My Assignment 1 Title (.)

SS1234 (.)

September 26, 2020 (.)

I. Introduction

The aim of this assignment is to find out a regression equation that could calculate the height of a person precisely and reproducibly from forearm length.

Collection of sample data

During the course live section, students were randomly assigned to groups.The data was collected within the groups #2,10 students self-reported their height and forearm length within the group.

Variables in this Assignment:

Dependent variable: Height(cm)

Independent variable: Forearm Length(cm) (The distance between the crease of your elbow and the crease of your wrist on your dominant hand)

The reason that forearm length was choosen to be the independent variable is that individual's forearm length is better to measure compare to height, most people can measure their forearm length without other's help.

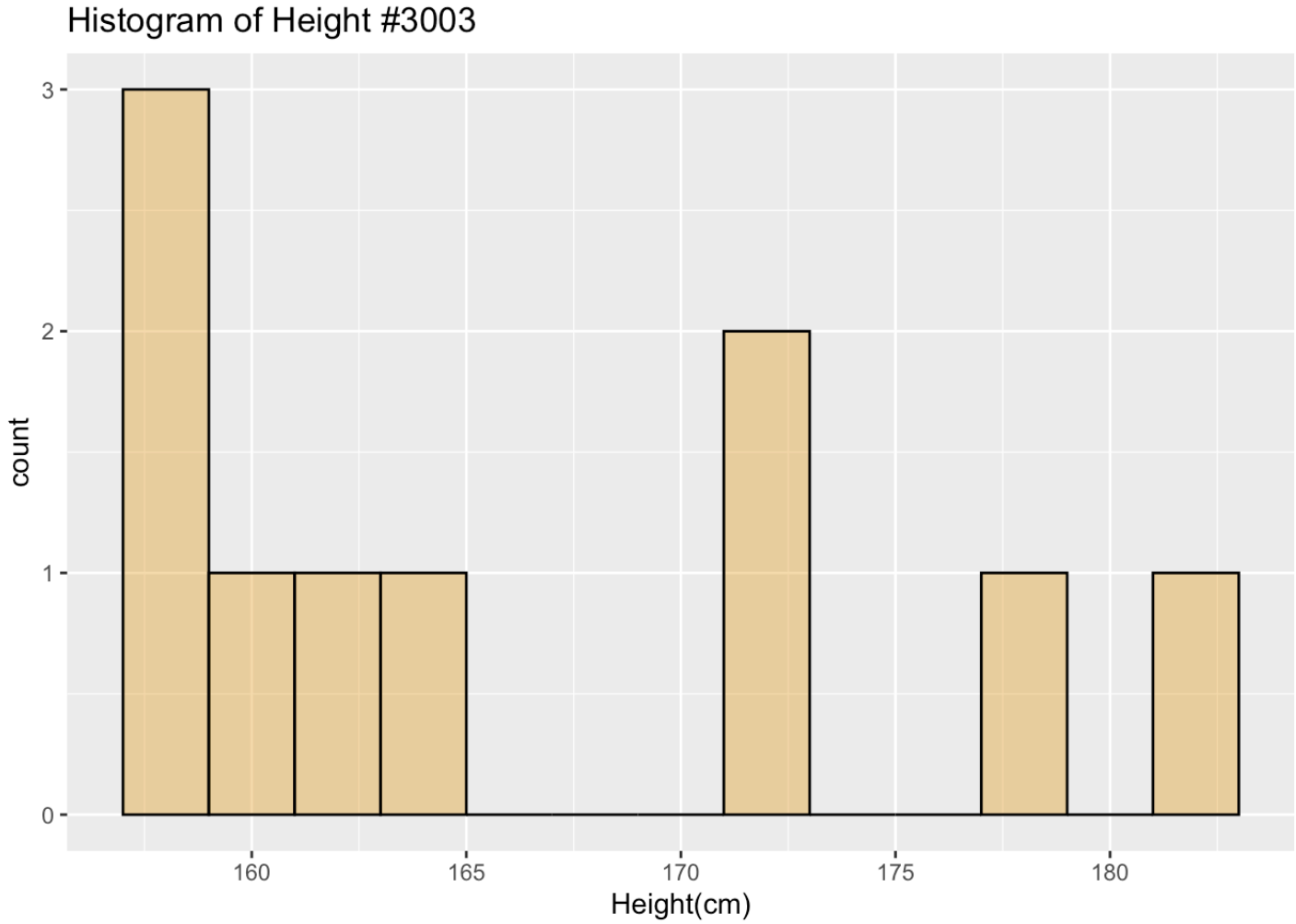
II. Exploratory Data Analysis

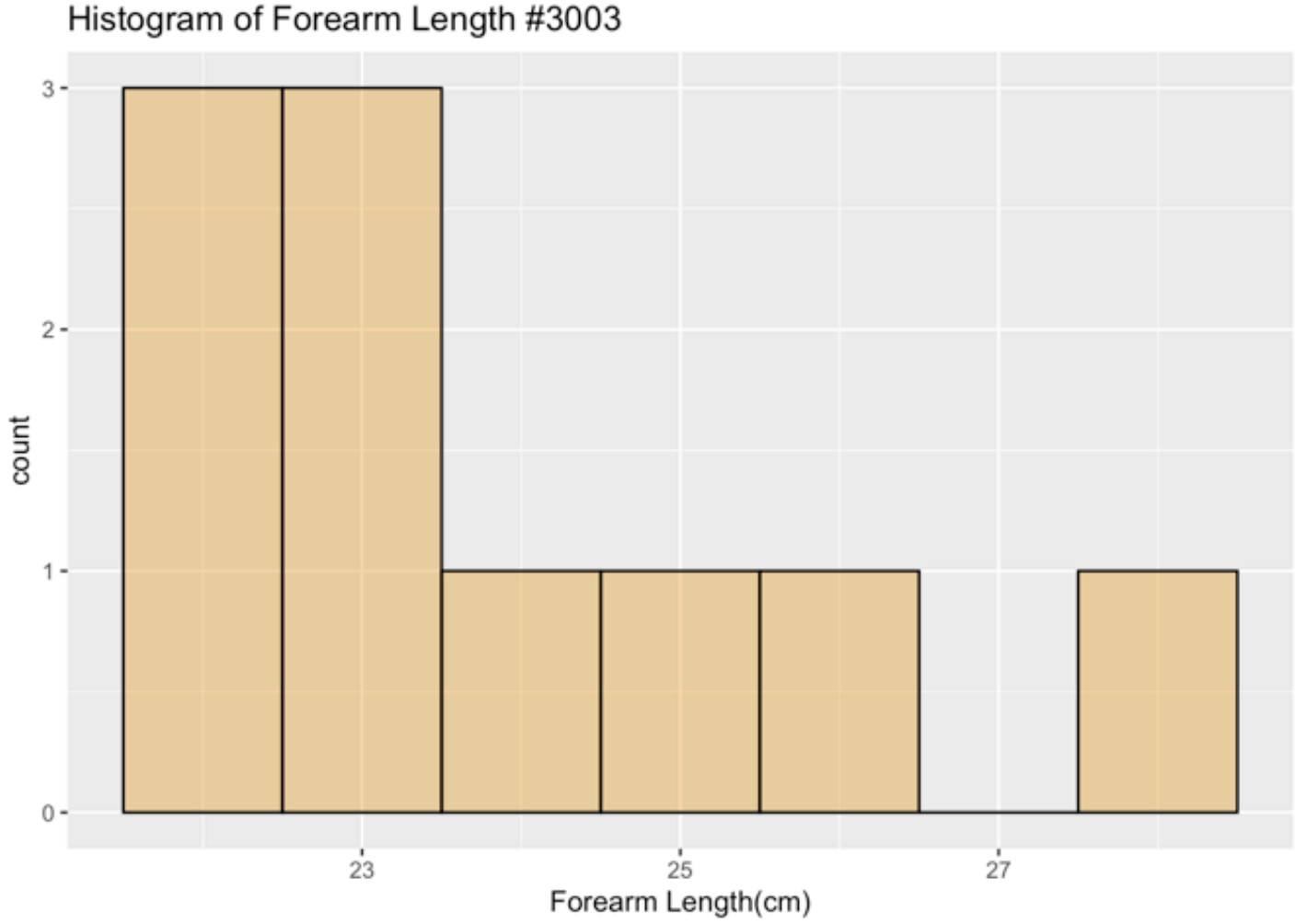
Group 2 Data

```
## 'data.frame':    10 obs. of  2 variables:
##  $ height : num  173 171 182 158 165 ...
##  $ forearm: num  24 28 26 22 25.4 23 22 22 23 23
```

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
count
Height(cm)
0
1
2
3
160
165
170
175
180

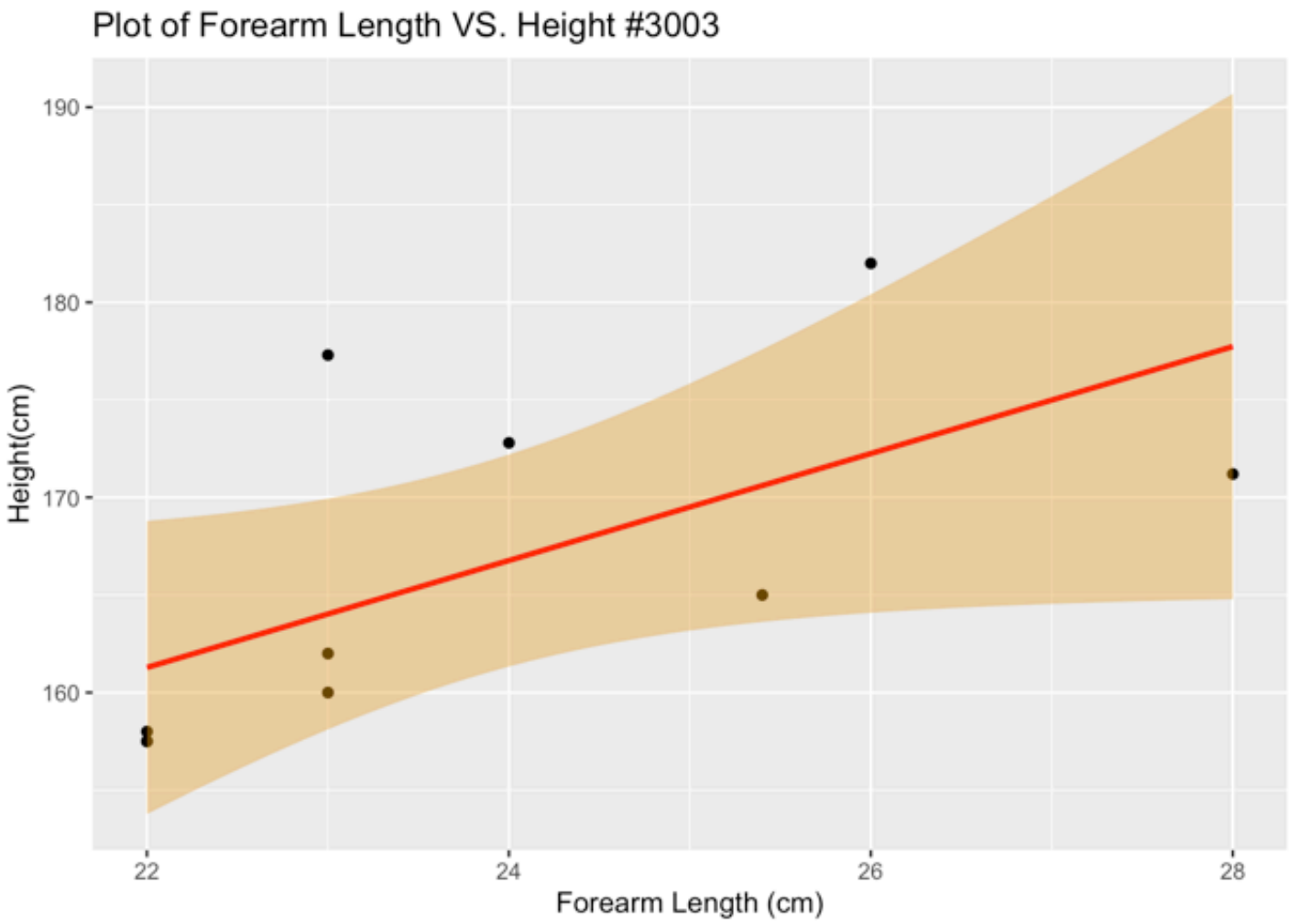

Histogram of Forearm Length #3003
count
Forearm Length(cm)
0
1
2
3
23
25
27

Height Distribution

This distribution seems like a bimodel 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 mediium 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 skewed right, with medium of 23cm and range 6cm.



The relation between forearm length and height are weark, 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.

```
##
## 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 $y = 100.97 + 2.742x$.

Y = Height of the individual

X = Arm length of the individual

The intercept in the scatterplot is 100.97, which is where the line of best fit crosses the height axis. The slope of the line in the scatterplot is 2.74. Thus, we say that each additional centermeter increase is predicted to add 2.742cm to the height.

p-value is 0.05579 which means that we failed to reject the null hypothesis.

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

IV. Discussions and Limitations

Limitation

Gender and age could be facts that influence the result.

Males tend to be taller than female.We can see it from histogram of height, the data is spread into two clusters.it could also effect the mean of height and leads to an

The study should included the age group because as the bone growth stops after 21 year, so the ratio between height & limbs will not change after that age.

Response bias should also be a fact since the data was collected by self-report.

Variables to explore a simple linear regression model.

Shoe size and Height could be variables to explore a simple linear regression model,