Jeux de données CreditBancaire

Ymca DOLBRUN

2022-12-09

# Importation des données

library(broom)

## Warning: package 'broom' was built under R version 4.2.2

library(car)

## Warning: package 'car' was built under R version 4.2.2

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.2.2

library(ggplot2)  
  
 df<-read.csv("CreditBancaire.csv")

## Regression du nombre de jours de retards sur les variables Credit et Type

Reg <-lm (Jours~Credit+Type,data = df)  
 Reg

##   
## Call:  
## lm(formula = Jours ~ Credit + Type, data = df)  
##   
## Coefficients:  
## (Intercept) Credit TypeConsommation TypeProduction   
## 8.519e+01 -5.484e-06 -1.486e+01 1.498e+01

### Affichons les résultats détaillés de la regression

summary(Reg)

##   
## Call:  
## lm(formula = Jours ~ Credit + Type, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -96.26 -69.22 -49.06 17.29 270.22   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.519e+01 1.275e+01 6.681 1.63e-10 \*\*\*  
## Credit -5.484e-06 1.182e-05 -0.464 0.643   
## TypeConsommation -1.486e+01 1.472e+01 -1.009 0.314   
## TypeProduction 1.498e+01 2.532e+01 0.591 0.555   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 104 on 241 degrees of freedom  
## Multiple R-squared: 0.007512, Adjusted R-squared: -0.004842   
## F-statistic: 0.608 on 3 and 241 DF, p-value: 0.6104

#### Extraction du tableau des coefficients qui inclut aussi les statistiques de Student et les probabilités.

Coeff<- tidy(Reg)  
 Coeff

## # A tibble: 4 x 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 85.2 12.8 6.68 1.63e-10  
## 2 Credit -0.00000548 0.0000118 -0.464 6.43e- 1  
## 3 TypeConsommation -14.9 14.7 -1.01 3.14e- 1  
## 4 TypeProduction 15.0 25.3 0.591 5.55e- 1

##### Affichage des colonnes des coefficients estimés et les valeurs de la statistique de Student

COEFF<-data.frame(tidy(Reg)$estimate,tidy(Reg)$statistic)  
 COEFF

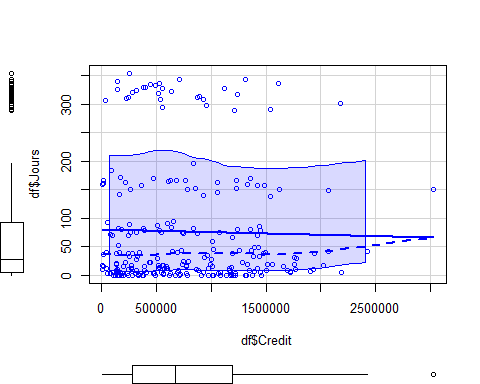
## tidy.Reg..estimate tidy.Reg..statistic  
## 1 8.519082e+01 6.6809256  
## 2 -5.483559e-06 -0.4637417  
## 3 -1.485726e+01 -1.0090929  
## 4 1.497503e+01 0.5913659

#### La droite de regression:

DroiteReg= 8.519082e+01-5.483559e-06\* Credit -1.485726e+01\* TypeConsommation +1.497503e+01\*TypeProduction

###### Réalisation du graphique de nuage de points

scatterplot(df$Jours~df$Credit, data = df)



###### Création d’une fonction qui nous permet de régresser une variable y sur une autre notée x, sachant que ces deux variables sont deux colonnes d’un dataframe noté data.

FonctionRegression <- function(){  
 y= c(3,9,20,40,45,90)  
 x= c(30,40,50,60,70,80)  
 data= data.frame(y,x)  
 Coefx<- summary(lm(y~x, data))$coefficients[,1]  
 Coefx  
 StatStudent<-summary(lm(y~x, data))$coefficients[,3]  
 StatStudent  
   
 data1= data.frame(Coefx,StatStudent)  
   
 return(data1)  
   
 }  
  
 FonctionRegression()

## Coefx StatStudent  
## (Intercept) -53.971429 -3.357499  
## x 1.608571 5.762928