

A Model for Individual Height Estimation from Forearm Length in

XB3003

September 26, 2020

I. Introduction

The aim of the study is to find out the relationship between individual height and length of forearm and to derive linear regression equation to calculate the height from length of ulna

Collection of sample data

Quantitative data collection method was applied where the sample “students” were randomly selected into group of ten students during the course live sections. The data use here was from group #2, data collected was their height and forearm length.

Variables in this Assignment:

The continuous (response) variable of this research was: Height (cm)

The continuous (predictor) variable was: Forearm Length (cm) of the students.

The predictor variable forearm length was selected because individual forearm does not depend on other sources to be measured an individual can measure his fore arm length but the for the height one have to depend on other people or other sources of measurement.

II. Exploratory Data Analysis

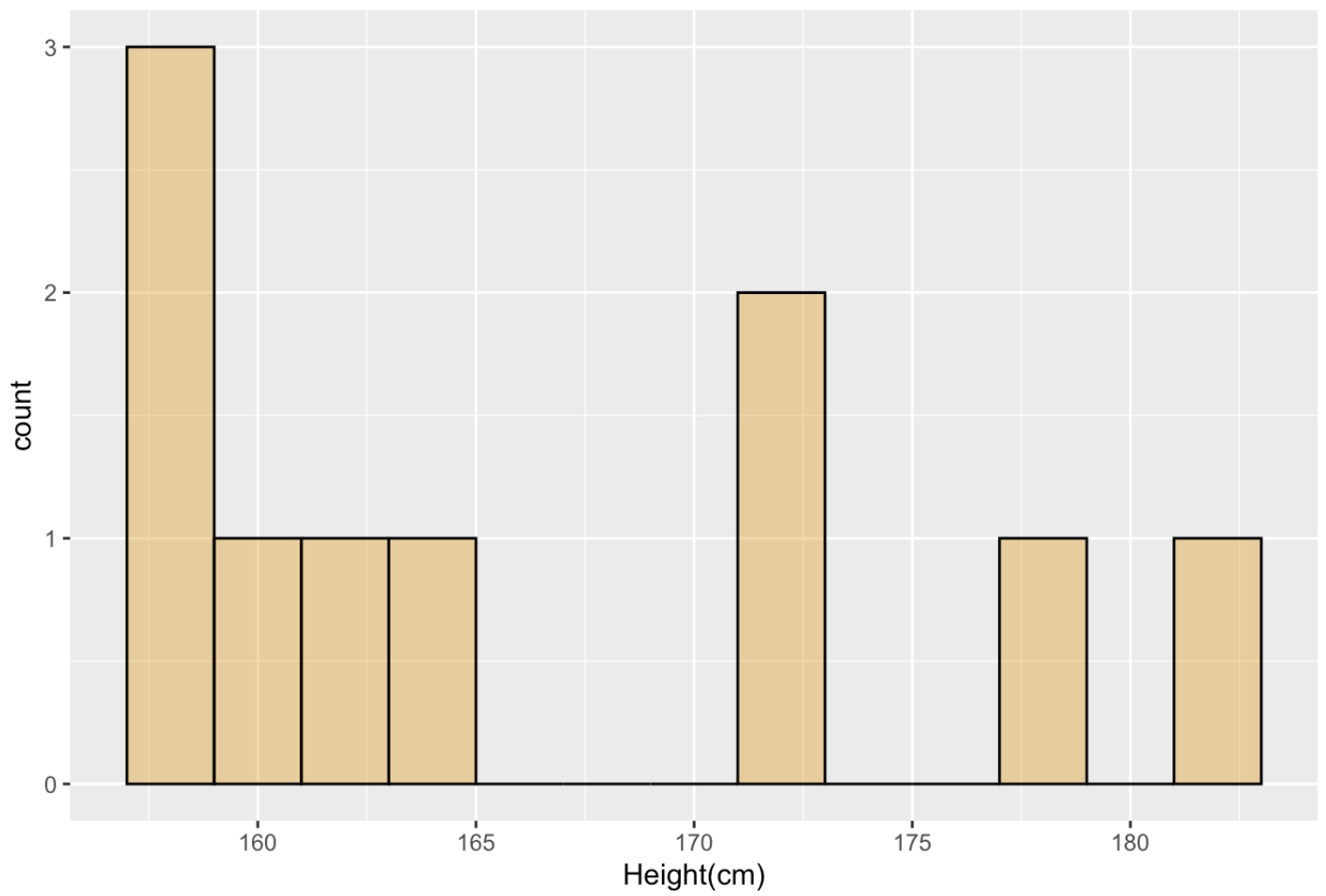
Group 2 Data

##	Height	Forearm
## 1:	172.8	24.0
## 2:	171.2	28.0
## 3:	182.0	26.0
## 4:	157.5	22.0
## 5:	165.0	25.4
## 6:	162.0	23.0
## 7:	158.0	22.0
## 8:	157.5	22.0
## 9:	160.0	23.0
## 10:	177.3	23.0

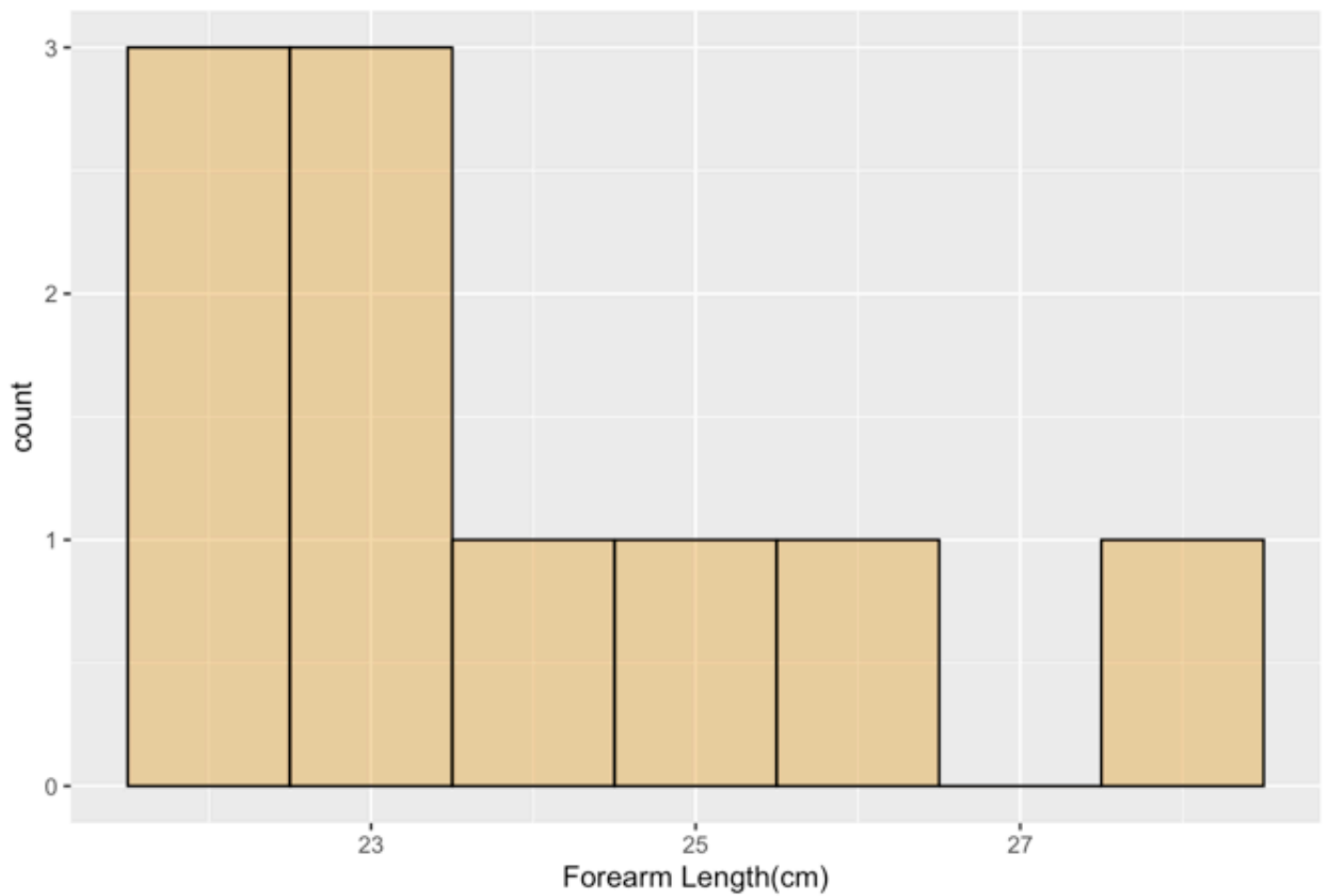
Summary and Histograms

##	Height	Forearm
##	Min. :157.5	Min. :22.00
##	1st Qu.:158.5	1st Qu.:22.25
##	Median :163.5	Median :23.00
##	Mean :166.3	Mean :23.84
##	3rd Qu.:172.4	3rd Qu.:25.05
##	Max. :182.0	Max. :28.00

Histogram of Height #3003



Histogram of Forearm Length #3003

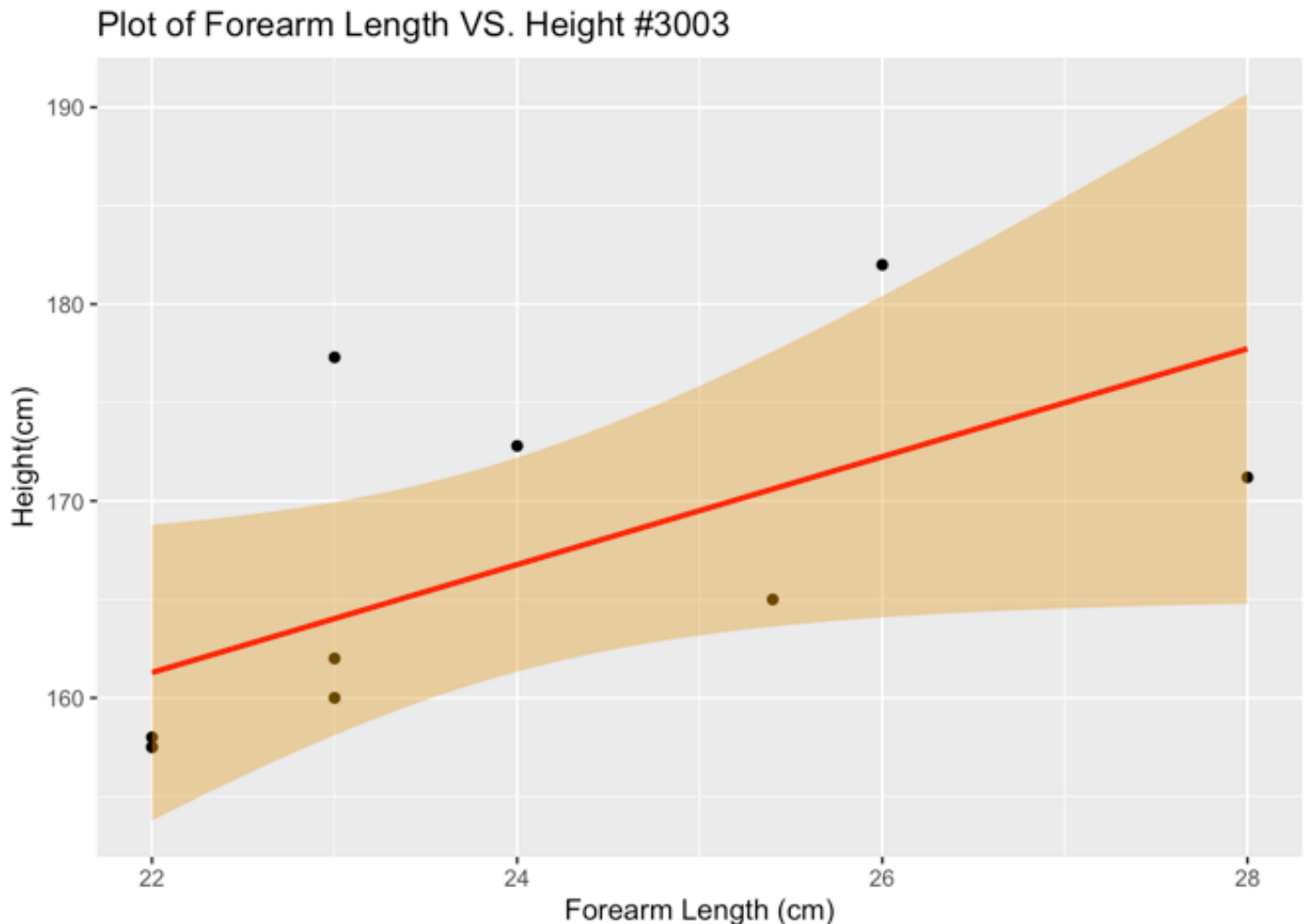


Height Distribution

This distribution seems to be bimodal with 2 clusters, they are both skewed right. The cluster in the right having a medium of about 161cm and the cluster in the right having a medium of about 177cm. The overall range is 24.5, the left cluster has a range of 7.5cm and right one has a range of 10.8cm. Obviously there has no outlier.

Forearm Length Distribution

This distribution is not normal, it is skewed right, with medium of 23cm and range 6cm.



The relation between forearm length and height have a moderate positive linear relationship

The point with forearm length 23cm and height 177.3cm falls far above the line, it could be a potential outlier, with forearm length 23cm we would expect a height of closer to 164cm.

A random pattern of residuals supports a linear model.

There are 3 points that falls outside the Linear regression 95% confidence band

```
##
## Call:
## lm(formula = Height ~ Forearm, data = group2data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.535  -3.967  -3.535   4.017  13.273
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   100.970      29.327   3.443  0.00879 **
## Forearm         2.742       1.226   2.236  0.05579 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.429 on 8 degrees of freedom
## Multiple R-squared:  0.3846, Adjusted R-squared:  0.3076
## F-statistic: 4.999 on 1 and 8 DF,  p-value: 0.05579
```

III. Methods and Model

Model: The Linear Regression (SLR)

The equation of the least squares line of best fit is $\hat{y} = 100.97 + 2.742x$.

Y = Height of the individual

X = Arm length of the individual

$\beta_0 = 100.97$, the Y-intercept, can be interpreted as the value that would predict for Y if X = 0, which is where the line of best fit crosses the height axis. We would expect an average height of 100.97 cm for individual with a forearm length = 0. However, there is no forearm length in the data that is near to zero, it is an extrapolated value.

The slope(β_1) of the line in the scatterplot is 2.74. Thus, we say that each additional centimeter increase is predicted to add 2.742 cm to the height.

Hypotheses:

H₀: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

We have significant level at 0.05

The p-value is 0.0557. The result is not significant at $p < .05$.

Since p-value is greater than 0.05, data failed to reject null hypothesis, Retention of the null hypothesis H₀: $\beta_1 = 0$ indicates no significant evidence that a change in forearm length is associated with a change in height.

The estimates of the regression parameters

$r = 0.62$ indicating a moderate positive linear relationship between the two variables.

R-squared = 0.3846, 38.46% of variability of height can be predicted by the length of forearm.

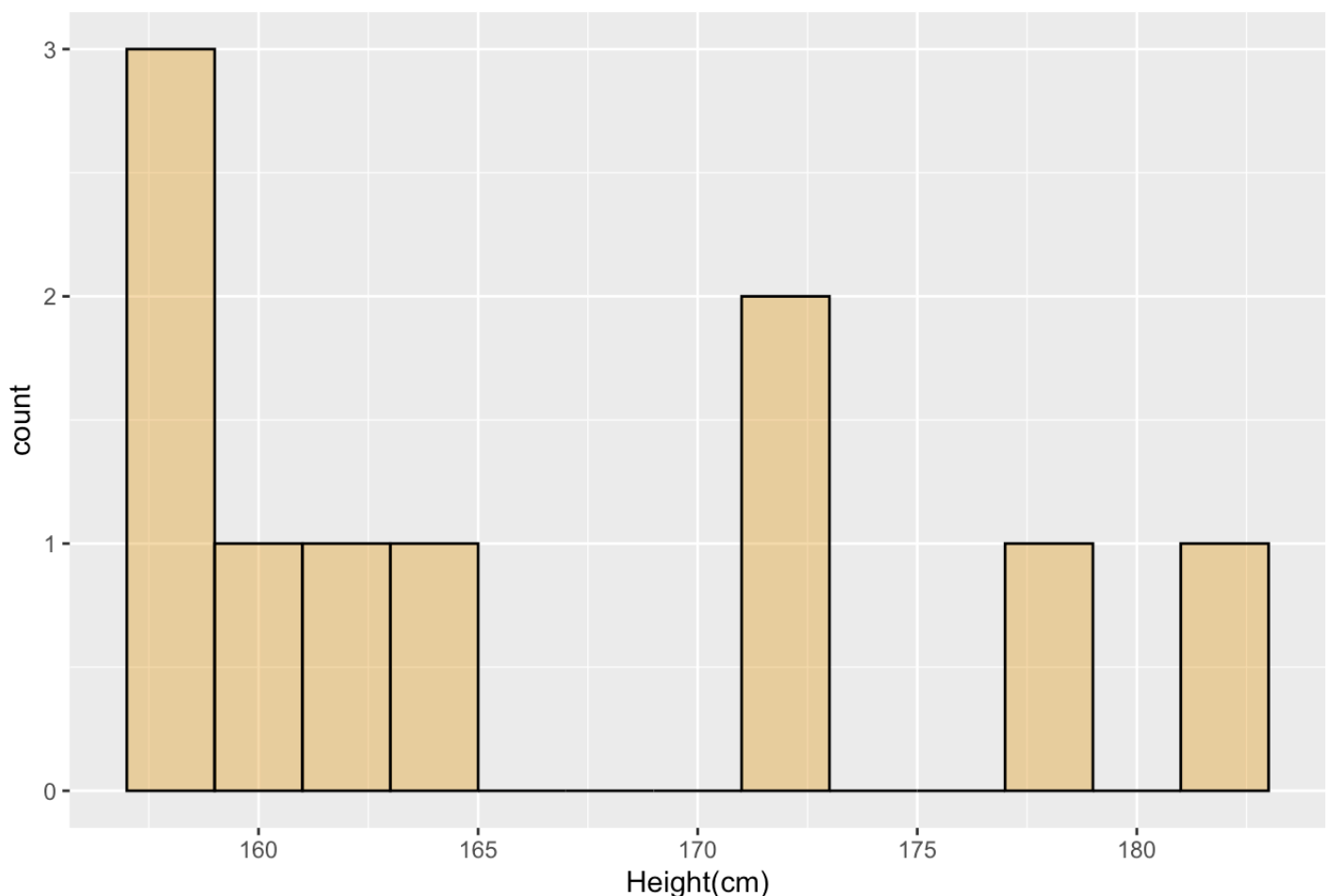
IV. Discussions and Limitations

Limitation

The lurking variable which could have the true relationship of the height and arm length variable is the age, because as the bone growth of an adult stops after 21 year, implying that the ratio between the height and forearm length might remain constant after the age of 21 years

Some of the limitations that could have influenced the outcome of the results is gender and reporting bias. It is factual representation of the height data in the histogram that males tend to be taller than female, we can from the histogram, a bimodal with 2 clusters.

Histogram of Height #3003



accuracy of the self report height and forearm

Variables to explore a simple linear regression model.

The number of hours a student studies affects the score she got on an exam

could be another pair of variables to explore a simple linear regression model. The amount of time studied is the independent variable (continuous), because it's what the student changed, so it's on the x-axis. The score the student got on the exam is the dependent variable (continuous), because it's what changed as a result of the independent variable, and it's on the y-axis.