A Study on the Prediction of the Quantity of Talents: A Case Study of Shandong Province

Sun Chenxi [1], Zhao Shuai [2,*], and Liu Yuchuang [3]

[1] School of Science, Shandong Jianzhu University, Jinan, 250101 China
[2] School of Civil Engineering, Shandong Jianzhu University, Jinan, 250101 China
[3] School of Thermal Engineering, Shandong Jianzhu University, Jinan, 250101 China
Email: 2993576476@qq.com (Z.S.)
*Corresponding author

Manuscript received December 11, 2023; revised December 24, 2023; accepted January 17, 2024; published April 17, 2024.

***Abstract*—Taking the sustainable development of R&D talents as an example, this paper proposes a prediction model based on Pearson correlation analysis and stepwise regression analysis. Using this model, it identifies the main factor affecting the quantity of technology-driven talents in Shandong Province from four dimensions: economic development, residents' living standards, social development, and technological activities, which is the added value of the secondary industry. The model predicts a potential talent outflow in Shandong Province in the future. By comparing the influencing factors on the quantity of technology-driven talents in Shandong Province with those in Jiangsu Province and Guangdong Province, the paper reveals the advantages and disadvantages of Shandong Province in attracting technology-driven talents and provides two recommendations: (1) Increase R&D research funding reasonably (2) vigorously develop the secondary industry. Predicting the trend of the number of R&D talents can help countries to plan for strategy. This article will provide suggestions for the economic development of Shandong Province through the prediction of the number of talents in Shandong Province.**

***Keywords*—R&D talents, Pearson correlation coefficient test, stepwise regression analysis**

I. Research Status

Existing research mainly focuses on predicting the total talent demand at the regional or industry level. For example, at the regional level, (Yao *et al*., 2021) studied the talent demand in Sichuan, while (Zhang *et al*., 2020) studied the talent demand in Henan Province; at the industry level, (Liu Zhimin *et al*., 2020) predicted talent demand in the agricultural science field, (Li *et al*., 2020) studied the talent demand in Hubei Province's automotive and steel industries, and (Chen *et al*., 2020) studied the talent demand in Hainan Province's large health industry. These studies have predicted talent demand at the meso level from different regions or industries and have achieved positive results. However, for national-level development strategies, predicting the talent demand in Shandong Province can provide support for strategy formulation and implementation, thus having a certain research significance.

II. A Combined Prediction Model Based on Pearson Correlation Test and Stepwise Regression

In order to better predict the changes in the number of R&D talents, we established a multiple linear prediction model for the changes in the number of R&D talents based on a combination model of Pearson correlation test and stepwise regression.

A. Influencing Factors

Based on the research of relevant documents, it can be seen that the factors that affect the stock of R&D personnel in Shandong Province are mainly composed of four parts: economic development, residents' living standards, social development, and technological activities. The specific factors are shown in Table 1.

Table 1. Influence factor

Type	Influence factor
Scientific activities	R&D investment in Shandong Province X_0/100 million yuan
	Shandong Province's R&D expenditure as a proportion of GDP X_1
Resident life	Per capita disposable income of urban residents in Shandong Province X_2/yuan
social development	Shandong Province Technology Market Transaction Value X_5/10000 yuan
	The total output value of construction enterprises in Shandong Province is X_6/10000 yuan
	Total fixed assets investment of Shandong Province is X_7 billion yuan
economic development	The added value of the secondary industry in Shandong Province is X_3/100 million yuan
	The added value of the tertiary industry in Shandong Province is X_4/100 million yuan
	Shandong Province GDPX_8/100 million yuan

Due to the different data units in the raw data, directly analyzing the data may lead to inaccurate conclusions. We used normalization processing to scale all factors to the same unit level, making them easier to compare and analyze.

B. Pearson Correlation Test

This study uses Pearson correlation coefficient for correlation analysis to accurately screen the main influencing factors of the number of scientific and technological talents in Shandong Province. The Pearson correlation coefficient is used to measure the degree of linear correlation between two variables X and Y, with a value between −1 and 1. The larger the absolute value, the stronger the correlation. The calculation formula for Pearson correlation coefficient (Yao, 2019; Zhang and Zhao, 2009) is:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{s_X}\right)\left(\frac{Y_i-\overline{Y}}{s_Y}\right) \quad (1)$$

Among them, r is the Pearson correlation coefficient, X and

S_X is the average and standard deviation of the sample, X_i is the actual value of the sample (i=0, 1,... 8).

This study empirically analyzes the influencing factors of scientific and technological talents in Shandong Province. The calculation results are shown in Table 2:

Table 2. Table of Pearson correlation coefficient test results in Shandong Province

Variable	Correlation coefficient R	Correlation
X_3	0.990	Correlation
X_8	0.988	Correlation
X_0	0.978	Correlation
X_2	0.970	Correlation
X_1	0.965	Correlation
X_7	0.963	Correlation
X_6	0.960	Correlation
X_4	0.951	Correlation
X_5	0.724	Correlation

From Table 1 above, it can be seen that when R>0.97, the correlation between X_i and Y is good; When 0.96<R<0.97, the correlation between X_i and Y is average; When R<0.96, the correlation between X_i and Y is poor. Therefore, X_3, X_8, X_0, X_2 have a good correlation with Y; The correlation between X_1, X_7, X_6 and Y is average; The correlation between X_4, X_5 and Y is poor.

C. Stepwise Regression Analysis

To further accurately screen the main influencing factors of the number of scientific and technological talents in Shandong Province, this study conducted factor screening through stepwise regression analysis.

This article conducts stepwise regression analysis on various factors using MATLAB, and the specific results are shown in Fig. 1.

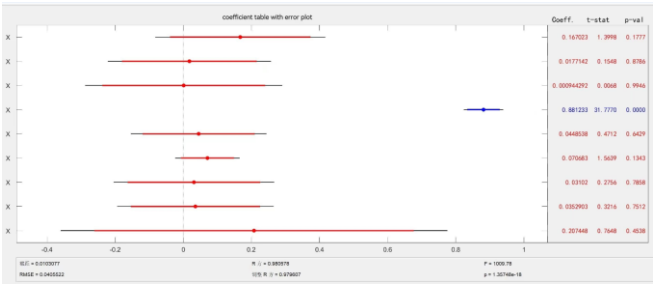

Fig. 1. Stepwise regression analysis chart of Shandong Province.

Generally, we believe that when the t-test value of the influencing factor is greater than 2, the p-value corresponding to the t-test is less than 0.05, indicating that X has a significant influencing factor on Y. From Fig. 1, it can be seen that only the t-test value of X_3 is greater than 2, and the p-value corresponding to the t-test is 0. So among the many factors that affect scientific and technological talents in Shandong Province, only X_3 plays a significant role.

The results obtained through stepwise regression analysis are shown in Table 3.

Table 3. Summary table of stepwise regression analysis results

Summary of stepwise regression results	
method	stepwise
Total variable situation	X_1, X_0, X_2, X_4, X_8, X_5, X_3, X_7, X_6
Preserve variables	X_3
Abandoning variables	X_1, X_0, X_2, X_4, X_8, X_5, X_7, X_6

D. Establishment of Linear Prediction Model

From the above combination model, it can be seen that the main influencing factor of Y in Shandong Province is X_3. For the data X_3, a linear function fitting was performed to obtain the following function expression:

$$Y_{new} = 0.0103 + 0.8812 \times x_3. \quad (2)$$

According to the normalized data of X_3, the grey prediction model is used to predict the data for the next five years, as shown in Table 4.

Table 4. Prediction table of influencing factors

Year	X_3
2022	1.095
2023	1.164
2024	1.235
2025	1.308
2026	1.384

Note: Data missing for 2022

By incorporating the predicted X_3 data for the next five years into the functional relationship expression of Y, the normalized values of Y for the next five years can be derived.

The specific results are shown in Table 5:

Table 5. Prediction tables for Ynew and Y_0

Year	Ynew	Y_0
2022	1.109858	772400.2
2023	0.943411	728691.1
2024	0.999340	728210.4
2025	0.998821	727351.8
2026	0.998946	726585.3

E. Model Verification

To test the accuracy of the above multiple linear prediction model, we conducted relative error analysis using Eq. (3) for Y_0 and Y values in 2011, 2016, and 2021. The specific results are as follows:

$$D = \frac{|Y_0 - Y|}{Y} \quad (3)$$

After testing, the relative error D is less than 0.07, so the accuracy of the model is good

III. Suggestions

In order to promote the growth of the number of scientific and technological talents in Shandong Province and optimize the structure of its scientific and technological talent team, we conducted in-depth research on the policies related to scientific and technological talents introduced by Jiangsu Province and Guangdong Province, compared and analyzed the advantages and disadvantages of Shandong in introducing scientific and technological talents, and proposed reasonable and effective suggestions for better constructing the policy system for scientific and technological talents in Shandong Province.

A. Analysis on the Policy of Technological Talents in Jiangsu Province

Jiangsu Province ranks fifth in terms of the number of people with a college degree or above in the country, and third in terms of per capita GDP, which is a leading position in the country. By analyzing the influencing factors of technological talents in Jiangsu Province, reference can be

provided for Shandong Province.

Combining Eqs. (1) to (3) yields the following results:

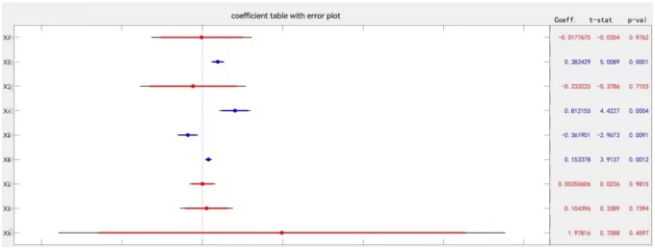

Fig. 2. Stepwise regression analysis chart of Jiangsu Province.

Generally, we believe that when the t-test value of the influencing factor is greater than 2, the p-value corresponding to the t-test is less than 0.05, indicating that X has a significant influencing factor on Y. Obviously, as shown in the above figure, the t-test values for X_3, X_1, X_5, and X_4 are greater than 2, and the p-value corresponding to the t-test is 0. So among the many factors that affect technological talents in Jiangsu Province, the effects of X_3, X_1, X_5, and X_4 are significant.

Table 6. Summary results of progressive regression in Jiangsu Province

Summary of stepwise regression results	
method	stepwise
Total variable situation	X_3, X_1, X_5, X_2, X_8, X_4, X_7, X_0, X_6
Preserve variables	X_3, X_1, X_5, X_4
Abandoning variables	X_2, X_8, X_7, X_0, X_6

Using Eq. (1), the following result can be obtained:

Table 7. Table of Pearson correlation coefficient test results in Jiangsu Province

variable	**Correlation coefficient R**	**Correlation**
X_3	0.992	Correlation
X_1	0.986	Correlation
X_2	0.984	Correlation
X_6	0.982	Correlation
X_8	0.981	Correlation
X_7	0.976	Correlation
X_0	0.969	Correlation
X_4	0.967	Correlation
X_5	0.842	Correlation

As shown in the above table, the correlation between Xi and Y is good when R > 0.97; When 0.96 < R < 0.97, the correlation between Xi and Y is average; When R<0.96, the correlation between Xi and Y is poor. Therefore, X_3, X_1, X_2, X_6, X_8

The correlation with Y is good; The correlation between X_1, X_7, X_6 and Y is average; The correlation between X_4, X_5, and Y is poor.

According to the above combination model, the main factors affecting Y in Jiangsu Province are X_1 and X_3.

B. Analysis of Guangdong Province's Science and Technology Talents Policy

Guangdong Province ranks 13th in terms of the number of people with a college degree or above in the country, and 7th in terms of per capita GDP, slightly better than Shandong Province. By analyzing the influencing factors of technological talents in Guangdong Province, reference can be provided for Shandong Province.

Combining Eqs. (1) to (3) yields the following results:

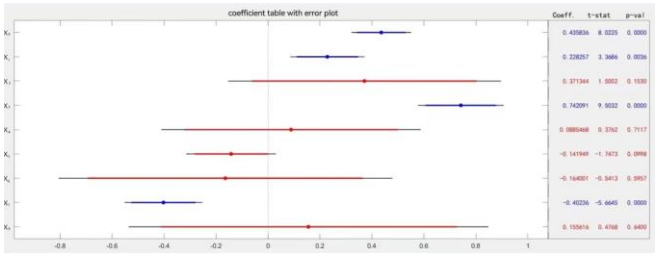

Fig. 3 Stepwise regression analysis chart of Guangdong Province.

Generally, we believe that when the t-test value of the influencing factor is greater than 2, the p-value corresponding to the t-test is less than 0.05, indicating that X has a significant influencing factor on Y. Obviously, as shown in the above figure, the t-test values for X_3, X_0, X_1, and X_7 are greater than 2, and the p-value corresponding to the t-test is 0. So among the many factors that affect technological talents in Jiangsu Province, the effects of X_3, X_0, X_2, and X_7 are significant.

The results obtained through stepwise regression analysis are shown in Table 8.

Table 8. Summary results of gradual regression in Guangdong Province

Summary of stepwise regression results	
Method	stepwise
Total variable situation	X_0, X_4, X_1, X_2, X_5, X_8, X_3, X_7, X_6
Preserve variables	X_7, X_1, X_3, X_0
Abandoning variables	X_4, X_2, X_5, X_8, X_6

Using Eq. (1), the following result can be obtained:

Table 9. Table of Pearson correlation coefficient test results in Guangdong Province

Variable	**Correlation coefficient R**	**Correlation**
X_3	0.990	Correlation
X_8	0.988	Correlation
X_0	0.978	Correlation
X_2	0.970	Correlation
X_1	0.965	Correlation
X_7	0.963	Correlation
X_6	0.960	Correlation
X_4	0.951	Correlation
X_5	0.724	Correlation

Generally, we believe that the correlation between X_i and Y is good when R > 0.97; When 0.96 < R < 0.97, the correlation between X_i and Y is average; When R < 0.96, the correlation between Xi and Y is poor. From the above table analysis, it can be seen that X_0, X_3, X_8, X_2 have a good correlation with Y; The correlation between X_1, X_7, X_6 and Y is average; The correlation between X_4, X_5, and Y is poor. Based on the results of stepwise regression analysis, it can be concluded that the main factors affecting Y in Guangdong Province are X_3, X_0, X_1, and X_7.

Therefore, by drawing on the factors that affect the number of technology-based talents in Jiangsu Province and Guangdong Province, Shandong Province can adjust from

three aspects: X_0, X_1, and X_7.

C. Research on Suggestions for Shandong Province's Policy on Science and Technology Talents

Based on the previous research on Jiangsu and Guangdong provinces, it can be seen that X_1, X_0, and X_7 are factors that Shandong Province can learn from. Below, a linear multiple fitting model Y is constructed based on a total of 6 factors, including X_0 that affects the growth of scientific and technological talents in Shandong Province and X_1, X_0, and X_7 that can be used for Ynew:

$$Y'_{new} = 0.0378 + 0.5024 \times x_0 - 0.1856 \times x_1 + 0.9030 \times x_3 - 0.3532 \times x_7. \quad (4)$$

By adjusting the multivariate linear prediction model for scientific and technological talents in Shandong Province as mentioned above:

Make predictions and the results are as follows:

Table 10. Ynew Forecast Table

Year	Ynew	Y_0
2021	1.109858	772300
2022	0.94311	728691
2023	0.99934	728210
2024	0.998821	727351
2025	0.998946	726585

It is not difficult to find that since 2022, there has been a loss of scientific and technological talents in Shandong Province. To avoid this situation as much as possible, we provide suggestions for the development of Shandong Province through the following analysis.

According to the formula, X_0 and X_3 are positively correlated with the number of scientific and technological talents in Shandong Province. Therefore, it is necessary to increase R&D research funding and develop the secondary industry in Shandong Province to avoid the loss of scientific and technological talents in Shandong Province. However, it is necessary to increase R&D research funding and develop the secondary industry in Shandong Province within a reasonable range. Below we provide specific values for R&D research funding and the increase in the secondary industry.

Under other unchanged conditions in 2022, taking X_0 as an example, it is necessary to increase research funding by at least 0.4% to avoid the loss of scientific and technological talents in Shandong Province. Taking X_3 as an example, the output value of the secondary industry needs to increase by at least 448.2 billion yuan to avoid the loss of scientific and technological talents in Shandong Province. Below we provide corresponding measures.

Regarding R&D Funding: Increase government investment in research funding, as government investment in research is a key driver for cultivating technology-based talents and promoting research activities. Therefore, the input of R&D funding can be increased by increasing the budget for research funding, establishing special funds, or providing tax incentives. Additionally, increasing its share of GDP can be achieved. Strengthen government-business cooperation, as the government actively promotes cooperation between government, universities, and enterprises to jointly invest in and carry out research projects. Promote the transformation and application of scientific and technological achievements to attract more technology-based talents to participate in research and development activities. Encourage innovative research from universities and research institutions, as they are the main forces for cultivating technology-based talents and promoting research. This can be achieved by providing more research funding, establishing a research project reward mechanism, and improving research conditions to encourage universities and research institutions to strengthen innovative research and attract more researchers to participate. Improve the treatment and development opportunities for researchers to attract technology-based talents. This can be achieved by improving researchers' salary treatment, providing a good working environment and development opportunities, including providing more research projects and project funding, and supporting researchers' participation in international academic exchanges and cooperation. Promote the combination of industry, education, and research to actively promote the cooperation between industry, universities, and research institutions to carry out R&D projects. This cooperation can provide opportunities for solving practical problems, helping technology-based talents transform research results into innovative products and technologies, and promoting a virtuous cycle of technological innovation and talent cultivation.

Regarding the output value of the secondary industry: Strengthen industrial agglomeration effects by promoting the clustering of related industries in industrial parks, science parks, and other ways. Build a good industrial ecological environment to attract enterprises to concentrate their development and form a complete closed loop of industry chains and value chains to enhance industrial competitiveness and influence. Promote collaborative industry development: strengthen the collaborative development with other related industries to form industrial integration and complementary effects. Encourage the secondary industry to cooperate and exchange with the primary and tertiary industries to promote resource sharing and complementary advantages, and promote the extension and expansion of industry chains. Strengthen market development and brand building by actively conducting market research and marketing promotion activities to expand domestic and foreign markets and increase the sales channels and influence of the secondary industry. At the same time, focus on brand building to enhance product and corporate image, and enhance market competitiveness. Encourage secondary industry enterprises to participate in international cooperation and exchanges to expand overseas markets. Promote the introduction and export of technologies by strengthening industrial cooperation with other countries and regions to improve the international competitiveness of Shandong's secondary industry.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Sun Chenxi and Liu Yuchuang conducted the research; Zhao shuai and Sun Chenxi analyzed the data; Zhao shuai and Liu Yuchuang wrote the paper; all authors had approved the final version.

REFERENCES

Chen, Y. Y., Hei, Q. M., Liu, C. P, etc, 2017. Construction of a multiple regression prediction model for talent demand in the big health industry of Hainan Province: From the perspective of principal component analysis and trend extrapolation combination prediction method. ***China Market***, (29): 101–107.

Du, Q. State of song wei definition of technological talents and related statistical issues. ***China Science and Technology Forum***, 200(5): 137–141.

Yao, J. 2019. ***Analysis of Sichuan, Shaanxi, and Shanghai Science and Technology Progress and Countermeasure*s**, 36(14): 46–52.

Zhang, J. Y., Zhao, T. 2009. Prediction of technology talent demand based on BP neural network. ***Research on Technology Management***, 29(8): 501–502.

Li, X. Y., Wang, Y. J. 2014. Research on the demand for overseas high level talents in key industries: Taking the automobile and steel industries in Hubei Province as an example. ***Science and Technology Progress and Countermeasures***, 31(8): 135–139.

Liang, S. Z., Chen, Z. 2020. Research on demand prediction of technological innovation talent resources in the Guangdong Hong Kong Macao Greater Bay Area. ***Contemporary Economy***, 514(10): 114–118

Liu, Z. M., Zhao, X. N. 2019. Prediction of agricultural talent demand in China from 2020 to 2025: Analysis based on grey GM (1,1) model. ***China Agricultural Education***, 20(5): 42–51.

Liu, Q, Xiao, J, Shi, Z, *et al.* 2020. Talent and urban development based on gray prediction model. Inovative Computing: Yang, CT., Pei, Y., Chang, JW. (eds) ***Innovative Computing: IC 2020***. Springer Singapore, 1117–1126.

Luo. J., Qu, J. W., Yang, R. J. 2023. Prediction of the demand for Shaanxi science and technology talents under the combination model. ***Technology and Industry***, 23(05): 19–24.

Min, X, L. 2005. Talent demand analysis based on grey prediction model GM (1,1). ***Science and Technology Management Research***, (6): 72–74,77.

Nasar, N., Ray, S., Wadud, A., *et al.*, 2020. Prediction model for talent management analysis through e-HRM. Recent Advances in Operations Management Applications. Select Proceedings of CIMS 2020. Singapore, Springer Nature Singapore, 193–207.

Qu, Q. Z., Wang, J. J., Tang, M. X., etc. 2021. Prediction of China's demand for scientific and technological talents during the 14th five year plan period based on a combination model. ***Science and Technology Management Research***, 41(21): 129–135.

Ren, J., Ji, M., Chen, Z. N. Analysis and prediction of talent demand in Hangzhou City-Based on GM (1,1) and ARIMA models. ***Modern Information Technology*.** 20(04): 9–13+16.