Dean&Francis

ISSN 2959-6157

Linear Regression Analysis Between the Height of Parents and Children

Yuetong Hao

Chongqing Nankai Secondary School, Chongqing, 400030, China

Z5596448@ad.unsw.edu.au

Abstract:

Height is a subject of interest in understanding familial relationships and growth patterns. A sample of 20 boys and 20 females, ages 15 to 19, together with their parents, were gathered for the study, which examines the link between parents' height and that of their offspring. By treating the parents' height as the independent variable (X) and the children's height as the dependent variable (Y), regression models were developed to predict height differences across genders. The models were also cross-tested to compare their predictive power. The optimal regression equations yielded relatively low R^2 values: 0.096 for the height of father and son, 0.122 for mother and son, 0.078 for father and daughter, and 0.091 for mother and daughter. Despite the relatively low R^2 values, these findings underscore the complex nature of height inheritance, suggesting that multiple genetic and environmental factors contribute to height variations. To give a more thorough picture of the factors influencing height, subsequent research might examine these extra variables.

Keywords: height; linear regression model; correlation analysis.

1. Introduction

Because height is a quantitative trait and has many influencing factors, the study of genetics has been progressing slowly. Early studies focused on the heritability of height, and little is known about the number and localization of height-related genes. In recent years, the genetic studies of height have evolved dramatically with the completion of the human genome project and the application of statistical and molecular biology methodologies. [1] The average height of Chinese men and women between 18 and 44 was 169.7 cm and 158 cm, respectively, according to a study on the Nutrition and Chronic Diseases Status of Chinese Residents 2020 [2]. A person's height and family genetic factors account for 60%~80%, the remaining 20%~40% rely on acquired factors, such as nutrition, sleep, exercise, environment, disease, mood, and so on [3]. If genetic factors cause parents to be short, their children's future height growth potential may be relatively small [4]. However, it should be noted that short parents are not equivalent to low genetic height.

Based on the multivariate linear regression model, this study aims to analyze the following questions with known data: the height of the children signifi-

cantly correlates with the height of the mother and father. Is there a linear relationship between the height changes of father and son, father and daughter, mother and son, and mother and daughter; Solve the best fit linear equation between each group.

2 Definition

Using a minimum square function known as a linear regression equation, linear regression is a regression strategy that models the connection between one or more independent and dependent variables. This function creates a linear combination of the regression coefficients, which represent one or more model parameters. Simple regression occurs when there is a singular independent variable, whereas multiple regression involves numerous independent variables.

Linear regression employs a linear prediction function to model the data while simultaneously estimating the unknown model parameters from the data. These models are called the linear models. Based on the historical sample data, establish the prediction model of multivariate linear regression, and predict the regression parameters in the future time [5]. The primary linear regression model asserts that the conditional mean of y, given a certain value of x, is an affine function of x. Instead of focusing on the joint probability distribution of x and y, linear regression highlights the conditional probability distribution of y given a certain value of x.

When fitting linear regression models, the least squares approximations are commonly used; alternatively, they might use techniques like minimal absolute error regression, which involves minimizing the "fit defect" in certain specifications, or the least squares loss function in bridge regression. Instead, those nonlinear models may be tailored to using least squares approximations. There is no way to connect "least squares" and "linear models," even if the two concepts are closely related.

An approach to statistical analysis that use the concepts of regression analysis to determine the quantitative connection between two or more variables. Compared with the fitting and blessing value methods, the regression method can not only get the function expression of the close data points but also get the statistical information of the deviation, such as the mean and variance of the deviation.

The advantage is that it is good at acquiring linear or nonlinear relationships in the data set, where some well-defined variables are available and a simple predictive model is needed. Its prediction is fast, performs well on small datasets, with interpretative results, and is easy to illustrate.

3 Materials and Methods

3.1 Investigated Subject

The sample was from Chongqing, China, with 40 high school students aged from 15 to 19, including 20 males and 20 females.

3.2 Methods

①Make a questionnaire;
②The questionnaire was randomly distributed to 40 families.
③Collect and organize data;
④Analyze the data according to the least squares method, and analyze and verify the obtained conclusions.

3.3 Process

①Collect data;
②Use the linear regression mapping method;
③ Using the least squares method, the most fitting regression line equation was calculated and the linear regression analysis was conducted;
④discuss the results of the analysis, test the association of the conclusions obtained by the results and the actual problems, and form a report.

3.4 Statistical Analysis of the Data

The Excel in the WPS was used to record and preliminarily organize the questionnaire clerk, and descriptive statistics were made for the height data with the SPSS 27 software. The association of height with parental height was analyzed using correlation analysis. The optimal regression equation was established by stepwise regression analysis in multiple regression analysis.

4. Results and Analysis

4.1 The Height of the Son and Corresponding Father

ISSN 2959-6157

Table 1. Result of the Linear Regression Analysis (LRA) (*n*=20)

	Non-standardized Coefficients (NSC)		Standardized Coefficient (SC)	*T*	*P*	Collinearity Diagnostic (CD)	
	B	Standard Error (SE)	*Beta*			VIF	Tolerance (T)
Constant	131.102	32.950	-	3.979	0.001**	-	-
Father's Height	0.272	0.191	0.311	1.424	0.171	1.000	1.000
R 2	0.096						
Adjust R 2	0.049						
F	*F*(1,19)=2.028, *p*=0.171						
D-W Value	1.599						
Note: Dependent Variable (DV) = the Height of Son							
*p<0.05 **p<0.01							

According to Table 1, the influence of father's height on the son's height was not significant (p = 0.171), and the standardized coefficient (Beta = 0.311) indicated a positive correlation, but not statistically significant. Furthermore, the model does not show a multicollinearity problem. Y = 131.102 + 0.272X.

4.2 The Height of Son and Corresponding Mother

Table 2. Result of the LRA (*n*=20)

	NSC		SC	t	p	CD	
	B	SE	*Beta*			VIF	Tolerance
Constant	124.598	32.931	-	3.784	0.001**	-	-
Mother's Height	0.331	0.204	0.349	1.622	0.121	1.000	1.000
R 2	0.122						
Adjusted R 2	0.075						
F	*F*(1,19)=2.632, *p*=0.121						
D-W	1.721						
Note: DV = the Height of Son							
* p<0.05 **p<0.01							

Table 2 indicated that maternal height was not statistically significant (p = 0.121) and had minimal explanatory power within the model (R^2 = 0.122). Adjusted R^2 = 0.075, signifying that mother's height, as an isolated predictor, had a limited capacity to forecast the height of the son. This was not statistically significant despite a positive relationship (Beta = 0.349). Y = 124.598 + 0.331X

4.3 The Height of the Daughter and Corresponding Father

Table 3. Result of the LRA (*n*=20)

	NSC		SC	*t*	*p*	CD	
	B	SE	*Beta*			VIF	Tolerance
Constant	113.139	40.573	-	2.789	0.012*	-	-
Father's Height	0.299	0.236	0.279	1.267	0.221	1.000	1.000
R 2	0.078						
Adjusted R 2	0.029						
F	*F*(1,19)=1.605, *p*=0.221						
D-W Value	1.986						

	NSC		SC	*t*	*p*	CD	
	B	SE	*Beta*			VIF	Tolerance
Note: DV = Height of Daughter							
*p<0.05 **p<0.01							

According to Table 3, the influence of paternal height on the daughter's height was not statistically significant (p = 0.221). In the study with a sample size of 20, the paternal height coefficient (0.299) and the standardized coefficient (Beta = 0.279) were small, and there was no collinearity problem (VIF = 1.000). Y = 113.139 + 0.299X.

4.4 The Height of the Daughter and the Corresponding Mother

Table 4. Result of the LRA (n=20)							
	NSC		SC	*t*	*p*	CD	
	B	SE	*Beta*			VIF	Tolerance
Constant	93.297	51.648	-	1.806	0.087	-	-
Mother's Height	0.441	0.320	0.302	1.379	0.184	1.000	1.000
R 2	0.091						
Adjusted *R* 2	0.043						
F	*F*(1,19)=1.902, *p*=0.184						
D-W Value	2.435						
Note: DV = Height of Daughter							
*p<0.05 **p<0.01							

According to Table 4, the study found that the correlation between maternal height and daughter height was not statistically significant (p = 0.184) with a sample size of 20, despite the maternal height coefficient (0.441) and the standardized coefficient (Beta = 0.302) suggesting a positive association. Furthermore, the model does not show multicollinearity problems (VIF = 1.000). Y = 93.297 + 0.441X.

4.5 The Height of Son and Corresponding Parents

Table 5. Result of the LRA

	NSC		SC	*t*	*p*
	B	SE	*Beta*		
Constant	86.563	41.618	-	2.080	0.052
Father's Height	0.186	0.190	0.212	0.975	0.343
Mother's Height	0.370	0.225	0.357	1.642	0.118
R 2	0.214				
Adjusted *R* 2	0.127				
F	F=2.453, p=0.114				
D-W Value	1.594				
Note: DV = the Height of Son					
* *p*<0.05 ** *p*<0.01					

The model's explanatory power (R^2 = 0.214) suggests that parental height accounts for about 21.4% of the variance in their son's height, whereas the adjusted R^2 of 0.127 reflects a reduction in this proportion after accounting for sample size (Table 5). The comprehensive significance test of the model (F = 2.453, p = 0.114) failed to achieve the conventional threshold for statistical significance (p > 0.05), indicating that the model may lack adequate sta-

ISSN 2959-6157

tistical support for the linear correlation between parental height and son's height. The constant term (B = 86.563, t = 2.080, p = 0.052) was close to the significant level, but the effects of father's height (B = 0.186, t = 0.975, p = 0.343) and mother's height (B = 0.370, t = 1.642, p = 0.118) were not statistically significant. Furthermore, the D-W value of 1.594 is proximate to 2, signifying the absence of significant auto-correlation among the residuals.

4.6 The Height of the Daughter and Corresponding Parents

Table 6. Result of the LRA

	NSC		SC	*t*	*p*
	B	SE	*Beta*		
Constant	36.242	65.462		0.554	0.587
Father's Height	0.314	0.230	0.293	1.370	0.188
Mother's Height	0.460	0.313	0.315	1.471	0.159
R 2	0.177				
Adjusted *R* 2	0.085				
F	F=1.933, p=0.174				
D-W Value	2.145				
Note: DV = Height of Daughter					
*p<0.05 **p<0.01					

According to Table 6, the model has a low explanatory power with an R2 value of 0.177, indicating that parental height accounts for just 17.7% of the variance in the daughter's height. The corrected R2 value is 0.085, which further suggests that the model's explanatory power is constrained. The model did not reach a statistically significant level, according to the model's overall significance test (F = 1.933, p = 0.174), suggesting that there was insufficient statistical support for the linear association between the heights of the parents and daughters. Specifically, the predictor variables, constant term (B = 36.242, t = 0.554, p = 0.587), father's height (B = 0.314, t = 1.370, p = 0.188) and mother's height (B = 0.460, t = 1.471, p = 0.159) did not reach statistical significance for daughter height. The D-W value of 2.145, which is near to 2, indicates that the model's residual sequences do not significantly auto-correlate.

5. Conclusion

Although not statistically significant, the positive association between parental height and child height (positive Beta coefficient) was seen in all models. This aligns with prior findings indicating that height remains affected by genetic factors. Although studies have shown that parental height is a highly correlated variable, child height may be influenced by other variables not considered in the model, such as nutrition, lifestyle, and environmental factors. In addition, Max pointed out that regional differences can significantly affect the variation in average human height. The reason why the results of this paper are not significant may be that the sample size is too small and not representative, and the children are in the age range of 15 to 19 years old, which may not be fully mature. Future studies could consider collecting larger sample sizes and adding populations of different age stages to improve detection accuracy. It is suggested that future studies construct more complex multifactorial models to assess whether the relationship between these factors and height is significant.

References

[1] Huo S, Zheng X F. Research progress on the genetic genes influencing human height development. Journal of Chengde Medical University, 2006, (04): 415-418.

[2] Nutrition and Chronic Diseases Status of Chinese Residents 2020.

[3] Pan C L. Height growth: Seven parts are determined by nature, three parts depend on effort. Parents Must Read, 2024, (4): 46-51.

[4] Wang Y Q. Parents' height does not necessarily influence children's height. Chinese Health Care, 2013, (9): 1.

[5] Wang H W, Meng J. Prediction modeling method of multiple linear regression. Journal of Beihang University, 2007, 33(4): 5.