

# LINEAR REGRETION OF 'Orange' Datasheet using R-LANGUAGE

*Kunal Chowdhury*

## 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

> cor(Orange\$circumference, Orange\$age)

[1] 0.9135189

> #program to build a linear regrestion model

> 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)
```

```
> #predict the circumference of an orange given its age
```

```
> predict(model, data.frame("circumference"=100))
```

1

798.2035

```
> predict(model, data.frame("circumference"=250))
```

1

1970.603

```
> predict(model, data.frame("circumference"=70))
```

1

563.7235

```
> predict(model, data.frame("circumference"=80))
```

1

641.8835

```
> predict(model, data.frame("circumference"=999))
```

1

7824.786

```
> predict(model, data.frame("circumference"=303))
```

1

2384.851

```
> predict(model, data.frame("circumference"=245))
```

1

1931.523

```
> predict(model, data.frame("circumference"=334))
1
2627.147

> predict(model, data.frame("circumference"=101))
1
806.0195

> predict(model, data.frame("circumference"=10111))
1
79044.16

> #program to build another model; given circumference and predicting the age
> model<-lm(circumference~age, data=Orange)
> #predicting the age given the circumference
> predict(model,data.frame("age"=1))
1
17.50642

> predict(model,data.frame("age"=20))
1
19.53506

> predict(model,data.frame("age"=30))
1
20.60276

> predict(model,data.frame("age"=69))
1
24.7668

> predict(model,data.frame("age"=72))
1
25.08711

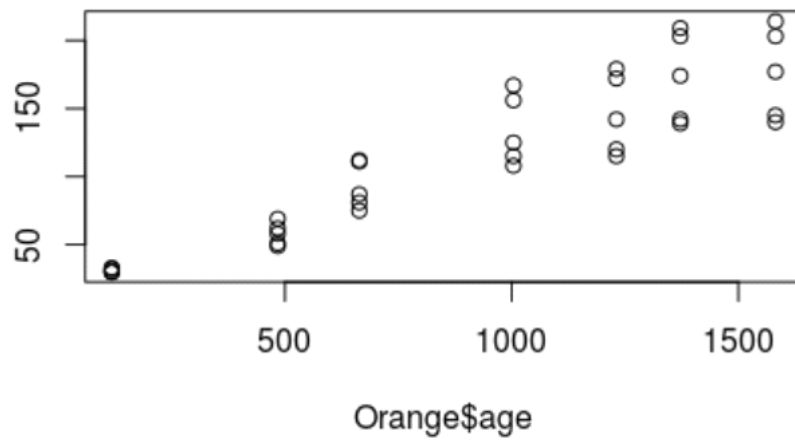
> predict(model,data.frame("age"=29))
```

```

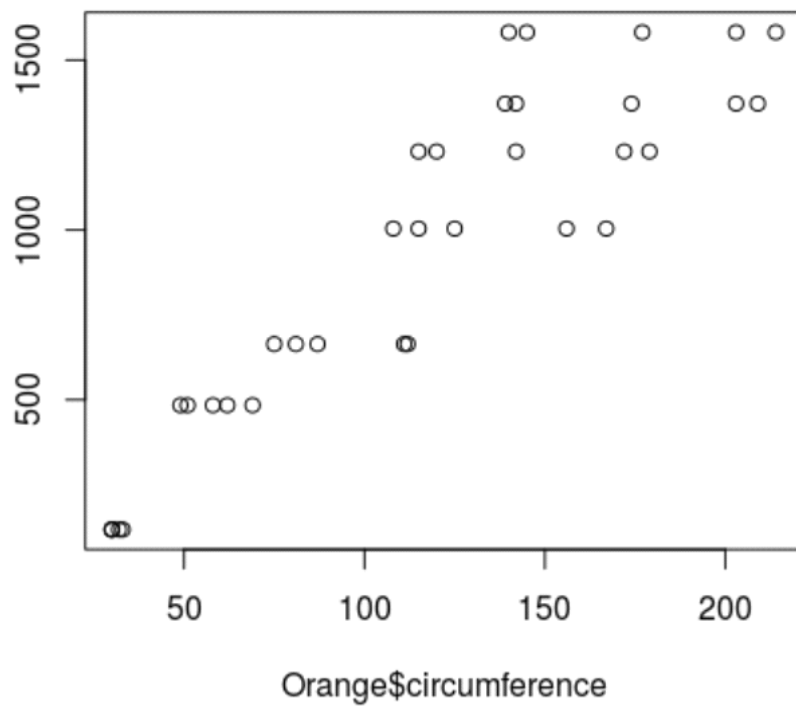
1
20.49599
> predict(model,data.frame("age"=13))
1
18.78766
> predict(model,data.frame("age"=2))
1
17.61319
> predict(model,data.frame("age"=5))
1
17.9335
> predict(model,data.frame("age"=32))
1
20.8163
> predict(model,data.frame("age"=7))
1
18.14704
>model<-lm(age~circumference,data=Orange)
> plot(Orange$circumference, Orange$age, xlab='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')
> abline(model,col="red",lty=2,lwd=3)

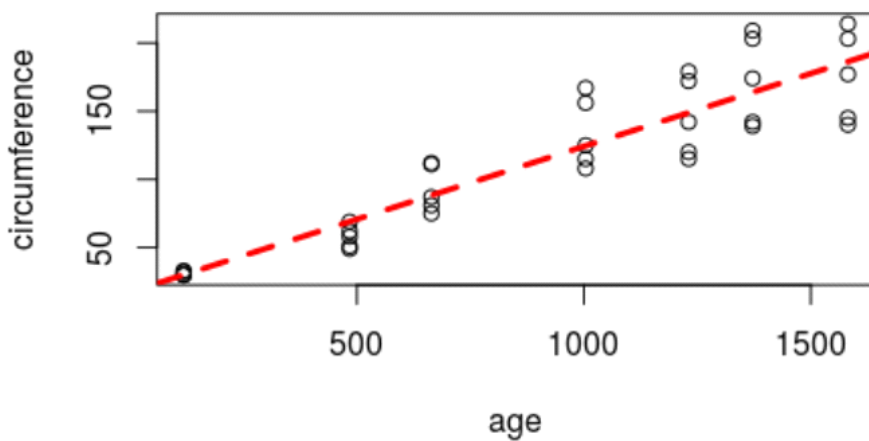
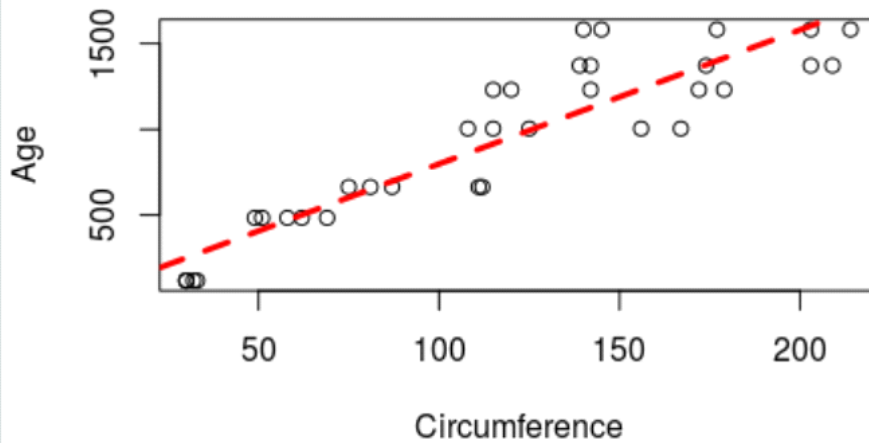
```

Orange\$circumference



Orange\$age





## Conclusion:

We have loaded the orange datasheet, found the correlation between the attribute and the target variable, built 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-language is much easier than python.