

Correlation and Regression

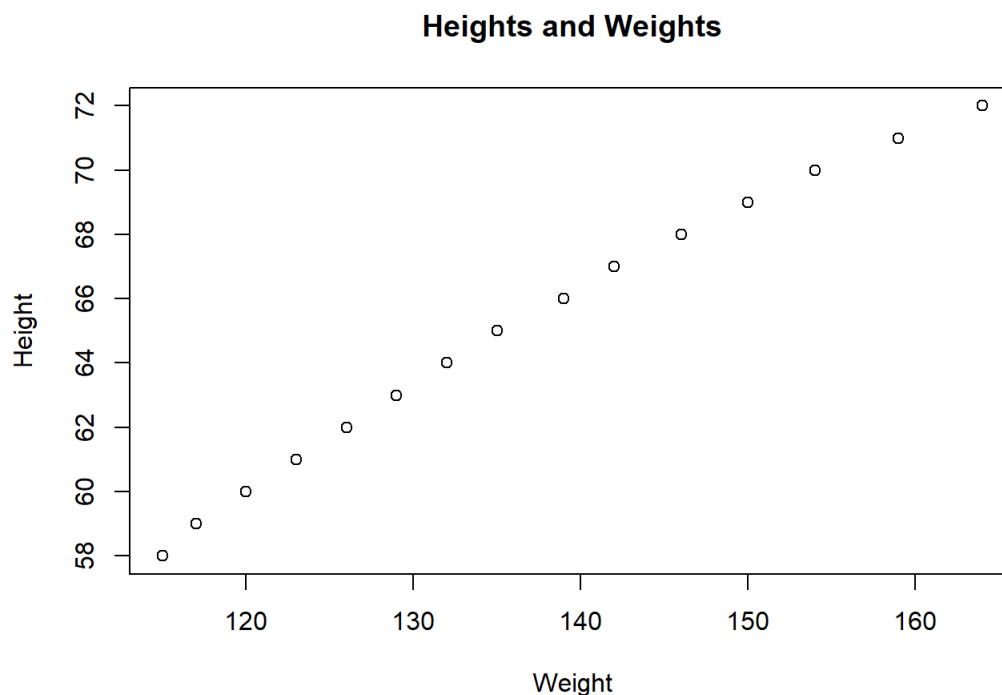
Chirantan Ganguly

27/04/2020

Scatterplots

Let's have a look at the relationship between height and weight through a scatterplot, using the R function `plot()`.

```
plot(women$weight, women$height, main="Heights and Weights", xlab="Weight", ylab="Height")
```



Calculating Correlation

We can calculate the correlation in R using the function `cor()`, which takes your two variables as its first argument.

```
cor(women$height, women$weight)
```

```
## [1] 0.9954948
```

Finding the Line

While Finding the line we use sum of the squared error we can do the same as follows:-

```
y1 <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
y2 <- c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)
y <- c(3, 2, 1, 4, 5, 10, 8, 7, 6, 9)
sum((y1-y)^2)
```

```
## [1] 36
```

```
sum((y2-y)^2)
```

```
## [1] 46
```

Finding The Regression Coefficients in R

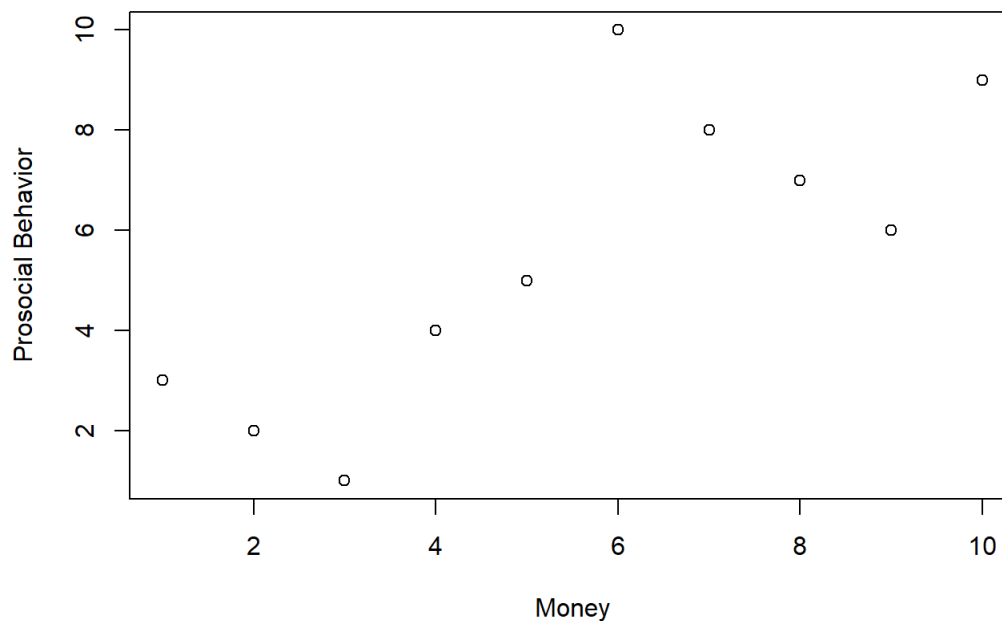
We can find the regression coefficients for our data using the `lm()` function

```
money <- c(1,2,3,4,5,6,7,8,9,10)
prosocal <- c(3, 2, 1, 4, 5, 10, 8, 7, 6,9)
lm(prosocial~money)
```

```
##
## Call:
## lm(formula = prosocial ~ money)
##
## Coefficients:
## (Intercept)      money
##      1.2000      0.7818
```

Using `lm()` To Add A Regression Line To Your Plot

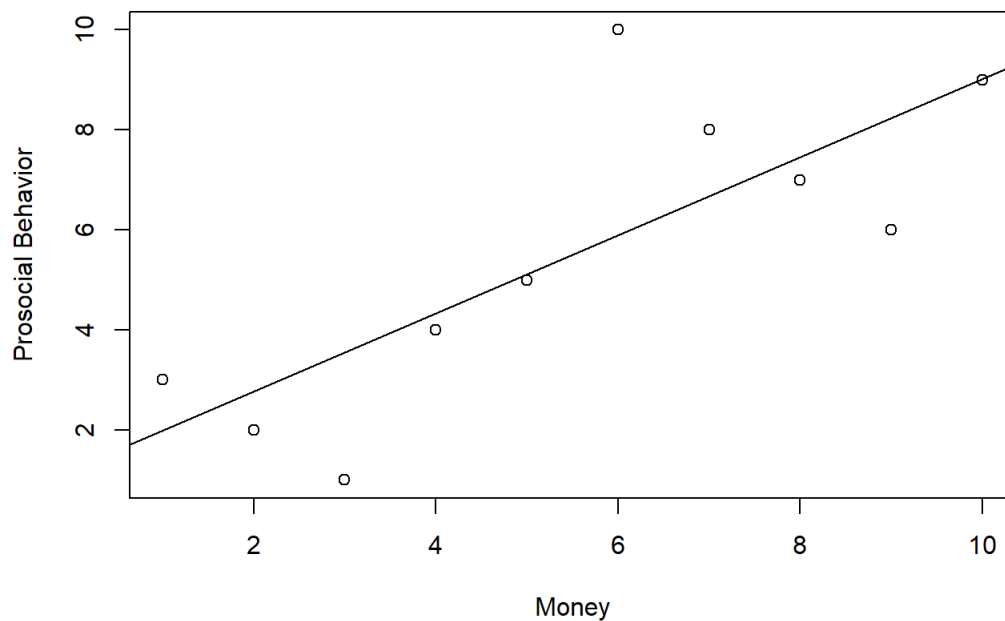
```
plot(money, prosocial, xlab = "Money", ylab = "Prosocial Behavior")
```



After you have created your

scatterplot, you can add a line using the function `abline()`

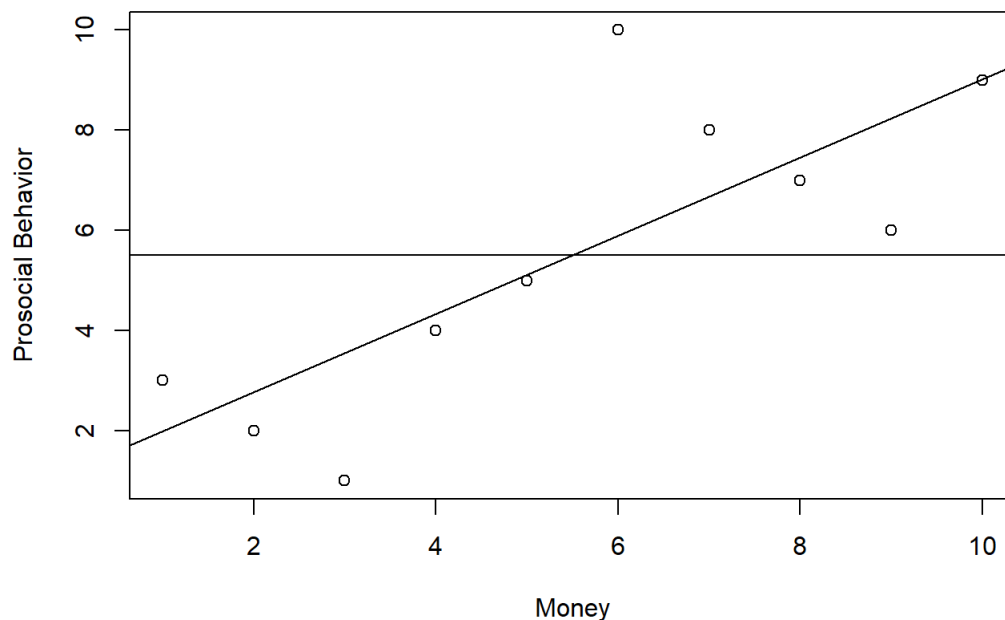
```
plot(money, prosocial, xlab = "Money", ylab = "Prosocial Behavior")
line <- lm(prosocial ~ money)
abline(line)
```



Adding a Line

We can use `abline()` to add any line we like, as long as the first argument is the intercept and the second is the slope.

```
plot(money, prosocial, xlab = "Money", ylab = "Prosocial Behavior")
abline(mean(prosocial), 0)
abline(line)
```



Calculating r^2

The reduction in error by using the regression line compared to the mean line tells us how well the independent variable (money) predicts the dependent variable (prosocial behaviour). This reduction is predicted by r^2 :-

```
r<- (cor(money,prosocial)) ^2
r
```

```
## [1] 0.6112397
```

Putting it all Together: Regression

Lets try to put all what we have learned together:-

```
money <- c(4, 3, 2, 2, 8, 1, 1, 2, 3, 4, 5, 6, 7, 9, 9, 8, 12)
education <- c(3, 4, 6, 9, 3, 3, 1, 2, 1, 4, 5, 7, 10, 8, 7, 6, 9)
r<-cor(money,education)
r
```

```
## [1] 0.5846627
```

```
line<-lm(money~education)
line
```

```
##
## Call:
## lm(formula = money ~ education)
##
## Coefficients:
## (Intercept)      education
##      1.5744         0.6731
```

```
plot(education,money,main="My Scatterplot")
abline(line)
```

My Scatterplot

