ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

Forecast of Grain Yield in Anyang City in 2020

-- Application of Multivariate Return Analysis based on SPSS

Haoran Fu[1, a], Huahui Li[2, b, *]

[1]School of Economics, Anyang Normal University, Anyang 455000, China.

[2]School of Mathematics, Anyang University, Anyang 455000, China.

[a]hao3681@foxmail.com, [b]Corresponding author Email: blhh8287@163.com

Abstract

In this paper, the grain output in 2020 is predicted by multivariate statistical analysis, and some suggestions are put forward: while ensuring the sown area, rational application of chemical fertilizer, laying plastic film as needed, and recycling the plastic film.

Keywords

Multivariate return analysis; SPPS; Factor analysis; Grain yield forecast.

1. INTRODUCTION

In statistics, statistical analysis is often used to clearly show the interdependence between variables. return analysis is the most commonly used analysis method in statistical analysis, which can be divided into two categories, namely, unary return analysis and multivariate return analysis. Obviously, the unary return analysis has only one independent variable, while the multivariate return analysis has two or more explanatory variables. When things change in real life, this change is generally influenced not only by one factor, but by many or more factors. Therefore, these factors will make things in change uncertain, and there will be some unexpected changes. In general data analysis, correlation coefficient is usually used to describe the instability of things. When it is necessary to use one or more observations to estimate and predict the development and change of another thing, the correlation coefficient has no effect on the result. However, after a lot of practical investigations, we can get the relationship between them according to the investigation results. return analysis is a concrete description of the relationship between these variables. The most common problems in daily life generally have multiple independent variables, so they will be solved by multivariate return analysis. Multivariate linear return analysis mainly solves several problems, which are: clearly indicating whether there is a certain correlation between several selected variables, if there is a correlation, what kind of correlation is there, and expressing this correlation in a correct numerical way. After factor analysis of several independent variables, find out several independent variables that have the closest relationship with the value of dependent variables, and form the combination that has the most significant influence on dependent variables; In this paper, the yield of grain crops in Anyang City in 2020 will be predicted by multivariate return analysis.

Food is the foundation of human survival. The output of grain is related to the survival and development of a country and a region. For a populous country, food is even more indispensable. Anyang City is a prefecture-level city in the northeast of Henan Province, which is located in the plain and is the junction of Hebei, Shandong and Henan provinces. Moreover, Anyang, as an

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

important national commodity grain production base and one of the main grain, cotton and oil producing areas in Henan Province, has a better agricultural development prospect. According to the data, the annual output of grain crops in Anyang has been on the rise in recent nine years, and it has been on the rise from 2009 to 2015. In 2016, the grain output in Anyang decreased compared with the previous year, breaking the trend of continuous rise. This paper will analyze the collected factors that may affect the yield of food crops from 2009 to 2017, and screen out several factors that have the greatest impact on the yield of food crops through gradual Return, and then establish a mathematical model according to these factors; The yield of grain crops from 2018 to 2019 is predicted and tested, that is, the actual value is fitted with the predicted value, and after constant attempts and establishment of a suitable mathematical model, the total yield of grain crops in 2020 is predicted and corresponding suggestions are put forward.

2. PRINCIPLE NARRATION

2.1. Multiple Linear Regression Model

2.1.1. General form of multiple linear regression model

The equation that can express the relationship between the explained variable y, a set of explained variables $x_1, x_2, \cdots, x_p$ and random error term ε is called multivariate linear return model.

The general expression of multivariate linear model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon = \beta_0 + \sum \beta_i x_i + \varepsilon \tag{1}$$

In equation 1, β_0, β_1, $\beta_2, \cdots$, β_p is $p+1$ unknown parameters, where β_0 is called return constant and β_1, $\beta_2, \cdots$, β_p is called return coefficient. Y is called the interpreted variable, that is, the dependent variable, while $x_1, x_2, \cdots, x_p$ is p general variables that can be measured and controlled, called the interpreted variable, also called the independent variable, and ε is the random error.

If n groups of observation data $x_{i1}, x_{i2}, \cdots, x_{ip}$, y_i ($i = 1, 2, \cdots, n$) are obtained in practical problems, the model of the linear return can be expressed as:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n \end{cases} \tag{2}$$

The above model can also be written in matrix form:

$$Y = X\beta + \varepsilon \tag{3}$$

In equation 3:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

$$Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix},X=\begin{bmatrix}1&\cdots&x_{1p}\\\vdots&\ddots&\vdots\\1&\cdots&x_{np}\end{bmatrix},\beta=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_p\end{bmatrix},\varepsilon=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix} \tag{4}$$

What is called the return design matrix is the matrix X mentioned in equation 4 above.

2.1.2. Basic assumptions of multivariate linear return

In return analysis, it is necessary to put forward several basic assumptions for return Equation 1, which aims to make the parameter estimation of the model more convenient. The basic assumptions are:

The explanatory variable $x_1, x_2, \cdots, x_p$ is definite and non-random, and the rank of return design matrix X is required to be $p+1$, that is, $rank(X)=p+1<n$. In this hypothesis, the matrix X is full rank, and the number of explanatory variables p is less than the number of sample sizes n.

(2) The variance of each random error is the same, and the expected value of the random error term is 0, which can be expressed as:

$$\begin{cases} E(\varepsilon_i)=0,(i,j=1,2,\cdots,n) \\ Cov(\varepsilon_i,\varepsilon_j)=\begin{cases}\sigma^2, i=j\\ 0, i\neq j\end{cases},(i,j=1,2,\cdots,n)\end{cases} \tag{5}$$

This assumption is also called Gauss-Markov condition. When the observed value is assumed to have no systematic error, the term of random error is expected to be 0. The meaning of covariance of 0 indicates that there is no correlation between random error terms of two randomly selected samples. Each error term has the same precision, that is, covariance and variance are the same.

The random error term satisfies the normal distribution, namely:

$$\begin{cases} \varepsilon_i \sim N(0,\sigma^2), i=1,2,\cdots,n \\ \varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n \text{ mutual independence}\end{cases} \tag{6}$$

2.2. Estimation of Multiple Linear Regression Model

2.2.1. Estimation of return coefficient

Coefficient $\beta_0,\ \beta_1,\ \beta_2,\cdots,\ \beta_p$ of multivariate return equation is an unknown parameter which needs to be estimated according to sample data. The least square method is used to estimate the coefficient $\beta_0,\ \beta_1,\ \beta_2,\cdots,\ \beta_p$, and the estimated value $\hat\beta_0,\ \hat\beta_1,\ \hat\beta_2,\cdots,\ \hat\beta_p$ is calculated so that the deviation square sum:

$$Q(\beta_0,\ \beta_1,\ \beta_2,\cdots,\ \beta_p)=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots-\beta_px_{ip})^2 \tag{7}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

When the minimum value is obtained, it is found that $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2, \cdots$, $\hat{\beta}_p$ satisfies the equation:

$$Q\left(\hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\beta}_2,\cdots,\ \hat{\beta}_p\right) = \min_{\beta_0,\beta_1,\beta_2,\cdots,\beta_p} \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2 \tag{8}$$

Generally, the following equations are called empirical equations:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p \tag{9}$$

2.2.2. Least square estimation

In multivariate return model, there are some uncertain random variables, and the least square estimator of return coefficient is one of them. When the basic assumptions are met, the expected value of the least square estimator is equal to the true value of the overall return coefficient, namely:

$$E(\hat{\beta}) = \beta \tag{10}$$

2.2.3. Population variance's estimate

Among the numerous data in the multivariate return model, there are also unknown parameters. In the model, the variance of return coefficient and random error term belongs to unknown parameters, and the random error term is not the true value but the estimated value. The estimation method is to estimate by using the sum of squares of residuals and degrees of freedom. The specific mathematical form is as follows:

$$\hat{\sigma}^2 = \frac{1}{n-p-1} SSE = \frac{1}{n-p-1}(e'e) = \frac{1}{n-p-1}\sum_{i=1}^{n} e_i^2 \tag{11}$$

In which $\hat{\sigma}^2$ is an unbiased estimator of the variance σ^2 of the error term.

2.3. Test of Return Model

Before return analysis, it is necessary to test the selected explanatory variables to determine whether these independent variables meet the requirements of the established model. Firstly, the explanatory variables are tested for multicollinearity in SPSS. After using the linear return equation obtained from the independent variable after multicollinearity test, it is necessary to carry out T test on return coefficient and F test on return equation.

2.3.1. Multiple collinearity test

In the linear return model, the variables may not be completely independent, and sometimes there will be explanatory variables with strong correlation, which is multicollinearity among variables. When there is multicollinearity among variables, it will make the variance of parameter estimation increase, make this return model lose some authenticity, or make the predicted value less accurate.

In order to improve the accuracy of the return equation, it is necessary to eliminate the explanatory variables with multicollinearity step by step.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

2.3.2. Significance test of return coefficient (T test)

Testing the independent variable corresponding to each return coefficient is the main function of T test in multivariate return analysis. To judge whether these selected explanatory variables can really have a significant impact on the dependent variables. T-test is mainly used to select an optimal combination, a combination composed of some suitable independent variables, which will have the most significant impact on the dependent variables.

2.3.3. F test of significance of return equation

After the multivariate return model is initially formed, it is necessary to judge whether there is a significant linear relationship in the return model. By calculating the ratio between regression sum of squares and residual sum of squares, we can judge whether there is a significant linear relationship in the return model. After variance analysis, because the regression sum of squares and the residual sum of squares will change when the sample size of the observed value and the number of independent variables change, it is not stable, so the ratio of these two sum of squares needs to be tested by F.

2.4. Test of Goodness of Fit

In multivariate return analysis, the reciprocal of the ratio of the sum of squares of total dispersion to the sum of squares of regression is called the sample complex determination coefficient R^2. Because in the multivariate linear return model, the number of variables selected in each return model is not necessarily the same. Therefore, it is necessary to adjust the sum of squares of total deviation and the sum of squares of residual error with their own degrees of freedom. The adjusted determination coefficient is denoted as $R_a{}^2$, and the adjusted determination coefficient $R_a{}^2$ will be used instead of the sample complex determination coefficient R^2 in return analysis. The closer the value of $R_a{}^2$ is to 1, the better the fitting effect of the return model is, and the closer $R_a{}^2$ is to 0 or less, the worse the fitting effect is.

2.5. Multivariate Linear Return Model Prediction

When the return model passes various tests, it can be used for forecasting. The result of bringing a given set of independent variables into return equation is the prediction result. The 95% and 99% confidence intervals of the dependent variable prediction results are:

$$(\hat{y}_0 - 2\hat{\sigma}, \hat{y}_0 + 2\hat{\sigma})\text{,}\quad (\hat{y}_0 - 3\hat{\sigma}, \hat{y}_0 + 3\hat{\sigma})$$

3. FORECAST OF GRAIN YIELD IN ANYANG CITY

3.1. Analysis of Factors Affecting Grain Crop Yield

In actual planting, there are many factors affecting grain yield, and the types are complex. According to the relevant literature and the analysis of the actual situation, this paper selects 10 factors that may affect the grain crop yield in Puyang, which are: Rural labor force (10,000 people), sown area of grain crops (1,000 hectares), grain yield per unit area (kg/mu), total power of agricultural machinery (10,000 kilowatts), effective irrigation area (1,000 hectares), agricultural fertilizer usage (tons), plastic film coverage area (hectares), pesticide usage (tons), agricultural intermediate consumption (10,000 yuan) and annual average precipitation (millimeters) In this paper, SPSS 25 is used to analyze the data of the 10 factors mentioned above.

3.1.1. Changes of rural labor force

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

From 2009 to 2017, the number of rural laborers in Anyang fluctuated greatly, and there was no obvious trend (Table 1). From 2012 to 2015, the number of rural laborers in Anyang City has been declining continuously, which is related to the fact that most of the young and middle-aged rural laborers choose to give up farming and go to cities to find jobs with higher incomes. In recent years, the number of rural laborers has increased, which is related to the relevant preferential policies. Using SPSS 25 to analyze the data of Anyang rural labor force from 2009 to 2017, the return equation is obtained:

$$y = 636.474 - 0.215x$$

3.1.2. Change of sown area of grain crops

The sown area of grain crops in Anyang showed an upward trend in the nine years from 2009 to 2017 (Table 1). This trend is related to a series of policies to benefit farmers to protect crop production in recent years. These preferential policies have stimulated farmers' enthusiasm for farming to a certain extent. At the same time, some relocation policies of "changing houses into cultivated land" have also increased the land area, greatly improved the land utilization rate and increased the sown area of food crops. return equation obtained by data analysis in SPSS 25:

$$y = -11250.036 + 5.785x$$

3.1.3. Grain yield change

During the nine years from 2009 to 2017, the grain yield per unit area in Anyang showed an overall upward trend with little fluctuation (Table 1). In recent years, with the development and progress of science and technology and the popularization of various efficient agricultural production methods, the per unit area yield of grain in Anyang has increased steadily. The return equation obtained by analyzing the data with SPSS 25 is:

$$y = -5466.244 + 2.933x$$

3.1.4. Total power change of agricultural machinery

2015 is a turning point in the change of total power of agricultural machinery in Anyang during the nine years from 2009 to 2017. From 2009 to 2015, the total power of rural machinery continued to rise, but from 2015 to 2017 when the data were collected, there was a slight decline (Table 1). Although agricultural machinery improves farming efficiency, in recent years, the increase of its use cost is also a reason why some farmers give up using machinery to assist farming. The return equation obtained by SPSS 25 analysis is:

$$y = 2211.732 - 0.89x$$

3.1.5. Change of effective irrigation area

From 2009 to 2017, when the data were collected, the effective irrigation area in Anyang City fluctuated little during the nine years, and the image showed an obvious upward trend (Table 1). Although Anyang is a city adjacent to the Yellow River, the cultivated land is not only irrigated by water from the Yellow River, but some farmers will choose to drill their own wells to improve irrigation efficiency and increase irrigation area. The return equation obtained after data analysis by SPSS 25 is:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

$$y = -2225.472 + 1.216x$$

3.1.6. Changes in the use of agricultural chemical fertilizers

From 2009 to 2017, the use of agricultural chemical fertilizers in Anyang showed an overall upward trend (Table 1). There are many kinds of agricultural chemical fertilizers. The main function of most agricultural chemical fertilizers is to improve the ability of crops to resist drought and cold to a certain extent. Reasonable application of nitrogen fertilizer can also promote the growth of food crops and improve the nutritional value of crops to a certain extent. Moreover, controlling the use of chemical fertilizers will increase the grain yield. Through the data analysis of SPSS 25, the return equation is as follows:

$$y = -4341315.683 + 2289.9x$$

3.1.7. Change of plastic film coverage area

Agricultural plastic film has the functions of raising ground temperature, maintaining water and soil, preventing drought and waterlogging, suppressing salt and protecting seedlings. The most common use of plastic film is to plant out-of-season crops when the weather gets cold, which mainly plays the role of heat preservation, thus improving the output quality of crops and being loved by farmers. This is also the reason why, although Anyang is located in the Central Plains where the average temperature is not low from 2009 to 2017, the area covered by plastic film is generally on the rise (Table 1). return equation is obtained after data analysis in SPSS 25:

$$y = -1158992.472 + 591.917x$$

3.1.8. Changes in pesticide usage

From 2009 to 2017, the amount of pesticides used in Anyang City tends to be stable but has a downward trend (Table 1). Pesticide is a kind of chemical agent with many kinds. In the daily agricultural production process, herbicides and plant growth regulators are the most used pesticides. Although controlling the amount of pesticides can effectively improve the quality and yield of grain, with the development and progress of society and the continuous improvement of people's living requirements, the use of pesticides will gradually decrease. The return equation obtained by analysis in SPSS 25 is:

$$y = 187275.267 - 90.933x$$

3.1.9. Change of agricultural intermediate consumption

Material consumption and production service expenditure are collectively referred to as agricultural intermediate consumption, in which material consumption mainly includes the purchase cost of seeds, fertilizers, pesticides, plastic films, etc. From 2009 to 2017, the intermediate consumption of agriculture in Anyang continued to increase and the trend was stable (Table 1). In recent years, although the sown area of grain has not increased by a large margin, the intermediate consumption of agriculture has also risen with the price increase, and the consumption cost has risen steadily. The return equation obtained by analyzing the data of SPSS 25 is:

$$y = -69299485.2 + 34703.283x$$

3.1.10. Variation of annual average precipitation

The annual average precipitation in Anyang has a stable overall change. From 2009 to 2017 (Table 1), there is a downward trend. In some years, the annual average precipitation is more or less than the annual average precipitation due to climatic reasons. After data analysis by SPSS 25, the return equation is as follows:

$$y = 2018.613 - 0.01x$$

Table 1. Influence factor

Year	Total output unit of grain crops: (tons)	Rural labor force (unit: 10,000 people)	Sown area of grain crops (unit: 1000 hectares)	Grain yield unit: (kg/mu)	Total power of agricultural machinery (unit: 10,000 kilowatts)	Effective irrigation area (unit: 1000 hectares)	Amount of agricultural chemical fertilizer used (unit: tons)	Covered area of plastic film (unit: hectares)	Amount of pesticides used (unit: tons)	Agricultural intermediate consumption (unit: 10,000 yuan)	Average annual precipitation (unit: mm)
2009	2687516	200.8	376.4	430	399.1	217.86	258800	28679	4388	406559	564.2
2010	2714573	206.5	379.75	430	407.407	217.52	258880	30368	4435	462602	711.6
2011	2735335	206	382.93	430	418.36	218.5	261232	31347	4493	500659	588.8
2012	2805443	206.9	387.91	437	432.9	220.44	263537	32574	4331	528400	389.9
2013	2822065	202.46	389.61	438	442.16	222.99	276262	35502	4396.2	563446	445.3
2014	2849408	204.08	393.2	435	450.91	224.57	277063	34466	4376	592699	516.5
2015	2905050	200.27	395.14	445	458.11	224.57	275730	33069	4126	601200	550.6
2016	2831644	200.32	419.94	450	409.06	225.52	268001	31269	3882	642950	550
2017	2883119	205.78	425.6	452	360.13	226.03	275678	35548	3611	725502	499.5

Data source: Statistical Yearbook of Anyang Municipal Bureau of Statistics 2009-2017, Water Resources Bulletin of Anyang Water Conservancy Bureau 2009-2017

3.2. Establish A Statistical Model

Establish grain crop yield y and rural labor force (10,000 people) x_1, grain crop sown area (1,000 hectares) x_2, grain yield (kg/mu) x_3, total power of agricultural machinery (10,000 kW) x_4, effective actual irrigation area (1,000 hectares) A multiple linear regression of x_5, agricultural fertilizer use (tons) x_6, plastic film coverage (hectares) x_7, pesticide use (tons) x_8, agricultural intermediate consumption (10,000 yuan) x_9, and annual average precipitation (mm) x_{10} model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10}$$

Where β_0, β_1, $\beta_2, \cdots,$ β_{10} is the parameter to be determined.

3.3. Correlation Analysis

In this experiment, 10 independent variables x_i ($i = 1, 2, \cdots, 10$) were selected, which may affect the dependent grain crop yield y. These independent variables are mainly selected according to relevant literature and factual experience. Therefore, in this paper, not every change of independent variables will have a significant impact on the dependent variables.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

Therefore, a correlation analysis is carried out at first, so as to eliminate some independent variables with low correlation with the dependent variables and improve the fitting degree of return equation. The relevant analysis results are as follows:

Table 2. Correlation analysis table

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
y	1	-.207	.700*	.810**	.234	.926**	.895**	.786*	-.632	.884**	-.479
x1	-.207	1	-.143	-.302	-.255	-.344	-.263	.159	.173	-.063	-.002
x2	.700*	-.143	1	.942**	-.423	.846**	.551	.494	-.946**	.929**	-.244
x3	.810**	-.302	.942**	1	-.253	.879**	.622	.511	-.934**	.910**	-.334
x4	.234	-.255	-.423	-.253	1	.058	.291	.156	.554	-.172	-.217
x5	.926**	-.344	.846**	.879**	.058	1	.880**	.720*	-.743*	.933**	-.424
x6	.895**	-.263	.551	.622	.291	.880**	1	.886**	-.445	.779*	-.461
x7	.786*	.159	.494	.511	.156	.720*	.886**	1	-.368	.747*	-.574
x8	-.632	.173	-.946**	-.934**	.554	-.743*	-.445	-.368	1	-.840**	.175
x9	.884**	-.063	.929**	.910**	-.172	.933**	.779*	.747*	-.840**	1	-.359
x10	-.479	-.002	-.244	-.334	-.217	-.424	-.461	-.574	.175	-.359	1

It can be seen from the correlation matrix that the yield of grain crops y and the sown area of grain crops (thousand hectares) x_2, grain yield (kg/mu) x_3, effective actual irrigation area (thousand hectares) x_5, agricultural fertilizer usage (tons) x_6, plastic film The absolute value of the correlation coefficients of coverage area (hectares) x_7, pesticide use (tons) x_8, and agricultural intermediate consumption (10,000 yuan) x_9 are all above 0.6. It can be explained that the selected x_2, x_3, x_5, x_6, x_7, x_8, x_9 have a high correlation with the explained variable, and the correlation coefficient between x_1 and the dependent variable y is -0.207, x_4 and the dependent variable The correlation coefficient of y is 0.234, and the correlation coefficient of x_{10} and dependent variable y is -0.497. The correlation between these three explanatory variables and the dependent variable is weak, so x_1, x_4, and x_{10} are eliminated.

3.4. Multiple Collinearity Analysis

After the independent variables x_1, x_4 and x_{10} are eliminated from the correlation matrix results, the remaining data are analyzed by return based on SPSS 25.

Table 3. Coefficient table

	Unstandardized coefficient		Normalization coefficient	t	Significant	B's 95.0% confidence interval		Collinearity statistics	
	B	Standard error	Beta			Lower limit	Upper limit	Tolerance	VIF
(Constant)	197796.353	223749.498		.884	.539	-2645210.579	3040803.284		
x2	-9641.221	377.596	-2.187	-25.533	.025	-14439.037	-4843.404	.014	72.015
x3	3606.300	444.011	.408	8.122	.078	-2035.390	9247.990	.040	24.789
x5	28503.856	2283.493	1.295	12.483	.051	-510.680	57518.391	.009	105.571
x6	-6.942	.790	-.726	-8.791	.072	-16.976	3.092	.015	66.847
x7	3.354	1.423	.104	2.358	.255	-14.722	21.430	.052	19.241
x8	-96.002	14.024	-.378	-6.846	.092	-274.194	82.190	.033	29.860
x9	1.182	.056	1.506	21.103	.030	.470	1.893	.020	49.968

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

It can be seen from the above coefficient table that the tolerance value of each explanatory variable is less than 0.1 and the VIF value is large. These data show that there is a correlation between the selected independent variables, that is, there is multicollinearity among these explanatory variables, and basically all of them are serious multicollinearity. Then the next step should be to start with the independent variable with the largest VIF value, eliminate these independent variables with multicollinearity, and then carry out return analysis after eliminating each independent variable. Until the independent variables with multicollinearity are independent of each other, or there is only low and reasonable multicollinearity among the variables. According to the VIF value from big to small, after removing x_5, x_9, x_3 and x_8 step by step, the final return coefficient table is as follows:

Table 4. Return coefficient table

	Unstandardized coefficient		Normalization coefficient	t	Significant	B's 95.0% confidence interval		Collinearity statistics	
	B	Standard error	Beta			Lower limit	Upper limit	Tolerance	VIF
(Constant)	358264.126	615322.480		.582	.586	-1223472.666	1940000.917		
x2	1309.809	874.150	.297	1.498	.194	-937.265	3556.884	.696	1.437
x6	7.349	3.564	.768	2.062	.094	-1.811	16.510	.197	5.070
x7	-1.334	11.480	-.042	-.116	.912	-30.843	28.176	.214	4.666

According to the return coefficient table, it can be known that the explanatory variables that the tolerance value and VIF value meet the condition that there is no multicollinearity are: sown area of grain crops (thousands of hectares) x_2, the amount of agricultural chemical fertilizer used (tons) x_6, and the area covered by plastic film (hectares) x_7.

3.5. Test of Return Equation

Table 5. Model summary table

model	R	R square	Adjusted R square	Error in standard estimation	Change statistics				
					R-square variation	F variation	Variance 1	Variance 2	Significant f variation
1	.929a	.863	.781	35443.262	.863	10.514	3	5	.013

According to the total number table of models, it can be seen that the complex correlation coefficient, decision coefficient and adjusted decision coefficient are all greater than 0.7, and the standard error is 35443.262, so it can be preliminarily judged that the model has a high degree of fitting.

Table 6. Analysis of variance table

	Sum of squares	Variance	Mean square	F	Significant
Regression	39624511897.893	3	13208170632.631	10.514	.013b
Residual	6281124201.663	5	1256224840.333		
Total	45905636099.556	8			

According to the above variance analysis table, F=10.514, p value is 0.013. The f value and p value show that the whole return equation is valid. That is to say, from 2009 to 2017, the linear relationship between the sown area of grain crops (thousands of hectares), the amount of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

agricultural chemical fertilizer used (tons) and the area covered by plastic film (hectares) on the yield of grain crops in Anyang City is significant.

The collected grain output in 2018 and 2019 is also put into the data, and the prediction results of return equation and the real values are used to make a multi-line line chart. It can be seen from the figure that the fitting degree of the model is relatively high, and the line chart is as follows:

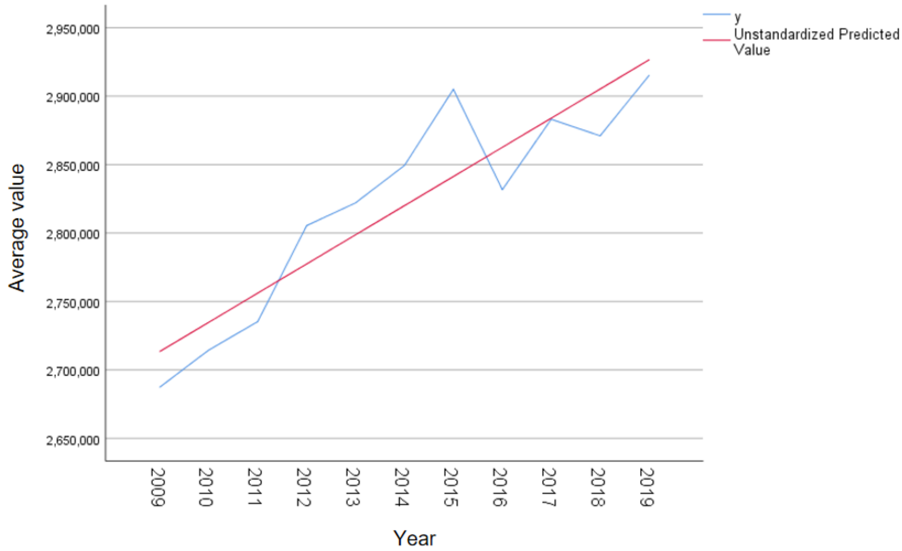

Figure 1. Fitting line chart

4. RESULT ANALYSIS

4.1. Analysis of Return Equation

Table 7. Coefficient table

	Unstandardized coefficient		Normalization coefficient	t	Significant	B's 95.0% confidence interval	
	B	Standard error	Beta			Lower limit	Upper limit
(Constant)	358264.126	615322.480		.582	.586	-1223472.666	1940000.917
x2	1309.809	874.150	.297	1.498	.194	-937.265	3556.884
x6	7.349	3.564	.768	2.062	.094	-1.811	16.510
x7	-1.334	11.480	-.042	-.116	.912	-30.843	28.176

From the above coefficient table, it can be concluded that the return equation of this model is:

$$\hat{y} = 358264.126 + 1309.809x_2 + 7.349x_6 - 1.334x_7$$

According to the return equation, we can know the influence degree of the finally selected factors on the explained variables. When the amount of agricultural chemical fertilizer and the area covered by plastic film remain unchanged, the grain crop yield in Anyang will increase by an average of 1309.809 units with each increase in the sown area of grain crops. The local film coverage area and the sown area of food crops remain unchanged, and the output of food crops will increase by an average of 7.349 units for each additional unit of agricultural chemical fertilizer use; When the sown area of grain crops and the amount of chemical fertilizer used in agriculture remain unchanged, the yield of grain crops in Anyang will decrease by 1.334 units on average for every unit increase of plastic film coverage area.

The 95% confidence interval of constant β_0 is $[-1223472.666, 1940000.917]$; The 95% confidence interval of return coefficient β_2 is $[-937.265, 3556.884]$; The 95% confidence interval of β_6 is $[-1.811, 16.51]$; The 95% confidence interval of β_7 is $[-30.843, 28.176]$.

4.2. Forecast Result

In SPSS 25, the year 2020 is put into the established multivariate return model for return analysis, and the predicted result can be obtained: the grain crop yield in Anyang in 2020 is 2947788.47273 tons.

5. SUGGESTIONS

5.1. Ensure the Sown Area of Grain

According to the return coefficient, it can be concluded that in the return model established in this paper, the independent variable that has the greatest influence on the grain crop yield is the sown area of the grain crop, and it has a strong positive correlation with the grain yield, which shows the mathematical significance that the larger the sown area, the more the yield. However, it is impractical to increase the sown area of grain crops for the sake of the rapid growth of grain output at present in Anyang City, so it is necessary to ensure the existing sown area of grain, ensure the grain output to the maximum extent, and try to maintain the slow growth rate of the existing grain output.

5.2. Rational Use of Agricultural Chemical Fertilizers

In the return model in this paper, the amount of agricultural chemical fertilizer used is the second most important factor, and the use of agricultural chemical fertilizer can increase the grain yield to a certain extent. Although agricultural chemical fertilizer is a kind of fertilizer which provides nutrition for crops through physical or chemical distribution, excessive use of agricultural chemical fertilizer will also lead to the death of crops. Therefore, the rational use of agricultural chemical fertilizers in the production process can maximize the effective role of chemical fertilizers, provide sufficient nutrition for food crops, ensure the quality of food, and help the increase of food output to the greatest extent.

5.3. Laying Mulch Film and Recycling It as Needed

Although there is a negative correlation between the area covered by plastic film and the yield of food crops in the return model in this paper, it is undeniable that in some areas or in the production process of some food crops, plastic film has a positive impact on the protection of food crops and the yield of food crops. The common agricultural mulch film is usually a kind of non-degradable plastic film, which damages the soil quality to a certain extent. Therefore, we should minimize the use of plastic film, lay plastic film according to the demand, and recycle the plastic film to protect the growing environment of food crops and lay a good "hardware" foundation for increasing the yield of food crops.

REFERENCES

[1] Xiang Shujian, Zhang Xueyi. Statistics. China Statistics Press, July 2015.

[2] Zhou Yongsheng, Xiao Yuhuan, Huang Runsheng. Guangxi Grain Yield Forecast Based on Multiple Linear Regression. Journal of Southern Agriculture, No. 9, 2011.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0017

[3] Liu Hongjun. Forecast analysis and recommendations of grain production in Neijiang City. Food Science and Technology and Economy, Issue 1, 2017.

[4] Zhang Xiaoling, Rao Zhijian. Food production forecast in Yunnan Province-Based on multiple linear regression analysis. Modern Business, Issue 4, 2017.

[5] Anyang Municipal Bureau of Statistics 2009-2017 Statistical Yearbook.

[6] Anyang Water Resources Bureau 2009-2017 Water Resources Bulletin.