Billion CNY)
2001	363.5	9 438.31	2 807.79	9 377.372 6
2002	413.5	10 552.06	3 509.29	11 497.532 7
2003	472.6	12 435.93	5 328.44	16 192.752 4
2004	537.5	15 490.73	7 629.04	22 269.063 9
2005	603.0	18 516.87	10 541.87	31 102.982 4
2006	643.7	22 077.36	11 136.06	42 225.745 5
2007	690.7	25 965.91	11 136.06	55 437.921 1
2008	733.6	31 072.06	15 435.93	72 156.673 3
2009	734.4	33 896.65	19 030.97	91 004.449 3
2010	775.7	39 169.92	23 276.69	104 655.116 7
2011	801.1	45 361.85	26 769.73	119 338.229 7
2012	819.5	50 013.24	31 255.96	132 978.589 2
2013	807.9	54 684.33	36 789.07	148 058.361 2
2014	934.6	59 426.59	42 495.55	162 301.575 6
2015	1 044.1	63 002.33	48 312.46	174 474.193 7

2.2. Correlation test.

For regression analysis models, the correlation between independent and dependent variables. Reflected the degree of influence of independent variables on dependent variables, therefore, established regression analysis. The model needs to calculate the correlation between the independent variable and the dependent variable, when the correlation. Models only have practical significance when they are very high. Calculate this article based on basic data. The correlation coefficient score between y and x1, x2, and x3 among the various indicators adopted. The values are 0.9549, 0.9404, and 0.9485, all exceeding 0.9, indicating the dependent variable. There is a strong correlation with the independent variable, and the selection of indicators is reliable.

2.3. Regression model.

A multiple regression model refers to the selection of independent variables x1, x2 ,... xi that affect a dependent variable y to be studied, and the establishment of a multiple regression equation:

$$y_i = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_i \quad (1)$$

3. Results and Discuss

In this study, SPSS software was used as the solver computational tools. According to the multiple regression model used in this article and Table 1. Design and establish a prediction model for industrial land use in Shandong Province based on basic data preparation. The specific operation steps and calculation results are as follows:

ISSN: 1813-4890

3.1. Data input.

Input data into SPSS as the following:

Y=[363.5 413.5 472.6 537.5 603 643.7 690.7 733.6 734.4 775.7 801.1 819.5 807.9 934.6 1044.1]

X1=[9438.31 10552.06 12435.93 15490.73 18516.87 22077.36 25965.91 31072.06 33896.65 39169.92 45361.85 50013.24 54684.33 59426. 59 63002. 33]

X2=[2807.79 3509.29 5328.44 7629.04 10541.87 11136.06 12537.02 15435.93 19030.97 23276.69 26769.73 31255.96 36789.07 42495.55 48312.46]

X3=[9377.3726 11497.5327 16192.7524 22269.0639 31102.9824 42225.7455 55437.9211 72156.6733 91004.4493 104655.1167 119338.2297 132978.5892 148058. 3612 162301. 5756 174474. 1937]

After calculation, the calculation result returned by SPSS is:

b0＝228.78，b1＝0.029，b2＝－0.003，b3＝－0.005

Which means the regression equation is:

y=228.78+0.029x_1-〖0.003x〗_2-0.005x_3 (2)

Meanwhile, the variance interpretation rate of the regression results is R^2=0.92, indicating that the fitting degree of the regression results is very good.

After the regression model is established, the independent variable of 2016 can be substituted into the model to calculate the industrial land use in 2016, and the result is 1111.4 square kilometers.

4. Conclusion

The key to using a multiple regression model for prediction is whether the independent variable is closely related to the dependent variable and can largely determine the changes in the dependent variable. The use of statistics closely related to industrial land use in reality to predict the amount of construction site land has been shown to be feasible through practical calculations, and the results obtained have high reliability. However, it should also be noted that with the continuous improvement of factor productivity and the continuous optimization of land use efficiency, the growth of land output efficiency in the long run is not linear, but rather reflects exponential growth. Therefore, using nonlinear models for prediction over a longer period of time may achieve more approximate results.

Acknowledgements

Shandong Jiaotong University Natural Science Foundation(N0.20220406).

References

[1] MENG, WEI, HE, CHUAN, ZHOU, ZIHAN, et al. Application of the ridge regression in the back analysis of a virgin stress field[J]. Bulletin of engineering geology and the environment,2021,80(3):2215-2235.

[2] ILIC, IGOR, GORGULU, BERK, CEVIK, MUCAHIT, et al. Explainable boosted linear regression for time series forecasting[J]. Pattern Recognition: The Journal of the Pattern Recognition Society,2021,120.

[3] DU, XIAOXIAO, ZARE, ALINA. Multiple Instance Choquet Integral Classifier Fusion and Regression for Remote Sensing Applications[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019,57(5):2741-2753.

[4] SJUR WESTGAARD, STEIN-ERIK FLETEN, AHLMAHZ NEGASH, et al. Performing price scenario analysis and stress testing using quantile regression: A case study of the Californian electricity market[J]. Energy,2021,214(Jan.1 Pt.B):118796.1-118796.11.

ISSN: 1813-4890

[5] MOTTAGHITALAB, AMIRREZA, KHANJARI, AMIR, ALIZADEH, REZA, et al. Prediction of affinity coefficient for estimation of VOC adsorption on activated carbon using V-matrix regression method[J]. Adsorption,2021,27(6):963-978.

[6] ZHANGJING YANG, WENBO WANG, PU HUANG, et al. Local median based linear regression classification for biometric recognition[J]. 2021,96(Pt.B):107509-1-107509-13.

[7] DELGADO, JOSE, ROSA, JULIO, PELAEZ, JOSE A., et al. On the applicability of available regression models for estimating Newmark displacements for low to moderate magnitude earthquakes. The case of the Betic Cordillera (S Spain)[J]. Engineering Geology,2020,274.

[8] SONG, YONGCHEN, ZHOU, HANG, WANG, PENGFEI, et al. Prediction of clathrate hydrate phase equilibria using gradient boosted regression trees and deep neural networks[J]. The Journal of Chemical Thermodynamics,2019,13586-96.