Sparks_Foundation_Task1_ScoresHours_Devanshi_Varshney

Devanshi Varshney

10/05/2021

```
library(dplyr)
library(ggplot2)
```

```
df<-read.csv("C:/Users/devan/Downloads/student_scores.csv")

head(df)
```

```
##   Hours Scores
## 1   2.5     21
## 2   5.1     47
## 3   3.2     27
## 4   8.5     75
## 5   3.5     30
## 6   1.5     20
```

Analysing the dataset

the range of the columns, mean, median, missing values will give an insight of the dataset I will work upon.

```
#insight
summary(df)
```

```
##      Hours           Scores
##  Min.   :1.100   Min.   :17.00
##  1st Qu.:2.700   1st Qu.:30.00
##  Median :4.800   Median :47.00
##  Mean   :5.012   Mean   :51.48
##  3rd Qu.:7.400   3rd Qu.:75.00
##  Max.   :9.200   Max.   :95.00
```

```
#structure
str(df)
```

```
## 'data.frame':    25 obs. of  2 variables:
##  $ Hours : num  2.5 5.1 3.2 8.5 3.5 1.5 9.2 5.5 8.3 2.7 ...
##  $ Scores: int  21 47 27 75 30 20 88 60 81 25 ...
```

```r
#check for missing values
colSums(is.na(df))
```

```
##  Hours Scores
##      0      0
```

Creating a linear regression model

the model will show the relation between the marks scored by a student and the number of hours of study he/she had put in daily.

```r
model<-lm(Scores~Hours, data=df)
model
```

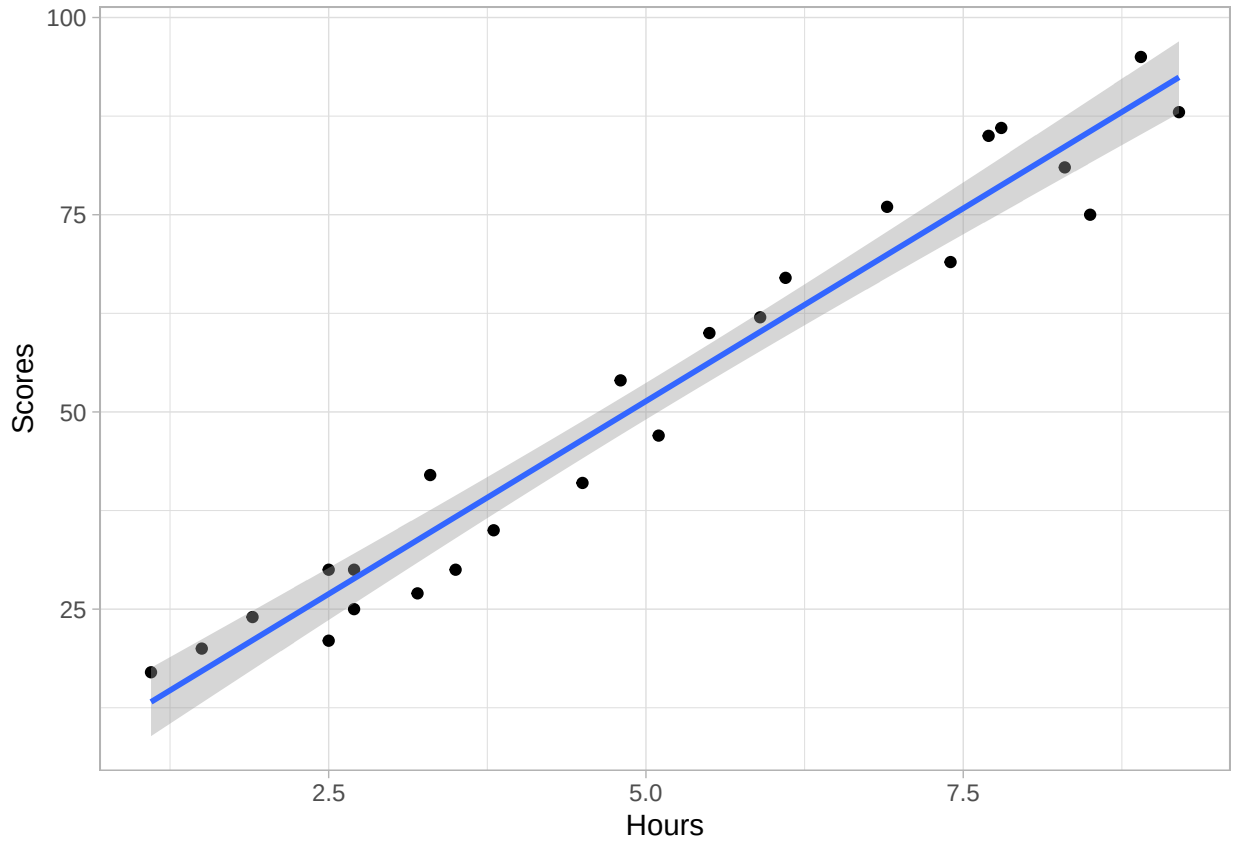
```
##
## Call:
## lm(formula = Scores ~ Hours, data = df)
##
## Coefficients:
## (Intercept)        Hours
##       2.484        9.776
```

Visualizing the Line od Regression

to see how our data is distributed around the regression line.

```r
ggplot(df, aes(Hours, Scores))+geom_point()+stat_smooth(method = lm)+theme_light()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Answer of the question in the task

we will use the above created model to form the linear equation by using the intercept and slope of the model. Then replace hours with 9.25 as per the question to get the answer.

```
summary(model)
```

```
##
## Call:
## lm(formula = Scores ~ Hours, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.578  -5.340   1.839   4.593   7.265
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4837     2.5317   0.981    0.337
## Hours         9.7758     0.4529  21.583   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.603 on 23 degrees of freedom
## Multiple R-squared:  0.9529, Adjusted R-squared:  0.9509
## F-statistic: 465.8 on 1 and 23 DF,  p-value: < 2.2e-16
```

```r
hrs<-9.25
scr<-2.484+(9.776*hrs)
scr
```

```
## [1] 92.912
```

Result: If a student studies for 9.776 hours, he will score approximately 92.912 ~ 93 marks