LINEAR REGRETION OF 'Orange' Datasheet using R-language

Sasanka Medhi

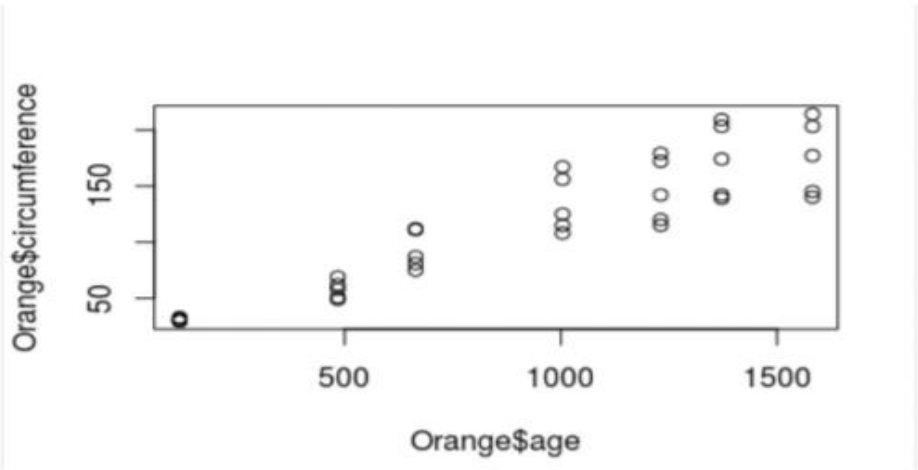
Linear Regression:

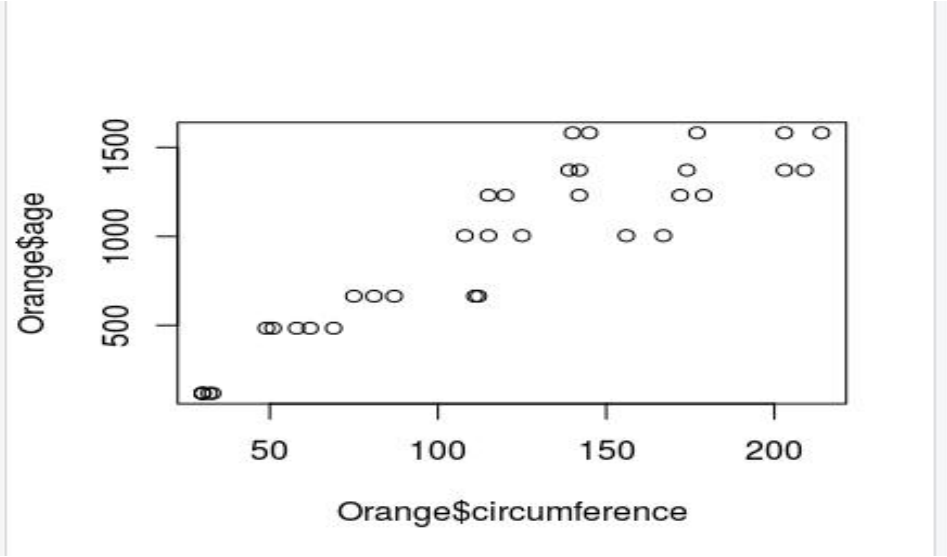
Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling. For example, it can be used to quantify the relative impacts of age, gender, and diet (the predictor variables) on height (the outcome variable). Linear regression is also known as multiple regression, multivariate regression, ordinary least squares (OLS), and regression. This post will show you examples of linear regression, including an example of simple linear regression and an example of multiple linear regression.

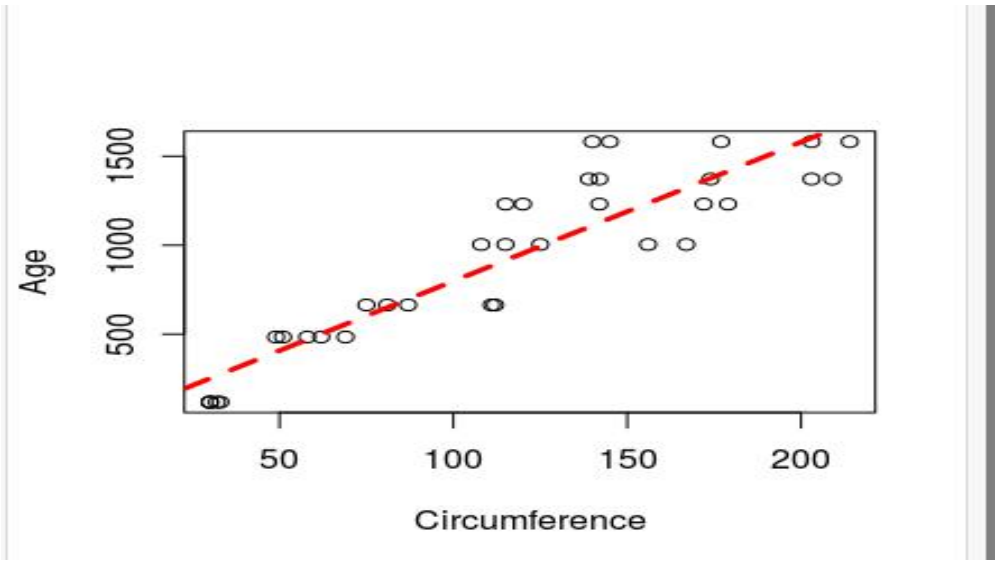
R-code for prediction:

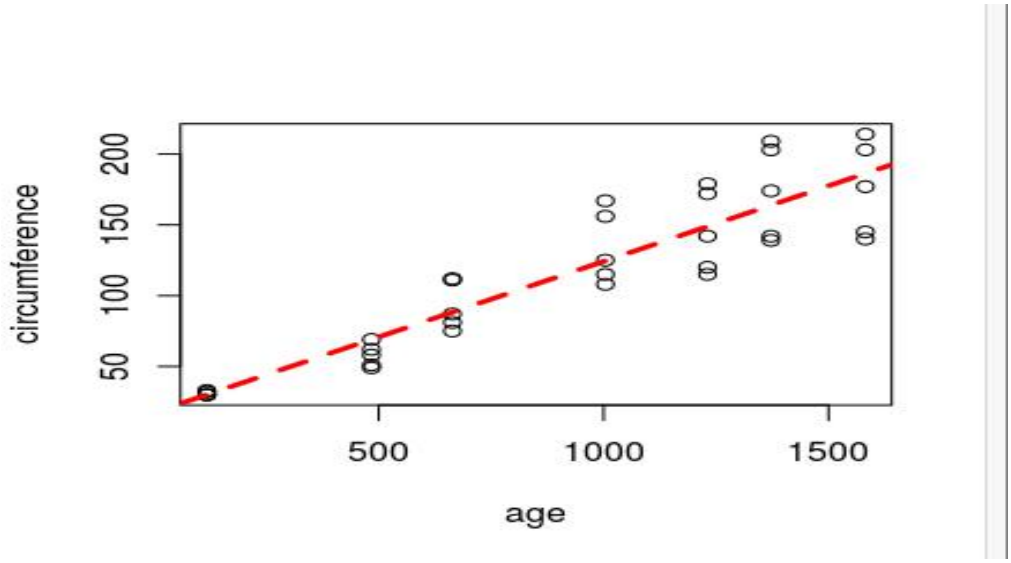
```
> #loading the "Orange" datasheet, finding the class, and printing first 6 data instances> data
("Orange")> head(Orange)  Tree  age circumference
1    1  118           30
2    1  484           58
3    1  664           87
4    1 1004          115
5    1 1231          120
6    1 1372          142> #Finding the correlation between attribute and the target variable> c
or(Orange$circumference, Orange$age)[1] 0.9135189> > #program to build a linear regretion mode
l> model<-lm(age~circumference,data=Orange)> #program to find the summary of the mode> summary
(model)
Call:
lm(formula = age ~ circumference, data = Orange)

Residuals:
    Min      1Q  Median      3Q     Max
-317.88 -140.90  -17.20   96.54  471.16

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    16.6036    78.1406   0.212    0.833
circumference   7.8160     0.6059  12.900 1.93e-14 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 203.1 on 33 degrees of freedom
Multiple R-squared:  0.8345,    Adjusted R-squared:  0.8295
F-statistic: 166.4 on 1 and 33 DF,  p-value: 1.931e-14
> #plotting> plot(Orange$age,Orange$circumference)> plot(Orange$circumference,Orange$age)> #pr
edict the circumference of an orange given its age> predict(model, data.frame("circumference"=1
00))       1
798.2035 > predict(model, data.frame("circumference"=122))       1
```

```
970.1554 > predict(model, data.frame("circumference"=142))       1
1126.475 > predict(model, data.frame("circumference"=400))       1
3143.003 > predict(model, data.frame("circumference"=1450))      1
11349.8 > predict(model, data.frame("circumference"=2000))      1
15648.6 > predict(model, data.frame("circumference"=2870))       1
22448.52 > predict(model, data.frame("circumference"=2134))       1
16695.94 > predict(model, data.frame("circumference"=219304))      1
1714096 > predict(model, data.frame("circumference"=99999))       1
781608.6 > #program to build another model; given circumference and predicting the age> model<-
lm(circumference~age, data=Orange)> predict(model,data.frame("age"=1))       1
17.50642 > predict(model,data.frame("age"=10))       1
18.46735 > predict(model,data.frame("age"=15))       1
19.00121 > predict(model,data.frame("age"=20))       1
19.53506 > predict(model,data.frame("age"=100))       1
28.07668 > predict(model,data.frame("age"=150))      1
33.4152 > predict(model,data.frame("age"=200))       1
38.75372 > predict(model,data.frame("age"=500))       1
70.78481 > predict(model,data.frame("age"=1000))     1
124.17 > predict(model,data.frame("age"=9999))       1
1084.996 > model<-lm(age~circumference,data=Orange)> plot(Orange$circumference, Orange$age, xl
ab='Circumference', ylab='Age')> abline(model,col="red",lty=2,lwd=3)> model<-lm(circumference~
age,data=Orange)> plot(Orange$age, Orange$circumference, xlab='age', ylab='circumference')> ab
line(model,col="red",lty=2,lwd=3)
```









Conclusion:

We have loaded the orange datasheet, found the correlation between the attribute and the target variable, build the model and predicted the age given its circumference and predicted the circumference given its age and plotted the following graphs. We also saw that R-langiage is much easier than python