Nuclear weapon quantity prediction based on the lasso regression model

Jinlin He[1, #], Liyun Chen[2, *, #], Hao Xu[2, #] and Lihua Huang[2, #]

[1]College of Electronic Information and Automation, Civil Aviation University of China, TianJin, China

[2]College of Applied Science and Technology, Hainan University, DanZhou, China

*Corresponding author: c1972359255ly@sina.com

[#]These authors contributed equally to this work

Abstract. Once a nuclear weapon explodes, its power will be devastating. In this paper, we analyze the current situation of nuclear weapons in the world, and firstly, we analyze the relationship between the number of nuclear-armed countries and the year using linear regression to find out the prediction model, from which we get that in 2123, the number of nuclear-armed countries in the world will reach 9. Then, this paper analyzes the Pearson correlation between the number of nuclear weapons of each country and the year and uses lasso regression to model the change in the number of nuclear weapons of each country with the year. Special attention is paid to the fact that the number of nuclear weapons in South Africa in the next 100 years is 0, indicating that the number of nuclear-armed countries will be reduced to 9 in 2123, which is consistent with the prediction of the number of nuclear-armed countries through linear regression. In 2123, the total number of nuclear weapons in the world will reach 29,293. To derive a limited total number of nuclear weapons in the world, a nuclear bomb detonation location destructive force model is developed in this paper to derive the maximum destructive range of nuclear bombs. Combined with the current number of nuclear bombs in the world, the calculation results show that the current number of nuclear bombs in the world is not enough to destroy the earth, and the total number of nuclear bombs in the world should be limited to 44,563.

Keywords: Lasso regression; Nuclear weapons; Linear regression; Pearson correlation analysis.

1. Introduction

Nuclear weapons are hugely lethal weapons associated with nuclear reactions, including hydrogen bombs, atomic bombs, neutron bombs, etc. It is one of the most powerful weapons in the history of mankind. The instantaneous explosion temperature of an atomic bomb can reach tens of millions of degrees. The explosive yield of a nuclear bomb is about tens to hundreds of thousands of tons of TNT equivalent. The explosion of an atomic bomb and its radiation range can destroy a city. Because of the terrible power of the atomic bomb, many countries wanted to use it to deter other countries and protect themselves from foreign invasion. After World War II, countries around the world began frantically researching and manufacturing atomic bombs, even creating destructive weapons such as the “Big Ivan”.

In this paper, by analyzing the number of nuclear weapons possessed by each country and using lasso regression models to predict the number of nuclear weapons in each country in the next 100 years, and by combining the destructive power of nuclear weapons, we calculate the number of nuclear weapons that the world should limit at present in order not to destroy the earth.

2. Prediction of the number of nuclear-armed countries based on the linear regression model

According to statistics, ten countries have ever possessed nuclear weapons, namely China, France, India, Israel, North Korea, Pakistan, Russia, South Africa, the United Kingdom, and the United States. The next predicts the countries that will have nuclear weapons in the next 100 years.

2.1. Model Building

Linear regression is a statistical analysis method that uses regression analysis in mathematical statistics to determine the quantitative relationship between two or more variables. The expression is $y = w'x + e$, with e being the mean value of the error obeying a normal distribution.

In regression analysis, only one independent variable and one dependent variable are included, and the relationship between the two can be approximated by a straight line. This type of regression analysis is called univariate linear regression analysis. If the regression analysis includes two or more independent variables and there is a linear relationship between the dependent and independent variables, it is called a multiple linear regression analysis [1].

In general, linear regression can be obtained by least squares or gradient descent methods to obtain its equation, which can be calculated as a straight line for $y = bx + a$. Taking the least squares method as an example, in general, there is more than one factor affecting y. Assuming that there are $x_1, x_2, \dots, x_k$, k factors, the following linear relationship can be considered.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{1}$$

The *n* sets of observations $(x_{t_1}, x_{t_2}, \dots, x_{t_k})$ are obtained from n independent y and $x_1, x_2, \dots, x_k$, obtained from simultaneous observations with $\mathrm{t} = 1, 2, \dots, \mathrm{n}\ (\mathrm{n} > \mathrm{k} + 1)$. They satisfy the following relations:

$$y = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t \tag{2}$$

Where the uncorrelated ones are random variables with the same distribution. To represent the above equation in terms of a matrix, let

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

Using least squares to obtain a solution.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

Among them, $(X^T X)^{-1} X^T$ is called the pseudo-inverse.

The slope *b* is calculated as follows.

$$u(b) = \frac{\sigma}{\sqrt{\sum (x - \bar{x})^2}} \tag{5}$$

Then the intercept a is as follows.

$$U(a) = U(b) \sqrt{\frac{\sum x_i^2}{n}} \tag{6}$$

2.2. Model Testing

Table 1. Linear regression model test analysis table

Linear regression analysis results n=31									
	non-normalized coefficients		standardized coefficient	t	P	VIF	R²	Adjustment R²	F
	B	standard error	Beta						
constant	-87.755	10.497	-	-8.36	0.000***	-	0.744	0.735	F=84.171 P=0.000***
Year	0.048	0.005	0.862	9.174	0.000***	1			
dependent variable：Possession									

***、**、* represents the significance level of 1%、5%、10% respectively.

The analysis of the F-test results showed that the significant P-value of 0.000 **** was significant at the level, rejecting the null hypothesis that the regression coefficient was 0, so the model met the requirements. Meanwhile, the VIF is less than 10, so the model does not have multiple co-problems and the model is well constructed.

The final prediction model is as follows.

$$y=-87.755 + 0.048*year \tag{7}$$

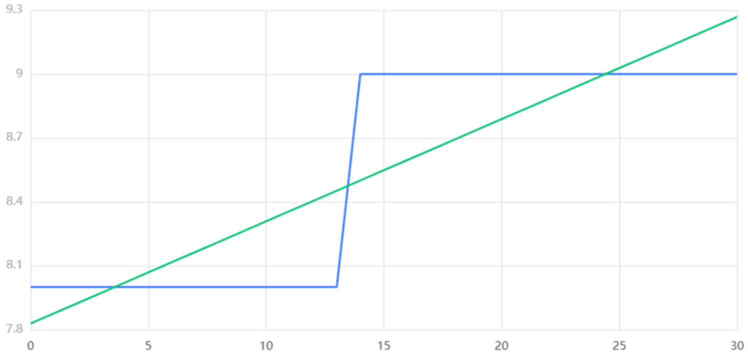


Fig. 1 Fitting effect diagram

We rounded the predictions to get nine nuclear weapon states in 2123.

3. Forecasting the number of nuclear weapons based on the lasso regression model

3.1. Pearson Correlation Analysis

To test whether there is a statistically significant relationship between the number of nuclear weapons possessed by each country and the year in the next 100 years, we used Pearson correlation analysis to correlate the number of nuclear weapons possessed by each country and the year.
The Pearson correlation coefficient is defined as follows [3].

$$\rho_{XY} = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}\sqrt{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}} \tag{8}$$

When $\rho_{XY} = 0$, X and Y do not have a linear correlation; when $\rho_{XY} > 0$, X and Y have a positive linear correlation; when $\rho_{XY} < 0$, X and Y have a negative linear correlation; when ρ_{XY} is closer to ±1, the higher the correlation.

Table 2. Correlation coefficients

Country	China	France	India	Israe	North Korea	Pakistan	Russia	South Africa	United Kingdom	United States
Correlation	0.931	0.782	0.749	0.973	0.483	0.756	0.163	0.037	0.301	-0.189
Significance	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.155	0.749	0.007***	0.097*

Note: * * *, * * *, * represent 1 %, 5 %, 10 % significance levels, respectively.

From the above table, it can be seen that in the correlation analysis between the number of nuclear weapons and the year of the same country, China and Israel have a strong correlation with the year. France, India, and Pakistan have a strong correlation with year; South Korea and the UK have an average correlation with year; Russia and South Africa have a weak correlation; and the US has a negative correlation with year.

3.2. Model Building

Lasso regression is a compressive estimation method that replaces the least squares method. Its basic idea is to build an L1 regularization model. In the process of building the model, some coefficients are compressed and some coefficients are set to zero. When the model training is completed, these parameters with weights equal to zero can be discarded, which can make the model simpler and effectively prevent model overfitting. It is widely used for fitting multicollinearity data and variable selection [2].

Assuming the data $\{X_i, Y_i\}, X_i = \{X_{i_1}, \ldots, X_{i_m}\}$ and Y_i are the explanatory and corresponding variables corresponding to the ith observation, respectively. Consider a linear regression model as follows.

$$Y_i = \alpha_i + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i, \qquad \varepsilon_i \sim (0, \sigma^2) \tag{9}$$

In the usual regression structure, it is assumed that the observations are independent of each other, or that when the observations are given, Y_i is independent, i.e., Y_i is independent of the X_i condition, and x_{ij} is normalized, i.e., $\frac{1}{n}\sum_j x_{ij} = 0, \frac{1}{n}\sum_j x_{ij}^2 = 1$.

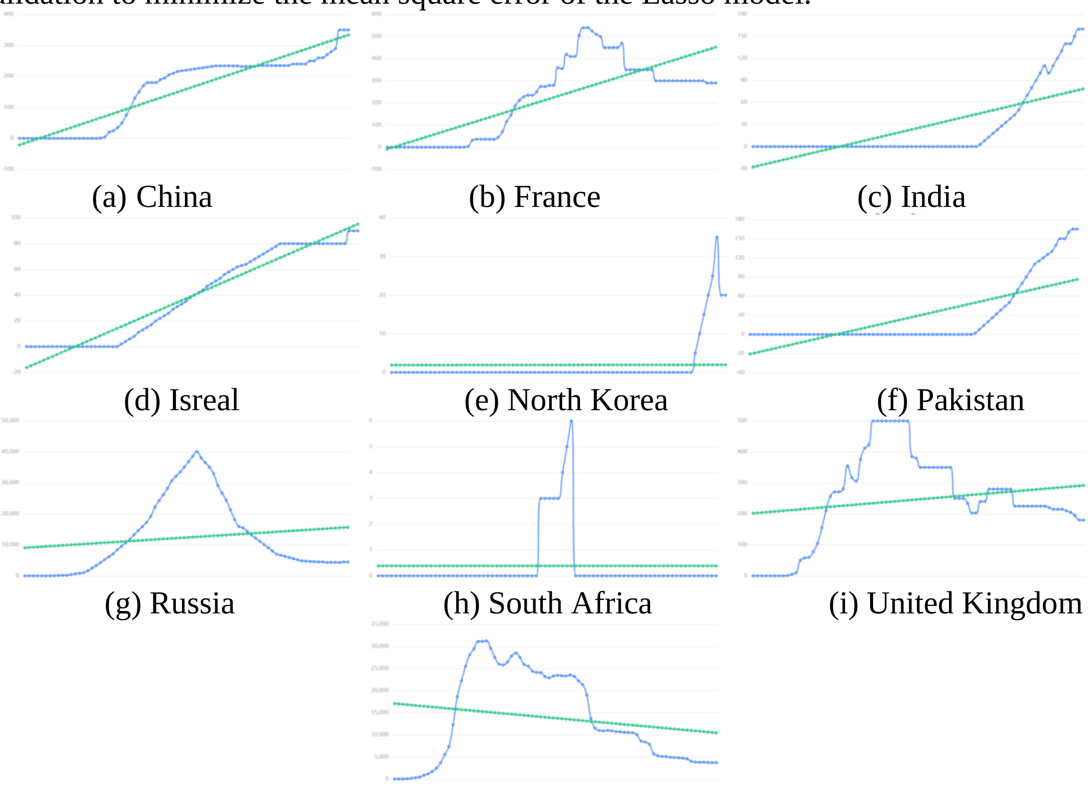
$$(\hat{a}, \beta) = \widehat{argmin} \left\{ \sum_i Y_i - \alpha_i - \sum_j \beta_j x_{ij})\text{^}2 \right.$$

$$\text{s.t.} \sum_j |\beta_j| \leq 1 \tag{10}$$

Here $t \geq 0$, it is a harmonic parameter. At this point, for all t, $\hat{\alpha} = \bar{y}$. Without loss of generality, it is assumed that α is omitted. The control of the harmonic parameter t makes the total regression coefficient smaller. If $t^0 = \sum_j |\beta_j|$, $t \leq t^0$ is made, it will cause some regression coefficients to shrink and converge to 0, and some coefficients to even equal 0.

3.3. Predicted results for the number of nuclear weapons

The variables selected by the model are determined by λ and the regression coefficient plot. The choice of the parameter λ determines the degree to which the regression coefficients are compressed, and different λ may produce different results. The variables with zero standardized coefficients can be considered to be excluded from the Lasso regression model. The value of λ is determined by cross-validation to minimize the mean square error of the Lasso model.



(a) China (b) France (c) India

(d) Isreal (e) North Korea (f) Pakistan

(g) Russia (h) South Africa (i) United Kingdom

(j) United States

Fig. 2 Prediction curve chart

Figure 2 shows the predicted curves for each country compared to the true values, where the true values are in blue and the predicted values are in green. It can be found that the number of nuclear weapons in South Africa in the next 100 years is 0, which means that the number of nuclear states in 2123 is 9, which is consistent with the previous prediction.

Table 3. Analysis of prediction results

Country	China	France	India	Israe	North Korea	Pakistan	Russia	South Africa	United Kingdom	United States
Predicted results	800	1077	241	242	21	263	24456	0	519	1673
R^2	0.867	0.611	0.555	0.946	0.003	0.565	0.026	0	0.076	0.036

The table shows that Russia will have the most nuclear weapons in the next 100 years, with 24,456; North Korea will have the least, with 21, except for South Africa, which will become a non-nuclear country. Thus, the number of nuclear weapons in the world will reach 29,293 by the year 2123.

4. The destructive power of nuclear weapons

Shock waves and optical radiation from nuclear explosions are the main destructive factors [5]. They both act as destructive factors within seconds to seconds after the explosion, which is also known as a transient destructive factor. The maximum range of shock waves and optical radiation does not exceed 20 km [4]. Assuming that the range of shock waves and light radiation equivalent to all nuclear weapons is 20 km and that the energy released by the nuclear explosion is not absorbed and reflected by the ground, an expression for the distance h from the explosion point to the ground in the range of influence S of the nuclear explosion can be obtained.

$$S = \left(\sqrt{20^2 - h^2}\right)^2 \times \pi \quad (\mathrm{km}^2) \tag{11}$$

Thus, when $h = 0$, the maximum area of impact of a nuclear explosion is $S = 20\pi$.

The surface area E of the Earth is $5.11 \times 10^8 km^2$ and the number of bombs required to destroy the Earth is assumed to be *N*.

$$N = \frac{E}{S} = \frac{5.11\times10^8}{20} = 2550000 \tag{12}$$

As of 2022, there are 9440 nuclear weapons in the world, assuming an impact of s when all of them explode.

$$s_{max} = 9440 \times 20\pi = 188S_{max} = 9940 \times 20\pi = 188800\pi < 5.11 \times 10^8\pi \tag{13}$$

So there are not enough nuclear bombs in the world to destroy the Earth.

Nuclear weapons detonated in desert areas can reduce the impact on humans and the environment. The area S_D of the world's deserts (including semi-deserts) is about 28 million square kilometers or 1/5 of the world's land area. therefore, the total number of nuclear bombs limiting world A should be as follows.

$$A = \frac{S_D}{S} = \frac{2.8\times10^6}{20\pi} \approx 44563 \tag{14}$$

Therefore, the total number of nuclear weapons in the world should be limited to 44,563.

5. Conclusions

In this paper, a linear regression model is built based on the current nuclear possession of countries in the world and predicts that there will be 9 countries possessing nuclear weapons in the next 100 years, and the model is well constructed as the p-value is much less than 0.05 and there is no multiple co-linearity problems after testing. Meanwhile, the lasso regression model established in the latter paper predicts that the number of nuclear weapons owned by South Africa in the next 100 years is 0, i.e., the number of nuclear-armed countries is reduced to 9, which is consistent with the results of linear regression and corroborates each other. Finally, this paper establishes a model of the destructive power of nuclear bomb detonation location and concludes that the current nuclear weapons in the world are not enough to destroy the earth, and the total number of nuclear weapons in the world must be limited to 44,563 in the future in order not to destroy the earth.

References

[1] Shan YZ, Xu HC, Shan WZ, et al. Multiple linear regression prediction method for submunition drop point dispersion [J]. Journal of Nanjing University of Science and Technology (Natural Science Edition), 2013, 37(5): 720-724.

[2] Zhang Lijin,Wei Xia Yan,Lu Jiaqi,et al. Lasso regression:from explanation to prediction[J]. Advances in Psychological Science, 2020, 28(10): 1777-1788, Post-Insert 1-Post-Insert 3.

[3] Z. H. She, J. H. Xu, C. Y. Yao, et al. Analysis of online shopping big data based on Pearson's correlation coefficients--Tenmall Bailunju flagship store transaction records as an example [J]. Journal of Hanshan Normal College, 2020, 41(3): 16-22.

[4] Sa-Graston. Nuclear weapons effects [M]. Beijing:National Defense Industry Press, 1966.

[5] Qiao Dengjiang, ed. Introduction to nuclear explosion physics [M]. Beijing: Atomic Energy Publishing House, 1988.05.