Analysis of Human Height Based on Footprint Multivariate Regression Analysis

Wenbo Yang
Department of Forensic Science & Technology, Criminal Investigation Police University of China, Shenyang 110854, China
ywb1029@163.com

Abstract. In the process of investigating criminal cases, it is usually necessary to infer the suspect's height according to the footprints on the scene. The current common method is to multiply the length of barefoot footprints by a coefficient of 7, and the accuracy of the results obtained by this method is less than 70% in the positive and negative 3cm of the interval. In this paper, the average foot length of 1200 people is collected by using mathematical statistics analysis method, and the average stride length of each person is added to the multiple regression analysis. The better results are obtained, which makes the accuracy rate of body height in positive and negative 3cm increase to about 87%.

1. Introduction
Footprint is a kind of trace material evidence with high occurrence rate in crime scene. The analysis of the time, the number and the nature of the case can provide direction for the investigation and narrow the scope of the investigation. The footprint analysis of the suspect's natural conditions, such as height, age, posture, sex, occupation and walking posture, is conducive not only to determining the key points, but also to identifying the person under certain conditions, providing expert opinions for litigation and trial. In most cases, the suspect's physiognomy is the first concern of investigators, especially the suspect's height.

Documents [1-3] introduced a variety of methods for analyzing suspect's height. Documents [1] listed several commonly used methods for analyzing suspect's height. Documents [2] gave a method for analyzing the height of incomplete footprint. Documents [3] gave a method for analyzing the height based on multiple features of barefoot. These methods have their own application background and advantages. In this paper, we try to add the factor of step characteristics to the analysis of height, and hope to improve the accuracy of height analysis.

2. Data Processing and Model Building
Under laboratory conditions, 1200 people were subjected to standardized barefoot stamping, step feature measurement and personal information collection. Firstly, the average foot length was calculated by using the measured left and right barefoot footprint length. Then, this group of data was processed, descriptive statistical analysis of height and foot length was made, and linear regression of height y and foot length x was made, and no intercept term was specified. The regression results are as follows:
Table 1 Linear regression of height y and foot length x without intercept term

| Parameter estimation |
|----------------------|
| Variable | Label | Free degree | Parameter Estimated value | standard error | t value | Pr > |t| | 99% Confidence limits |
|----------|-------|-------------|---------------------------|----------------|--------|--------|------------------------|
| x        | x     | 1           | 7.05528                  | 0.00622        | 1133.51 | <.0001 | 7.03922                | 7.07134 |

As a result: \( y = 7.05528x + \varepsilon \), here, \( \varepsilon \) is the error term. And the goodness of fit of the model is 0.9991, so we can see that the model seems reasonable.

When the range of height prediction deviates from the observed value more than 3 cm in advance, the model prediction is considered to be unsuccessful or successful. Therefore, when judging the accuracy of the prediction model, we must introduce a 0-1 variable \( Z \). When the prediction fails, \( Z \) takes a value of 0, and when the prediction succeeds, \( Z \) takes a value of 1. Thus, using the \( Z \) value of each sample in the data set, the mean value represents the prediction success rate of the model. The prediction results of the model with no intercept term are as follows:

Table 2 The prediction results

| Basic Statistic Measure \((z)\) |
|---------------------------------|
| Position | Variability |
|---------|-------------|
| mean value | 0.560100 | Standard deviation | 0.49658 |
| Median   | 1.000000 | Variance           | 0.24659 |
| Mode number | 1.000000 | Range              | 1.00000 |
|         |           | Quartile spacing   | 1.00000 |

It can be seen that when the criteria for predicting and judging accuracy are given beforehand, the prediction accuracy under this model is 56.01%, slightly higher than 50%, and the results are not very satisfactory.

In view of the limitations of the above models, we can improve the model from two aspects: one is to add intercept term to the original linear model, the other is to choose another model to regression height y and foot length x.

1) Add intercept term. The regression results are as follows:

Table 3 Linear regression of height y and foot length x with the intercept term

| Parameter estimation |
|----------------------|
| Variable | Label | Free degree | Parameter Estimated value | standard error | t value | Pr > |t| | 99% Confidence limits |
|----------|-------|-------------|---------------------------|----------------|--------|--------|------------------------|
| Intercept| Intercept | 1       | 89.87483                  | 2.69691        | 33.33  | <.0001 | 82.91693                | 96.83272 |
| x        | x     | 1           | 3.45226                  | 0.10821        | 31.90  | <.0001 | 3.17308                 | 3.73143 |

We can know that the linear equation of height y and foot length x is

\[ y = 89.87483 + 3.45226x + \varepsilon \]

here, \( \varepsilon \) represents the error term. The prediction accuracy of the model is as follows:
Table 4 The prediction results

| Basic Statistical Measure (z) | Position | Variability |
|-------------------------------|----------|-------------|
| mean value                    | 0.599332 | Standard deviation 0.49024 |
| Median                        | 1.000000 | Variance 0.24033 |
| Mode number                   | 1.000000 | Range 1.00000 |
| Quartile spacing              |          | 1.000000   |

The prediction accuracy of height under this model is 59.93%, which is slightly higher than 56.01% under no intercept. But it is still not obvious.

Another model is chosen to predict. Consider transforming height y into height growth rate log (y), foot length into foot growth rate log (x), and cross regression. The accuracy of regression is as follows:

Table 5 The accuracy of regression

|                          | foot length x | foot growth rate log (x) |
|--------------------------|---------------|--------------------------|
| height y                 | 59.93%        | 59.77%                   |
| height growth rate log (y)| 59.85%        | 59.68%                   |

It can be seen that the accuracy of the alternative model has not been improved, and the linear regression with intercept terms of height y and foot length x still has the highest accuracy in height prediction. When adjusting the range of fluctuation, the accuracy of prediction is 42.82% and 20.95% after adjusting 3 cm to 2 cm and 1 cm respectively.

After eliminating two outliers from the original data, the data are arranged in ascending order and scatter plots are made as follows:

Fig. 1 Scatter plots

The linear relationship between height y and foot length x is obvious, and it is unreasonable to use non-linear regression. Therefore, we consider adding the characteristics of the sample tester's stride length to the prediction model. Traditionally, the taller the longer the stride, the smaller the shorter the
stride. However, the accuracy rate of judging the height by the stride length alone is very low. If combined with the foot length, we hope to get a better result.

Predicting height by foot length and stride length. There are no abnormal values. The results of linear regression are as follows:

Table 6 Multivariate regression model results

| Variable | Label | Free degree | Parameter Estimated value | standard error | t value | Pr > |t| | 99% Confidence limits |
|----------|-------|-------------|---------------------------|----------------|---------|-------|-------|-----------------------|
| Intercept | Intercept | 1 | 70.40770 | 2.88563 | 24.40 | <.0001 | 62.94309 | 77.87231 |
| x1 | x1 | 1 | 3.17913 | 0.14697 | 21.63 | <.0001 | 2.79893 | 3.55932 |
| x2 | x2 | 1 | 0.32783 | 0.02315 | 14.16 | <.0001 | 0.26796 | 0.38771 |

We can get the regression function:

\[ y = 70.40770 + 3.17913x_1 + 0.32783x_2 + \varepsilon \]

here, y stands for height, x1 stands for foot length, x2 stands for stride, \( \varepsilon \) is a standard normal distribution. According to 3cm standard, the accuracy rate is 87.39%. According to 2cm standard, the accuracy is 68.14%. According to 1 cm standard, the accuracy is 37.61%.

3.Conclusion

In criminal cases, footprint analysis is widely used to analyze the age and height of suspects, and plays a vital role in the investigation and resolution of cases. However, due to the limitation of some methods, there will be a big error in estimating the height from the foot length, or even a very far difference from the real height, leading to the wrong direction of investigation. Therefore, scholars have been engaged in the work of estimating height at full length for a long time. In this paper, the factor of step feature is added to the traditional one-dimensional regression model. Although it effectively improves the accuracy of height, it is only applicable to those places that leave both step feature and foot feature, and can not be applied to all situations. Therefore, the researcher sincerely hopes that more scholars will join us in future research and get more abundant and effective results.

Acknowledgments

This work was financially supported by the Application of Multivariate Analysis in Barefoot Footprint Analysis (2015).

References

[1] Chengqing Tang, Jiaying Dong, Li Yao, Limin Shi, Analysis on Body Herght[J], Journal of Liaoning Police Academy, No. 5(Sum.33), Sept. 2005.
[2] Limin Shi, Yuting Zhu, Maosen Ban, Study on the Correlation Between Height and Measurements of Different Parts of Footprint[J], Journal of Criminal Investigation Police University of China, No. 6, Serial 140, 2017.
[3] Shennan Qiao, Measure the Bare Footprint Though Using Multiple Regression Method to Analyze Homogeneous Population Height[J], Journal of Hunan Police Academy, Vol. 29, No. 5, Oct. 2017.
[4] Kewal Krishan, Tanuj Kanchan, Heel-Ball(HB) Index: Sexual Dimorphism of a New Index from Foot Dimensions[J]. Forensic Sciences, 2012(1):172-175.
[5] Krishan K. Estimation of stature from footprints and foot outline dimensions in Gujjars of North India[J], Forensic Science, 2007(2-3):93-101.