

Exploration of the Relationship between the Forearm Length and the Height

M.Z

September 26, 2020

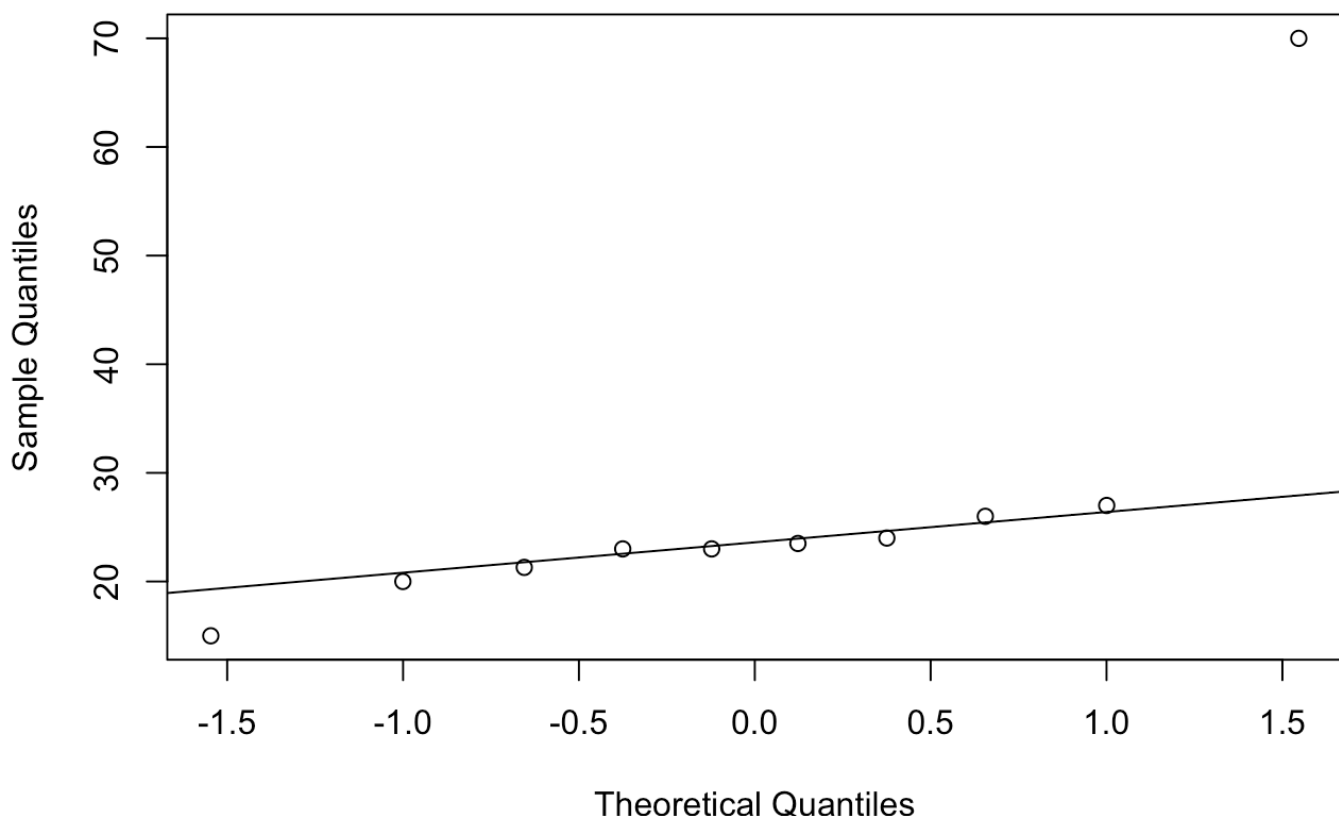
I. Introduction

In this assignment, I explored the relationship between the forearm length and the height. When we were supposed to collect data from our break-out room during the meeting, most of the students didn't know how to measure their forearm, thus didn't provide the data. As the result, the data I used in this assignment is randomly selected from the file a20a1data.csv, the data of which is collected by a survey conducted by the teaching team of STA302.

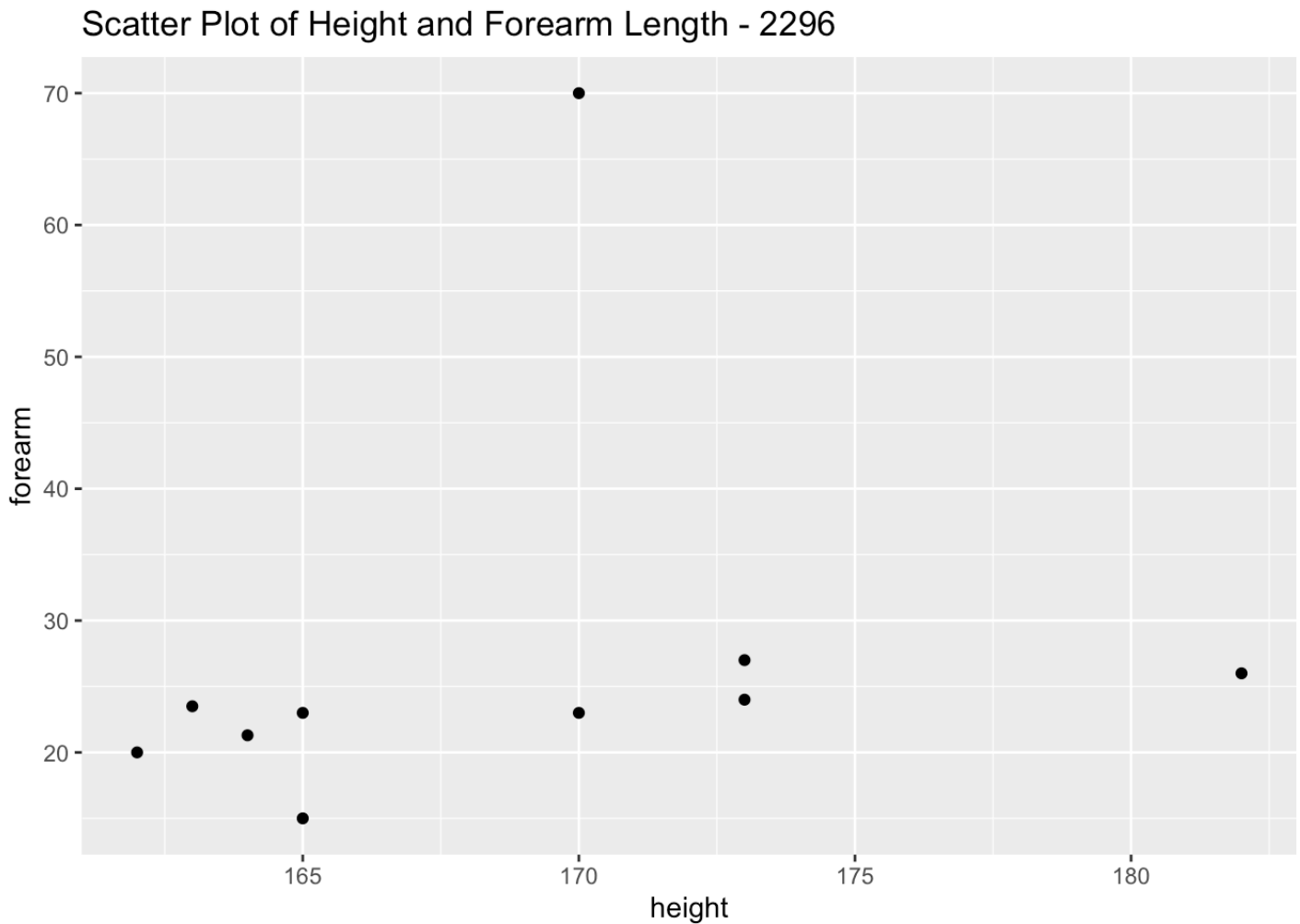
Between the forearm length and the height, the height is the explanatory variable, and the forearm length is the response variable. Because it is more likely that the forearm length is determined by the height, since the height is measured from the whole body, and forearm is only a small part of human body.

II. Exploratory Data Analysis

Q-Q Plot of Forearm Length - 2296



According to the Q-Q plot, we can observe that most of the data points seem to be well inline with percentiles of a normal distribution. The response variable is approximately normal.

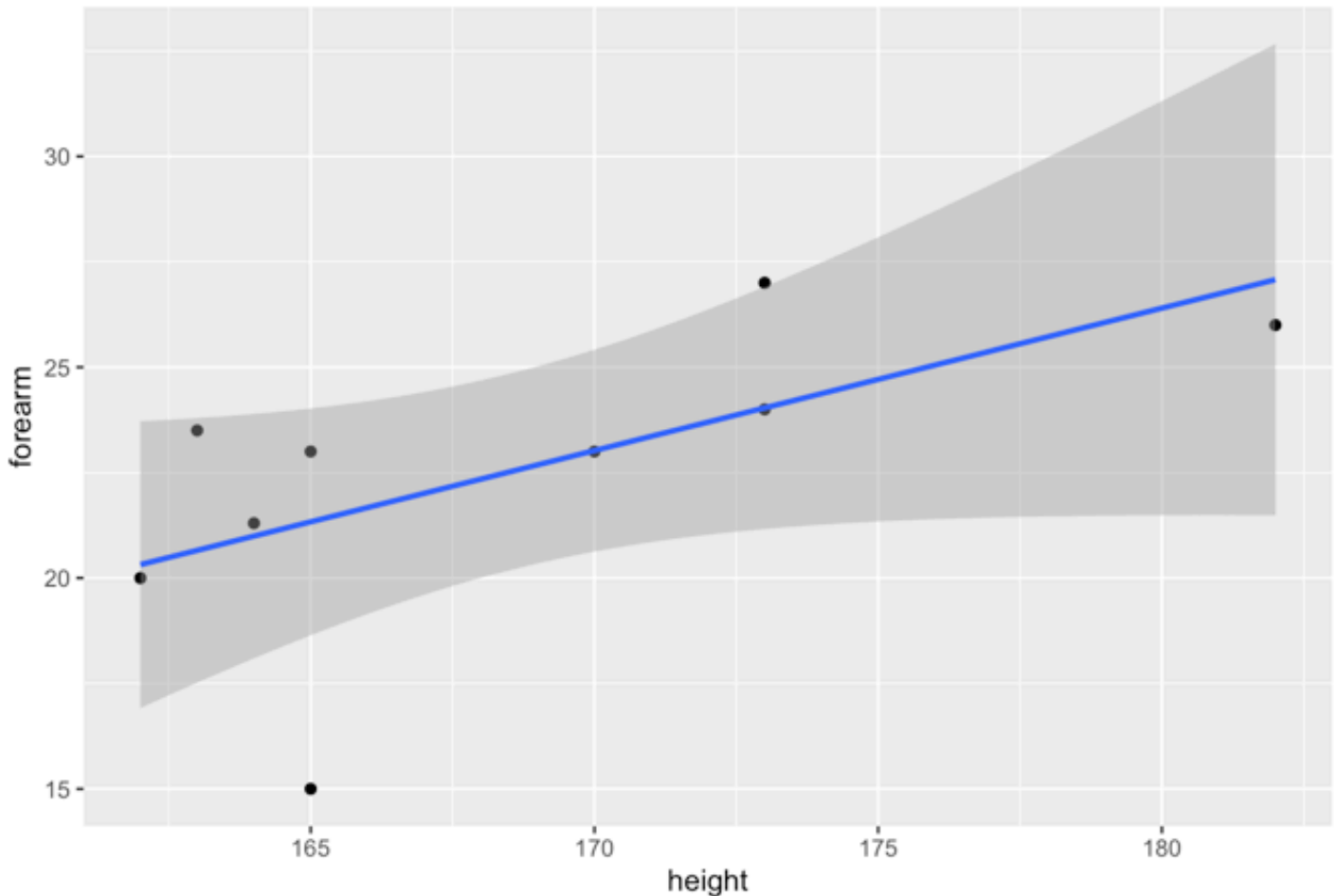


The scatter plot is used to visualize the 10 data points. There is one point being far away from other data points. This point was collected from a student with a 70 cm forearm. It is scientifically not likely for a human to have such a long forearm. As a result, this data point is considered as an outlier. According to the plot, it seems like the relationship between the two variables is linear.

Based on the 10 data points, the center of the data is 27.28 cm, which means that the average forearm length is 27.28 cm. The variance of the data is 236.4396. This large number indicates that the data pointes are very spread out from the mean. However, this might be the result of the outlier data point.

III. Methods and Model

Linear Regression Model of the Height and Forearm Length - 2296



```
##
## Call:
## lm(formula = forearm_height_1$forearm ~ forearm_height_1$height)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.3319 -0.3183 -0.0214  1.6681  2.9649
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -34.4198    26.9390  -1.278   0.2421
## forearm_height_1$height  0.3379     0.1597   2.116   0.0722 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.955 on 7 degrees of freedom
## Multiple R-squared:  0.39, Adjusted R-squared:  0.3029
## F-statistic: 4.476 on 1 and 7 DF, p-value: 0.07219
```

The outlier data point is temporarily removed while checking for the linear model. As the result shows, the regression parameters are not 0. The estimate of the intercept is -34.4198, which means that when a person has a height of 0 cm, his/her forearm length would be -34.4198 cm. This data is not practical in real life, since it is scientifically impossible for a human to have a height of 0 cm. The estimate of the slope is 0.3379, which indicates the relationship between the two variables, the forearm length and the height. It suggests that when a person's height is 1 cm taller, his/her forearm length is expected to be 0.3379 cm

longer. The p-value of the model is 0.07219. Since the p-value is not small enough, there is insufficient evidence to reject the null hypothesis, which is that there is no relationship between the forearm length and the height.

IV. Discussions and Limitations

Potential lurking variable may exist. One possible lurking variable might be the genes that determine how long people's each body part would grow. Genes which contain hereditary information of having higher height and longer arms would make some people taller and have longer forearm. At the same time, other people might have relatively lower height and shorter forearm, since their genes contain information of having those. Genes are the cause of both having lower/higher height and shorter/longer forearm length. People might overlook this factor and be affected while checking for the relationship between forearm length and height.

There are also some limitations of my fit. First, many students didn't know how to measure their forearm length properly, which led to getting inaccurate measurement. And these measurements ended up being the outlier data points, as the one outlier in my data. Second, when we were collecting the data, we only asked the students in this specific course to do the survey. The result might not be representataive for the whole population.

Another pair of variables I would like to explore the simple linear regression model of is the times people go out per week and the probability of them getting COVID-19. The response variable is the probability of people getting infected by COVID-19. Because when people go outside, there is a higher chance for them being exposed to the virus, which leads to the higher possibility of getting COVID-19. The two variables have a positive linear relationship.