

Research On The Relationship Between Fathers' Height And Sons' Height

Yiqi Liu(Simon),Yuetong Ding(Dennys), Yulin Bai(Lawrence)

26th May 2023

1 Introduction

Genetics is always popular, especially with the interplay between inherited traits and their manifestation in descendants. To be specific, the relationship between parents' characters and those of their children has got a lot of attention. In this research, our team delves into the question of the relationship between fathers' height and sons' height, trying to figure out the inheritance patterns of height.

Understanding the association between fathers' height and sons' height is scientifically significant. Height is an important trait in genetics, studying this can provide a huge insight into human biology or even healthcare, because height also links to a wide range of health outcomes, like cardiovascular disease, diabetes, and cancers. Besides, studying the inheritance of height can help us understand population evolution.

In this project, we expect to find a positive, linear correlation between fathers' height and sons' height. Previous studies have found that a strong relationship is around 60 percent to 80 percent. In other words, the father's height determines the son's height. So, we hypothesize that taller fathers are likely to have taller sons; shorter fathers are likely to have shorter sons.

2 Background Research

According to J M Tanner, H Goldstein, R H Whitehouse, they investigated the standards for children's height at ages 2-9 years allowing for heights of parents. They construct the charts of centile standards for boys and girls with their mid-parental height(the average height of mother and father). The 3rd centile is what they focused on, which is about the diagnosis of genetic small structure. Besides, The correlation coefficients are given at successive ages, from 1 month to 9 years. [1] On the other hand, instead of the children's height, we will focus on the final height that children can grow, calculating the correlation coefficients.

In addition to J M Tanner, Danny Zeevi and several researchers investigated the accurate prediction of children's target height from their mid-parental height. They used height data from 23 large nuclear families, comparing the final height of the children to the height predicted by the standard procedure. Also, they develop an alternative way to calculate the final height which includes the corrections of mid-parental height for age and sex. The results explained 40 percent of the variance in children's height and an inheritance rate of 80 percent, which increased 4 percent from the standard way.[2] Similarly, Pen-Hua Su estimates final height from parental heights and sex in Taiwanese. They developed a simple function for estimating the target height for Taiwanese children. He found out that the increases in height between the two generations were 1.49-3.19 centimeters for boys and 2.03-2.061 centimeters for girls. So, they built a new linear model to fit their data, increasing the accuracy of the target height in Taiwanese children.[3]

For our research, we intend to build our own linear model which is more precise and accurate, predicting the target height for boys.

3 Methods And Procedures

3.1 Data Collection

We collected the data through the Internet(Kaggle website). you can look at the data through [Data](#)

	Father (inch)	Son (inch)	Father (cm)	Son (cm)
1				
2	65	59.8	165.1	151.892
3	63.3	63.2	160.8	160.528
4	65	63.3	165.1	160.782
5	65.8	62.8	167.1	159.512
6	61.1	64.3	155.2	163.322
7	63	64.2	160	163.068
8	65.4	64.1	166.1	162.814
...
1072	73.3	67.9	186.2	172.466
1073	65.8	61	167.1	154.94
1074	67.7	59.8	172	151.892
1075	67	70.8	170.2	179.832
1076	71.3	68.3	181.1	173.482
1077	71.8	69.3	182.4	176.022
1078	70.7	69.3	179.6	176.022
1079	70.3	67	178.6	170.18

Figure 1: Fathers' height and Sons' height in inches and centimeters

The figure above gives the heights of fathers and their sons, based on a famous experiment by Karl Pearson around 1903. The number of cases is 1078. Random noise was added to the original data, to produce heights to the nearest 0.1 inch.

3.2 Inference

Because genetic factors are the most important when considering height, the height of the son is regarded as the dependent variable, and the corresponding father's height is the independent variable. After testing that the dependent variables are normally distributed and each dependent variable is independent, linear regression is performed. First, we ran a hypothesis test. H_0 is ρ equals zero, H_a is ρ not equal to 0. ρ here represents the linear correlation coefficient between the height of the father and the height of the son. As for the condition of the t-test: our data from the website is randomly selected, and it is bigger than 30; the dependent variable follows the normal distribution, which is supported by the box plot below; each of the data of the son's height is independent of each other. After t-testing with Python code, we will compare the resulting p-value with alpha. If it's less than alpha, then we can reject the null hypothesis and prove that H_a is true.

4 Results

According to Figure 2, the R square obtained by regression is 0.9983, indicating that 99.83 percent of the height variation of sons in these samples is caused by the height variation of fathers. As you can see from Figure 3, the residuals are evenly distributed above and below 0. The data calculated by the code also prove that the residual satisfies the normal distribution. Besides, the values below show that the correlation coefficient is about 1.0139, the constant is about 0.0001. Furthermore, the T-static value is 18.997, and the p-value is extremely small so the computer automatically estimates it to be 0. With this $p\text{-value}=0, \alpha=0.05$, we ensure that there is a linear relationship between the fathers' height and the sons' height.

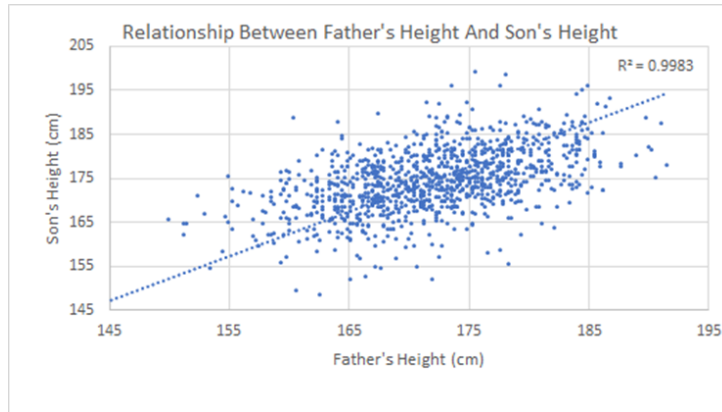


Figure 2: Relationship Between Fathers' Height and Son's Height

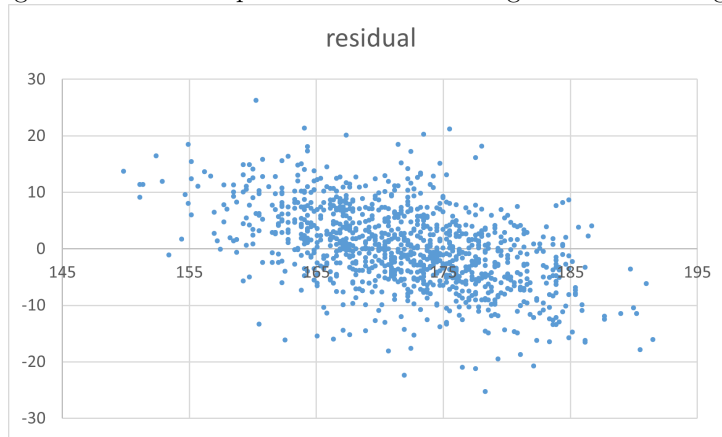


Figure 3: Residuals

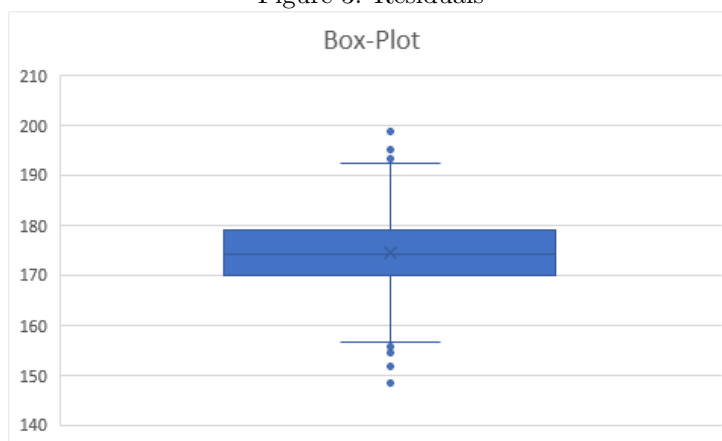


Figure 4: Box-plot of Fathers' Height and Sons' Height

5 Problems In The Projects

5.1 Limitations

Although we find a linear relationship between fathers' height and sons' height, there are still some limitations of our research. First, the data was from 1907 which will cause some bias. Since it was one hundred years ago, the fathers' height and sons' height may be different from nowadays height. As a new generation, people now might be higher than in the past because we have thorough medical equipment, advanced technology, and a rational diet.

5.2 Suggestions For The Future

Besides the limitations, there are also some recommendations for future work in the height area. The height is also influenced by other factors like sleeping time and nutrition. Therefore, in the future, not only investigate how parents' height affects children's height but also find the effect of sleeping and eating. Further work can do With parents' height as a control variable and find out a more accurate inheritance pattern.

6 Conclusion

From the above regression model and residual chart, it is not difficult to see that there is a significant linear relationship between the height of the father and the height of the son. From this, we can conclude that there is a strong positive correlation between the height of fathers and the height of sons.

References

- [1] JM Tanner, H Goldstein, and RH Whitehouse. Standards for children's height at ages 2-9 years allowing for height of parents. *Archives of disease in childhood*, 45(244):755–762, 1970.
- [2] Danny Zeevi, Adi Ben Yehuda, David Zangen, and Leonid Kruglyak. Accurate prediction of children's target height from their mid-parental height. *medRxiv*, pages 2022–10, 2022.
- [3] Pen-Hua Su, Shu-Li Wang, and Jia-Yuh Chen. Estimating final height from parental heights and sex in taiwanese. *Human biology*, 79(3):283–292, 2007.

[\[3\]](#) [\[1\]](#) [\[2\]](#)