Bangladesh J. Bot. **51**(4): 979-985, 2022 (December) Special DOI: https://doi.org/10.3329/bjb.v51i40.63842

MULTIPLE LINEAR REGRESSION AND CORRELATION ANALYSIS OF YIELD OF SUMMER MAIZE VARIETIES IN SHAANXI PROVINCE, CHINA

QILI HAO*

Shangluo Branch of Shaanxi Land Engineering Construction Group Co., Ltd.
Shaanxi Xi'an 710075 China

Keywords: Maize varieties, Yield, Multiple linear regression, Correlation analysis

Abstract

The main factors affecting the yield of mid-ripening maize varieties were clarified to provide reference for breeding and production of summer maize varieties in Guanzhong Plain, Shaanxi Province. According to the relevant agronomic trait data of 11 new maize varieties planted in the Guanzhong Plain summer maize regional trial introduced in Shaanxi Province in 2019, analysis of the main characters of the new maize varieties were carried out using the multiple linear regression method. It can provide reference for the selection of maize agronomic traits. The agronomic traits that were significantly related to yield of summer maize varieties in Shaanxi Province were 100-grain [1]weight, number of grains in row, seed rate, empty stalk rate and ear length, etc. When selecting maize varieties, under the premise of ensuring proper 100-grain weight and ear length, one should pay attention to the selection of varieties with high seed rate and low empty stalk rate. In the breeding of high-yielding maize varieties, lodging resistance should be enhanced, the rate of empty stalks should be reduced, the rate of seed production should be increased, 100-grain weight and ear diameter should be appropriately increased, and ear height should be reduced.

Introduction

With the rapid development of science, technology and economy, the demand for corn in China is also increasing year by year (Liu *et al.* 2017, Axel *et al.* 2021, Singh 2021). With the increasing uncertainty of international factors such as geopolitics, countries and people are paying more and more attention to food security. The conflict between the two major grain-producing countries, Russia and Ukraine, has further increased the country's sense of worry about grain production (Xiang *et al.* 2020). Yield remains an important goal pursued by corn breeders, national and individual producers at this stage. Shaanxi Province is the main summer corn producing area of China and a commercial grain production base (Ali *et al.* 2019). The selection of suitable crop varieties is of great significance for the promotion of crop planting, the increase of grain production and the increase of farmers' income (Qiu *et al.* 2015, Nafziger and Srinivasan 2021, Shahhosseini 2021). Shaanxi Province has a vast territory. Beishan and Qinling Mountains divide Shaanxi Province into three natural regions: the northern Shaanxi Plateau, the central Guanzhong Plain, and the southern Qinba Mountains (Nascimento 2020, Yue *et al.* 2021). The Guanzhong Plain has been known as a granary since ancient times (He and Zhou 2016, Forte *et al.* 2017, Ruis 2021, Simão and Johnston 2021, Lu *et al.* 2022). In the process of maize planting, it is necessary to study the main agronomic traits and their correlation with yield, to clarify the main factors affecting yield, and to provide a basis for selecting superior hybrids (Fernando 2020, Hou *et al.* 2020, Tao 2020). Analysis of the correlation between the main agronomic traits and yield of 11 maize varieties introduced in Guanzhong Plain, Shaanxi Province in 2019, would provide a reference for the selection of maize agronomic traits, and promote suitable planting varieties, increase crop yield, provide strong support to increase farmers' income.

*Author for correspondence: <64118717@qq.com>.

Materials and Methods

The experiment was set up in Ducun Town, Fuping County, Weinan City. The area has a semi-arid continental climate, and the rainfall distribution is extremely uneven throughout the year, mostly concentrated in July to September, accounting for 49% of the annual rainfall, and other seasons are relatively dry. The dry and wet seasons are distinct, and the dry season is longer than the wet season. Especially in spring, it is windy and rainy, and the evaporation is large. The annual evaporation is 1000-1300 mm, which is 2.0 to 2.3 times of the rainfall. The maximum evaporation (189.5 mm) was in June, the smallest (44. 9 mm) was in December. The frost-free period is 225 days, and the annual average temperature is 13. 4°C, the highest temperature in summer is 41.8 ℃, the lowest temperature in winter - 22 ~ -10°C, the regional conditions are suitable for the cultivation of crops such as corn and wheat.

The experimental materials 11 corn varieties are suitable for the Guanzhong summer planting area in the sixth batch of application for introduction of corn varieties in Shaanxi Province in 2019: AnNong 591, NongDan476, JinAo608, Wo Feng 9, Zheng Dan5176, Rui Feng 168, Zheng Da 1473, BoFa707, Xi Wang 3088, XianYu1653, XianYu1568.

Corn planting density was 60 000~67 500 plants/hm^2, randomly block arranged, in 5 rows of blocks, 6 m row length, and the middle 3 rows of actual receipts in each plot were used to calculate the yield. The experiment was repeated three times. The eight main agronomic traits selected for the study were ear length (cm), ear diameter (cm), bald tip length (cm), number of grains per row (grain/row), weight of 100-grains (g), and empty culm rate (%), inversion rate (%) and seeding rate (%).

According to the requirements of multiple linear regression theory, the yield of 11 corn hybrid combinations and 8 related agronomic traits were subjected to regression analysis, and t-test was used for significance test. Data analysis were performed using the R statistical programming language.

Results and Discussion

Since the dimensions of the inspection indicators are inconsistent, it is necessary to initialize the original data to make it dimensionless, so as to facilitate the comparison of data between different dimensions. The Z-Score standardization method was used to standardize the data. Applicable to large samples of more than 30, the mean of the standardized indicators is 0, and the variance is 1. The formula is as follows:

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} (i = 1,2,\cdots,n;\ j = 1,2,\cdots m) \tag{1}$$

Among them, $\bar{x}_j$ is the average value of index *j*, s_j is the variance of index *j*.

In order to investigate the dependence between the yield and agronomic traits of different corn hybrid varieties, it is necessary to establish a multiple linear regression model.

Suppose there is a linear relationship between variable Y and variables X_1, X_2, X_3, •••, X_p as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p + \varepsilon \tag{2}$$

Among them, β_0 is the regression constant, β_1, β_2, β_3•••, β_8 are the overall regression parameters. When p=1, formula (2) is called a univariate linear regression model, and when $p \geq 2$, it is called a multiple linear regression model. ε is a random error and obeys the distribution of $\varepsilon \sim N(0,\sigma^2)$.

Table 1. Normalization results of raw data.

Tested varieties	Yield	Ear length	Ear thickness	Ear tip-barrenness	Number of grains	100-grain weight	Empty stalk rate	Inversion rate	Seed yield
AnNong 591	0.042	1.222	2.083	0.498	0.909	0.392	0.278	6.354	-0.280
NongDan476	0.044	0.047	3.819	-2.488	-0.909	0.283	-1.667	-4.696	-1.212
JinAo608	0.026	0.047	5.556	-2.488	0	-0.215	0	-0.552	1.119
WoFeng 9	-0.032	0.987	2.083	-2.488	0	-0.121	3.889	-3.315	-1.678
Zheng Da5176	-0.004	2.162	-8.333	-1.493	0.909	-0.261	-0.833	3.591	-0.435
Rui Feng 168	0.034	0.517	-1.389	6.468	0.909	-0.775	-2.222	-1.934	0.031
Zheng Da 1473	0.063	1.222	3.819	3.483	0	-0.246	-0.278	-3.315	0.653
BoFa707	-0.028	-3.008	-3.125	-4.478	-0.909	-0.012	0.278	3.591	2.517
XiWang 3088	-0.028	-1.128	0.347	1.493	-0.455	0.283	-1.111	3.591	-1.523
XianYu1653	0.026	0.047	5.556	-2.488	0	-0.215	0	-0.552	1.119
XianYu1568	0.037	0.041	2.612	2.221	-1.234	-0.351	0.575	-0.513	2.312

The estimation method of parameter β adopts the least square estimation method, and its objective function is to minimize:

$$Q(\beta) = \sum_{i=1}^{n} \left\| y_i - x_i \beta \right\| \tag{3}$$

Among them, Y stands for yield, X_1 stands for 100-grain weight, X_2 stands for the number of grains in row, X_3 stands for seed rate, X_4 stands for empty stalk rate, X_5 stands for ear length, X_6 stand for ear tip-barrenness, X_7 stands for ear diameter, and X_8 stands for fold rate. When getting n sets of observation data (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, y_i) (i=1,2,3,…, n), The linear regression model can be expressed as

$$\begin{cases} y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \beta_3 X_{13} + \cdots + \beta_8 X_{18} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \beta_3 X_{23} + \cdots + \beta_8 X_{28} + \varepsilon_2 \\ \bullet\bullet\bullet\bullet\bullet\bullet \\ y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \beta_3 X_{n3} + \cdots + \beta_8 X_{n8} + \varepsilon_n \end{cases} \tag{4}$$

It is written in matrix form as $Y = X\beta + \varepsilon$, Usually X is called the design matrix, β is the regression coefficient vector.

From the established multiple regression model and the regression coefficients that have been obtained, the fitting test of the entire regression equation was carried out, and the R^2 test was used. Goodness of fit was calculated based on decomposing the square of the total deviation. The formula for calculating the sum of squares of the total deviation is:

$$SST = SSE + SSR \quad (5)$$

Among them, SSE is the residual sum of squares, SSR is the regression sum of squares, and SST is the total squared deviation. Calculated as follows:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (6)$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \quad (7)$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (8)$$

Among them, $\bar{y}$ is the mean value of the observed value of the sample, $\hat{y}$ is the estimated value, and the coefficient of determination R2 is the ratio of the sum of squares of the regression to the sum of squares of the total deviation. The calculation formula is:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (9)$$

R^2 reflects the goodness of fit of the regression line to the data, and its value is between [0,1]. The closer R^2 is to 1, the better the fitting result of the regression equation, and the closer R^2 is to 0, the worse the fitting of the regression equation.

The t-test is to test whether each regression coefficient in the regression model is significant, so that only those factors that have a significant impact on the dependent variable are retained in the model. When testing, first calculate the statistic t_1, and then look up the t distribution table according to the given significant level α and the degrees of freedom n-k-1 to obtain the critical value tα or $t_{\alpha/2}$, if t > t - α or $t_{\alpha/2}$, then the regression coefficient. There is a significant difference between β_i and 0, on the contrary, there is no significant difference. The calculation formula is:

$$t = \frac{\bar{x} - \mu}{\sigma_x / \sqrt{n-1}} \quad (10)$$

where $\bar{x}$ is the sample mean， μ is the population mean， σ_x is the sample standard deviation and *n* is the sample size.

Using the data in Table 1 to solve the reference factors, the results in Table 2 can be obtained.

Table 2. Model summary.

Model	R	R Square	Adjusted R Square	Std. Error of the estimate	Change Statistics					Durbin-Watson
					R-squared change	F value change	df1	df2	Significant F value change	
1	0.992	0.984	0.946	0.0182	0.984	15.461	8	2	0.0013	2.138

Table 2 showed how well the model fits. As can be seen from the table, the complex correlation coefficient of the model is 0.992, the determination coefficient is 0.984, the adjusted coefficient of determination was 0.946 and the standard error of the estimated value was 0.0182, and the Durbin-Watson test statistic was 2.138. When DW≈2, the residuals were independent.

Table 3. ANOVA.

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.011	8	0.001	15.462	0.0013
Residual	0.000	2	0.000		
Total	0.011	10			

Table 3 showed ANOVA results for the model. It was observed that the value of the F statistic corresponding to the model was 15.462, and the p value was 0.0013. When the significance level is 0.05, it can be considered that there is a linear relationship between crop yield and different traits.

Table 4. Parameter estimation results of multiple linear regression.

Regression coefficients	Parameter estimates	Standard error	*t*	*p*-value
$\beta 0$	33.675	3.8724	3.527	0.0487
$\beta 1$	0.1891	0.0137	3.172	<0.001
$\beta 2$	0.0321	0.0021	1.026	<0.001
$\beta 3$	0.0328	0.0073	2.186	<0.001
$\beta 4$	0.2157	0.0342	1.823	<0.001
$\beta 5$	0.1274	0.0225	3.524	<0.001
$\beta 6$	0.3251	0.0673	-3.434	<0.001
$\beta 7$	0.1325	0.0101	-3.052	0.0321
$\beta 8$	0.0132	0.0031	4.871	0.0211

According to the *P* value results presented in the table, it can be found that the parameters β_0, β_7, β_8 were significant at the 95% confidence level. Parameters β_1, β_2, β_3, β_4, β_5, β_6 are significant at the 99 per cent confidence level. The results showed that among the 8 main agronomic traits selected, factors such as 100-grain weight, number of grains per row, seed rate, empty stalk rate, ear length and other factors had a significant impact on yield, and were closely related to yield. Through the analysis, the multiple linear regression model about the yield was obtained.

$$Y = 0.1891X_1 + 0.0321X_2 + 0.0328X_3 + 0.2157X_4 + 0.1274X_5 + 0.3251X_6 + 0.1325X_7 + 0.0132X_8 + 33.675 \tag{11}$$

By calculating the correlation coefficient between yield and 8 main agronomic traits, it was found that the correlation between different traits and yield was quite different. The correlation between 100-kernel weight and yield exceeded 0.9, the strongest correlation. The number of grains per row and seedling emergence rate in these two traits have a strong correlation with yield, ranging from 0.8 to 0.9, while the correlation between empty culm rate, ear length, bald length, ear diameter and fold rate and yield was less than or equal to close to 0.8, indicating relatively weak correlations between these traits and yield.

Table 5. Correlation coefficients between yield and different traits of maize varieties.

Tested varieties	Yield	Ear length	Ear thickness	Ear tip-barrenness	Number of grains	100-grain weight	Empty stalk rate	Inversion rate	Seed yield
AnNong 591	0.78	0.872	0.902	0.728	0.923	0.947	0.798	0.729	0.88
NongDan476	1	0.825	0.622	0.614	0.945	0.909	0.867	0.769	1
JinAo608	0.994	0.83	0.624	0.795	0.946	0.995	0.879	0.793	0.996
WoFeng 9	0.804	0.864	0.63	0.793	0.98	0.915	0.86	0.717	0.803
Zheng Da5176	0.658	0.834	0.736	0.621	0.942	0.935	0.837	0.607	0.858
Rui Feng 168	0.895	0.847	0.693	0.726	0.838	0.949	0.78	0.508	0.897
Zheng Da 1473	0.783	0.828	0.55	0.786	0.932	0.925	0.793	0.777	0.881
BoFa707	0.792	0.818	0.733	0.608	0.931	0.924	0.835	0.736	0.892
XiWang 3088	0.792	0.818	0.733	0.708	0.931	0.924	0.835	0.736	0.892
XianYu1653	0.794	0.817	0.766	0.688	0.901	0.915	0.821	0.752	0.819
XianYu1568	0.837	0.804	0.742	0.701	0924	0..901	0.875	0.713	0.872

Eleven varieties of summer corn were selected in the Guanzhong Plain of Shaanxi Province in 2019 to conduct multiple linear regression analysis of main agronomic traits. Results showed that the agronomic traits closely related to yield of summer maize varieties in Shaanxi Province were 100-grain weight, number of grains per row, seed rate, empty stalk rate and ear length, etc that are the main influencing factors of hybrid breeding. It showed that when selecting corn varieties, on the premise of ensuring proper 100-grain weight and ear length, one should pay attention to selecting varieties with high seed rate and low empty stalk rate. In the selection of high-yielding maize varieties, lodging resistance should be enhanced, the rate of empty stalks should be reduced, the rate of seed production should be increased, the 100-grain weight and ear diameter should be appropriately increased, and the ear height should be reduced.

Acknowledgements

This study is financially supported by Shaanxi Province Natural Science Basic Research Program (2022JQ-457); The Internal Scientific Research Projects of Shaanxi Provincial Land Engineering Construction Group (DJNY2022-25).

References

Ali S, Xu YY, Ma XC, Ahmad I, Manzoor JQM, Akmal M, Hussain Z, Arif M and Cai T 2019. Deficit irrigation strategies to improve winter wheat productivity and regulating root growth under different planting patterns. Agr. Water Manage. **219**(1): 1-11

Axel H, Giuliana D, Francisco GA and Fernandez FA 2021. Sustainable production of microalgae in raceways: Nutrients and water management as key factors influencing environmental impacts. Cleaner Product. 287(1): 125005.

Fernando SG 2020. Investigation of azospirillum brasilense inoculation and silicon application on corn yield responses. Soil Sci. Plant Nutri. **102**(1): 1-13.

Forte A, Fiorentino N, Fagnano M and Fierro A 2017. Mitigation impact of minimum tillage on CO_2 and N_2O emissions from a Mediterranean maize cropped soil under low-water input management. Soil Tillage Res. **166**(1): 167-178.

He HJ, Zhou GS 2016. Climate-associated distribution of summer maize in China from 1961 to 2010. Agricul. Ecosyst. Environ. **232**(1): 326-335.

Hou HJ, Han ZD,Yang YQ, Abudu S and Li ZC 2020. Soil CO_2 emissions from summer maize fields under deficit irrigation. Environ. Sci. Pollu. Res. **27**(1): 4442-4449.

Liu Y, Qin Y, Ge Q, Dai J and Chen Q 2017. Reponses and sensitivities of maize phenology to climate change from 1981 to 2009 in Henan Province, China. Geograph. Sci. 27: 1072-1084

Lu JS, Ma LH, Hu TT, Geng CM and Yan SC 2022. Deficit drip irrigation based on crop evapotranspiration and precipitation forecast improves water- use efficiency and grain yield of summer maize. Sci. Food Agricul. **102**(2):653-663.

Nascimento LO 2020. Effect of reduced spacing on relationship of physiological, morphological and productive traits of corn yield. Australian J. Crop Sci. **14**(7): 1202-1208.

Nafziger ED, Srinivasan V 2021. Nitrogen deficiency and corn yield with delayed N application. Agron. J. **113**(4): 3665-3674.

Qiu QY, Wu LF, Ouyang Z, Li BB, Xu YY, Wu SS and Gregorich EG 2015. Effects of plant-derived dissolved organic matter (DOM) on soil CO_2 and N_2O emissions and soil carbon and nitrogen sequestrations. Appl. Soil Ecol. **96**(1): 122–130

Ruis SJ 2021. Corn residue baling and grazing impacts on corn yield under irrigated conservation tillage systems. Agron. J. **113**(3): 2387-2397.

Shahhosseini M 2021. Corn yield prediction with ensemble CNN-DNN. Front. Plant Sci. **12**(1): 709008-709008.

Simão LM and Johnston AM 2021. Winter wheat residue impact on soil water storage and subsequent corn yield. Agron. J. **113**(1): 276-286.

Singh R 2021. Effects of broiler litter application rate and time on corn yield and environmental nitrogen loss. Agron. J. **114**(1): 415-426.

Tao L. 2020. DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation. Environ. Res. Lett. **15**(3): 034016-034016.

Xiang KY, Yi L, Robert H and Hao F 2020. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration - a review. Agricul. Water Manage. **232**(1): 106043.

Yue HW, Jiang XV, Wei JW, Xie JL, Chen SP, Peng HC and Bu JZ 2021. A study on genotype × environment interactions for the multiple traits of maize hybrids in China. Agron. J. **113**(6): 4889-4899.

(Manuscript received on 15 June, 2022; revised on 21 September, 2022)