TP\_EX3

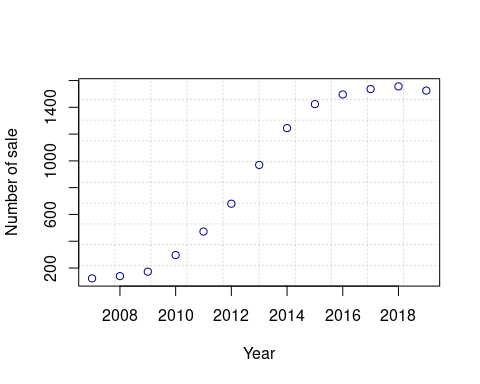
vhinyg

28/09/2020

tab = read.table("cellphonedata.txt", header = T,sep = ";",dec = ",")  
  
X = cbind(rep(1,length(tab[,1])-2),tab[1:(length(tab[,1])-2),1])

Affichages des données

Ytest = tab[1:(length(tab[,1])-2),c(2)]  
Yfut = tab[(length(tab[,1])-1):(length(tab[,1])),c(2)]  
Xtest = tab[1:(length(tab[,1])-2),c(1)]  
Xfut = Yfut = tab[(length(tab[,1])-1):(length(tab[,1])),c(1)]  
X1 = cbind(rep(1,length(tab[,1])-2),Xtest)  
plot(Xtest,Ytest,col = 'blue4',xlab = "Year",ylab = "Number of sale",panel.first = grid(10))



The trend of the curve let us think that a linear model doesn’t fit with data. The trend let us think about a or an function between the model and the variable However we are going to give a chance to the linear model then see his limitations on predictions

cor(tab)

## year EndUsers  
## year 1.0000000 0.9509108  
## EndUsers 0.9509108 1.0000000

However, because we are going to give a chance to the linear model then see his limitations on predictions. First it’s to notice that this cross correlations term is close to 1 because most of data are in the middle of the distribution

Estimation de and

tab1 = tab[1:(length(tab[,1])-2),]  
(tab1)

## year EndUsers  
## 1 2007 122.32  
## 2 2008 139.29  
## 3 2009 172.38  
## 4 2010 296.65  
## 5 2011 472.00  
## 6 2012 680.11  
## 7 2013 969.72  
## 8 2014 1244.74  
## 9 2015 1423.90  
## 10 2016 1495.96  
## 11 2017 1536.27  
## 12 2018 1556.27  
## 13 2019 1524.84

modreg = lm(EndUsers ~ . , data = tab1)  
summary(modreg)

##   
## Call:  
## lm(formula = EndUsers ~ ., data = tab1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -260.95 -126.01 -13.31 118.19 232.00   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -297979.58 23611.08 -12.62 6.92e-08 \*\*\*  
## year 148.47 11.73 12.66 6.71e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 158.2 on 11 degrees of freedom  
## Multiple R-squared: 0.9358, Adjusted R-squared: 0.9299   
## F-statistic: 160.2 on 1 and 11 DF, p-value: 6.706e-08

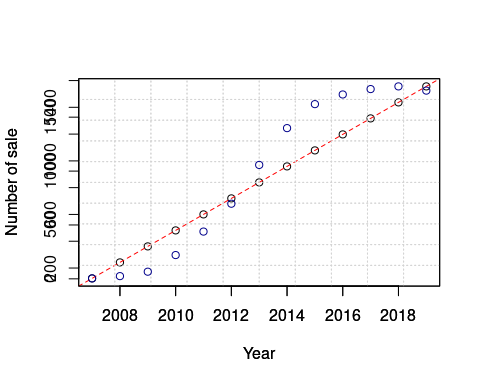
betah = modreg$coefficients  
betah #coefficient beta hat

## (Intercept) year   
## -297979.5765 148.4722

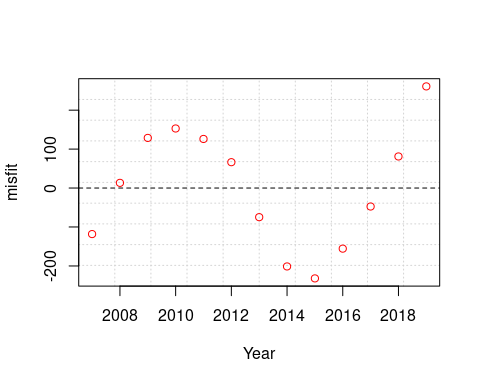
Yhat = X1%\*%betah

plot of the data estimated trend

{plot(Xtest,Yhat,xlab = "Year",ylab = "Number of sale",panel.first = grid(10))   
abline(betah[1],betah[2],col = 2,lty =2)  
par(new = T)  
plot(Xtest,Ytest,col = 'blue4',xlab = "Year",ylab = "Number of sale",panel.first = grid(10))}



{plot(Xtest,Yhat - Ytest,xlab = "Year",ylab = "misfit", col = "red",panel.first = grid(10))  
abline(0,0,col = 1,lty =2)}

 we can see how bad is the linear model, it doesn’t fit to data. And the trend show that a prediction will give data far from the reality.

Exemple let’s predict Year 2020 and 2021

Xpred = matrix(c(1,1,2020,2021),2,2)  
Ypred = Xpred%\*%betah  
misfit = Ypred - Yfut

We are going to look a for a new model.

ATTENTION !!!!!!

PAS FINI!!!!!!! We are oing to try the sigmoid function on these data if with We are going to estimate the best The inverse of is $f^{-1}(x) = -(1/) \* ln(-1) = ln(-1) $

We are going to estimate through a linear regression between and avec

Data

Ytestb = -log(1-(1/Ytest))  
dim(tab)

## [1] 15 2

Y1 = cbind(rep(1,length(Ytestb)),Ytestb)  
Y1

## Ytestb  
## [1,] 1 0.0082088788  
## [2,] 1 0.0072051612  
## [3,] 1 0.0058180290  
## [4,] 1 0.0033766704  
## [5,] 1 0.0021208916  
## [6,] 1 0.0014714324  
## [7,] 1 0.0010317576  
## [8,] 1 0.0008037035  
## [9,] 1 0.0007025432  
## [10,] 1 0.0006686906  
## [11,] 1 0.0006511392  
## [12,] 1 0.0006427686  
## [13,] 1 0.0006560216

tab2 = as.data.frame(cbind(Xtest,Ytestb))  
tab2

## Xtest Ytestb  
## 1 2007 0.0082088788  
## 2 2008 0.0072051612  
## 3 2009 0.0058180290  
## 4 2010 0.0033766704  
## 5 2011 0.0021208916  
## 6 2012 0.0014714324  
## 7 2013 0.0010317576  
## 8 2014 0.0008037035  
## 9 2015 0.0007025432  
## 10 2016 0.0006686906  
## 11 2017 0.0006511392  
## 12 2018 0.0006427686  
## 13 2019 0.0006560216

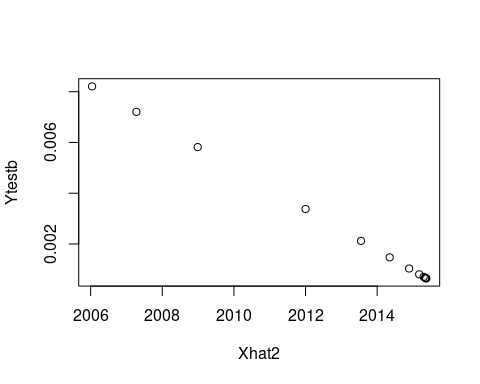
regmod = lm(Xtest ~ ., data = tab2)  
summary(regmod)

##   
## Call:  
## lm(formula = Xtest ~ ., data = tab2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5489 -1.8921 0.0107 0.9592 3.6445   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2016.1645 0.7916 2546.862 < 2e-16 \*\*\*  
## Ytestb -1233.2708 215.6790 -5.718 0.000135 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.041 on 11 degrees of freedom  
## Multiple R-squared: 0.7483, Adjusted R-squared: 0.7254   
## F-statistic: 32.7 on 1 and 11 DF, p-value: 0.0001345

betah2 = regmod$coefficients  
lambda= betah2[2]  
  
Xhat2 = Y1 %\*% betah2  
Xhat2

## [,1]  
## [1,] 2006.041  
## [2,] 2007.279  
## [3,] 2008.989  
## [4,] 2012.000  
## [5,] 2013.549  
## [6,] 2014.350  
## [7,] 2014.892  
## [8,] 2015.173  
## [9,] 2015.298  
## [10,] 2015.340  
## [11,] 2015.362  
## [12,] 2015.372  
## [13,] 2015.355

plot(Xhat2,Ytestb)



We have the for the simoid function now we can build the model

Yhat2 = 1/(1+exp(lambda\*Xhat2))  
Xhat2

## [,1]  
## [1,] 2006.041  
## [2,] 2007.279  
## [3,] 2008.989  
## [4,] 2012.000  
## [5,] 2013.549  
## [6,] 2014.350  
## [7,] 2014.892  
## [8,] 2015.173  
## [9,] 2015.298  
## [10,] 2015.340  
## [11,] 2015.362  
## [12,] 2015.372  
## [13,] 2015.355

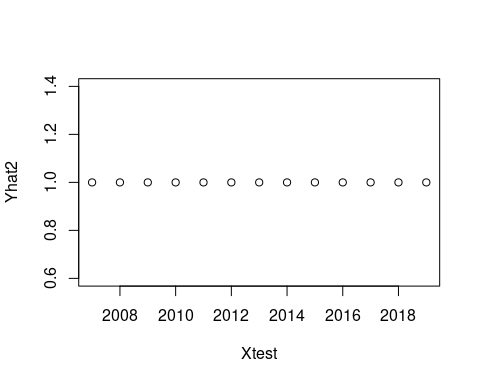
lambda

## Ytestb   
## -1233.271

Yhat2

## [,1]  
## [1,] 1  
## [2,] 1  
## [3,] 1  
## [4,] 1  
## [5,] 1  
## [6,] 1  
## [7,] 1  
## [8,] 1  
## [9,] 1  
## [10,] 1  
## [11,] 1  
## [12,] 1  
## [13,] 1

plot(Xtest,Yhat2)



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.