

MTH – 522: Advanced Mathematical Statistics

Homework 1

Name: Pranav Vinod

UMassD ID: 01984464

Regression on Pearson's Father-Son data.

Q1. Get the classical Pearson's father-son data by the following R commands.

Ans.

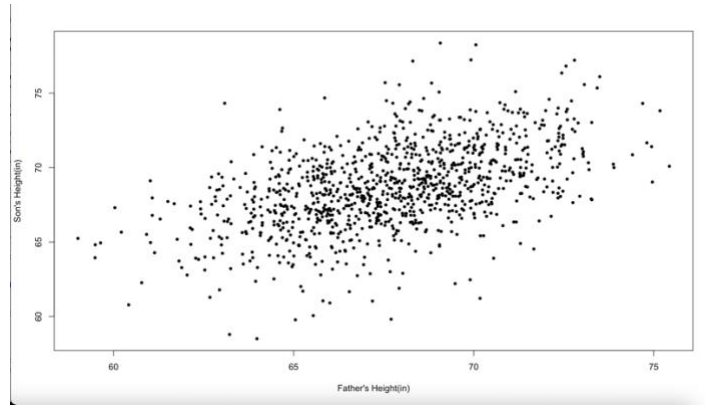
```
# a. Getting the data
library(UsingR)
data(father.son)
View(father.son)
```

fheight	sheight
65.04851	59.77827
63.25094	63.21404
64.95532	63.34242
65.75250	62.79238
61.13723	64.28113
63.02254	64.24221
65.37053	64.08231
64.72398	63.99574
66.06509	64.61338
66.96738	63.97944
59.00800	65.24451
62.93203	65.35102
62.67062	65.67002

Q2. Produce a scatter plot of father's height (x axis) Vs son's height (y axis).

Ans.

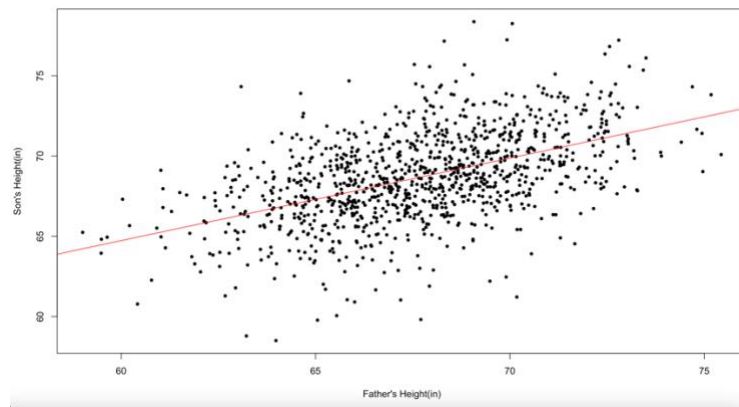
```
# b. Scatter Plot of father's and son's height
plot(father.son$fheight, father.son$sheight, xlab = "Father's Height(in)",
     ylab = "Son's Height(in)", pch=20)
```



Q3. Add the regression line on the scatter plot.

Ans.

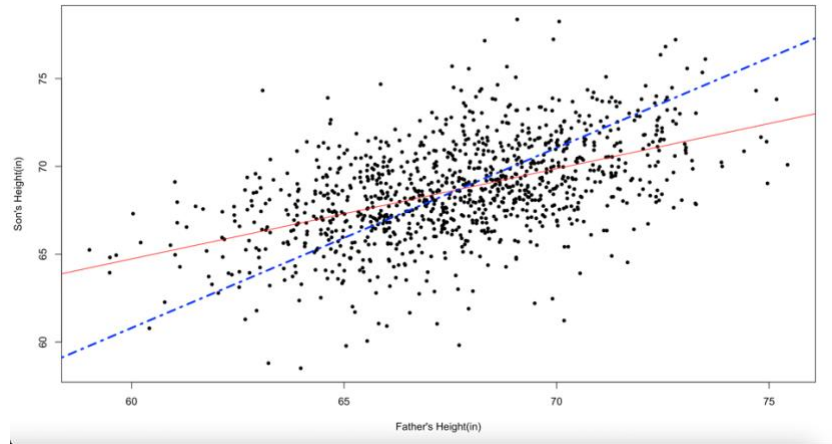
```
# c. Regression line on scatter plot
abline(lm(father.son$sheight~father.son$fheight, data=father.son), col="red")
```



Q4. Add [the SD line](#) (with blue color, different from the regression line) to the same plot.

Ans.

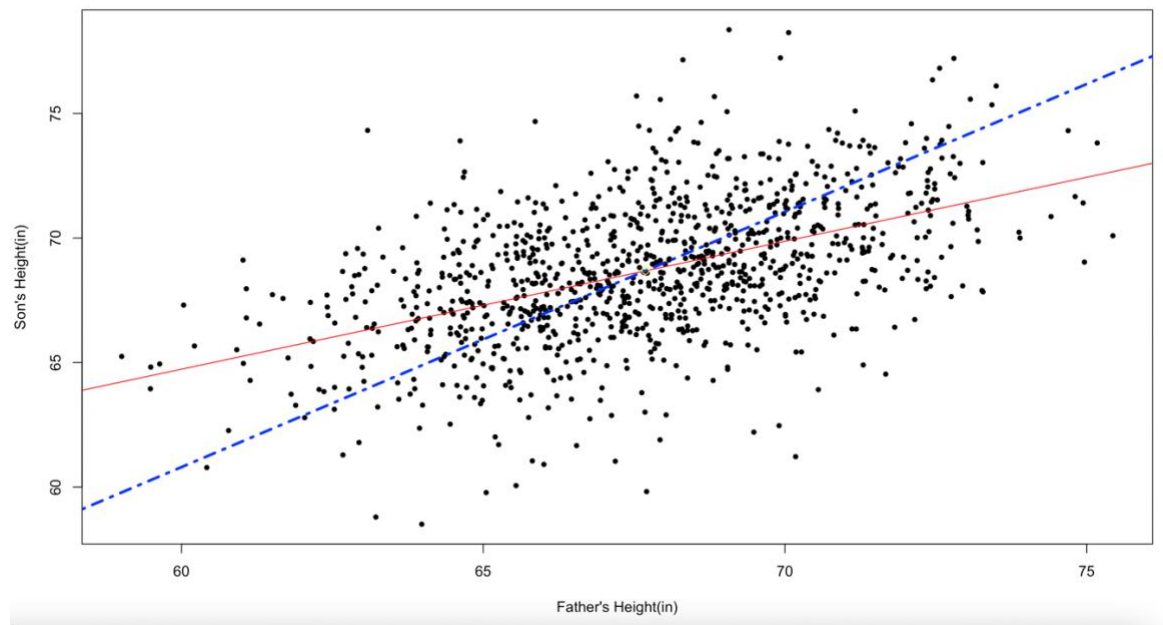
```
# d. Add SD line to plot
slope <- sd(father.son$sheight)/sd(father.son$fheight)
intercept <- (68.68407) - (67.6871)*1.025441
abline(a=intercept, b=slope, col="blue", lty=4, lwd=3)
```



Q5. Mark the center of regression.

Ans.

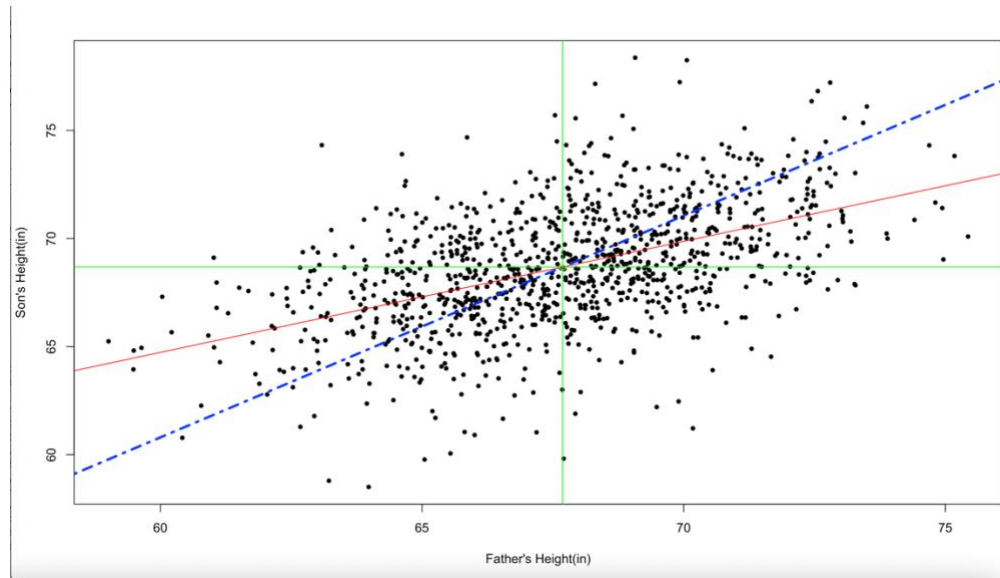
```
# e. center of regression
points(mean(father.son$fheight),mean(father.son$sheight),col="yellow")
```



Q6. Add horizontal and vertical lines (green color) through the center of regression.

Ans.

```
# f. Horizontal and vertical lines through center of regression
abline(h=mean(father.son$sheight), col="green")
abline(v=mean(father.son$fheight), col="green")
```



Q7. Report the linear regression output (including R^2 etc).

Ans.

```
# g. Linear regression output
model <- lm(father.son$sheight~father.son$fheight, data = father.son)
summary(model)
```

Call:

```
lm(formula = father.son$sheight ~ father.son$fheight, data = father.son)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.8772	-1.5144	-0.0079	1.6285	8.9685

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	33.88660	1.83235	18.49	<2e-16 ***
father.son\$fheight	0.51409	0.02705	19.01	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.437 on 1076 degrees of freedom
Multiple R-squared: 0.2513, Adjusted R-squared: 0.2506
F-statistic: 361.2 on 1 and 1076 DF, p-value: < 2.2e-16

The errors are as follows:

1. $RSE = 2.437$
2. $R^2 = 0.2513$
3. $F - Statistic = 361.2$