Examen OLS - STT5100

Arthur Charpentier

Hiver 2020

```
load("stt5100-intra.RData")
database$logPrix = log(database$Prix)
database$Prix <- NULL
str(database)
```

```
## 'data.frame':    1425 obs. of  41 variables:
##  $ Zone              : Factor w/ 7 levels "A (agr)","C (all)",..: 6 6 6 6 6 6 3 7 6 2 ...
##  $ Surface_Lot       : int  8396 11631 10456 14694 10400 9760 2998 6000 7400 8712 ...
##  $ Rue               : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Forme             : Factor w/ 4 levels "IR1","IR2","IR3",..: 1 1 1 1 4 4 4 4 4 4 ...
##  $ Utilities         : Factor w/ 3 levels "AllPub","NoSeWa",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Configuration     : Factor w/ 5 levels "Corner","CulDSac",..: 5 1 5 5 5 5 5 5 1 5 ...
##  $ Proxim_1          : Factor w/ 9 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 1 3 ...
##  $ Proxim_2          : Factor w/ 8 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ Logement          : Factor w/ 5 levels "1Fam","2fmCon",..: 1 1 1 1 1 1 5 1 1 1 ...
##  $ Style             : Factor w/ 8 levels "1.5Fin","1.5Unf",..: 6 6 3 3 6 6 6 1 3 1 ...
##  $ Int_Qualite       : int  7 8 6 8 6 6 6 6 7 4 ...
##  $ Int_Condition     : int  5 5 6 9 5 8 5 7 6 7 ...
##  $ Construction_Annee: int  2003 2004 1967 1977 1972 1964 2000 1940 1962 1900 ...
##  $ Renovation_Annee  : int  2003 2005 1967 2008 1972 1993 2000 1989 1962 1950 ...
##  $ Toit              : Factor w/ 6 levels "Flat","Gable",..: 2 2 4 2 2 4 2 2 2 2 ...
##  $ Exterieur         : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 14 7 9 14 15 9 15 15 9 ...
##  $ Maconnerie        : Factor w/ 6 levels "","BrkCmn","BrkFace",..: 3 3 3 3 5 3 3 5 5 5 ...
##  $ Ext_Qualite       : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 3 4 1 4 4 3 4 4 4 ...
##  $ Ext_Condition     : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 1 5 5 5 5 5 5 ...
##  $ Foundation        : Factor w/ 6 levels "BrkTil","CBlock",..: 3 3 2 2 2 2 3 2 2 5 ...
##  $ Chauffage         : Factor w/ 6 levels "Floor","GasA",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Chauff_Qualite    : Factor w/ 5 levels "Ex","Fa","Gd",..: 1 1 3 1 3 3 1 1 3 3 ...
##  $ Electricite       : Factor w/ 6 levels "","FuseA","FuseF",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ Surface_RdC       : int  847 1052 1829 1694 1012 1020 768 1014 1045 859 ...
##  $ Surface_Etage     : int  1139 1321 0 0 778 831 756 658 0 319 ...
##  $ Surface_Autre     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Toilettes         : int  2 2 2 2 1 2 2 1 1 1 ...
##  $ Chambres          : int  4 4 4 2 4 3 2 3 3 2 ...
##  $ Cuisine_Qualite   : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 3 4 1 4 3 3 3 4 4 ...
##  $ Pieces            : int  9 9 8 5 8 7 4 6 6 7 ...
##  $ Note              : Factor w/ 8 levels "Maj1","Maj2",..: 8 8 8 8 4 8 8 8 8 8 ...
##  $ Foyer             : int  1 1 0 1 1 1 0 1 0 0 ...
##  $ Allee             : Factor w/ 3 levels "N","P","Y": 3 3 3 3 3 3 3 3 3 1 ...
##  $ Garage_Ext_Surface: int  48 46 76 120 0 42 32 11 0 0 ...
##  $ Garage_Int_Surface: int  0 0 0 0 0 0 0 0 0 98 ...
##  $ Piscine_Surface   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Vente_Mois        : int  5 6 5 6 11 7 6 7 3 1 ...
##  $ Vente_Annee       : int  2008 2006 2010 2009 2006 2007 2006 2008 2006 2010 ...
##  $ Vente_Type        : Factor w/ 10 levels "COD","Con","ConLD",..: 10 7 10 10 10 10 10 10 10 10 ...
##  $ Vente_Condition   : Factor w/ 6 levels "Abnorml","AdjLand",..: 5 6 5 5 5 5 5 5 5 1 ...
##  $ logPrix           : num  12.3 12.5 12.3 12.7 12 ...
```

Variable y `logPrix`

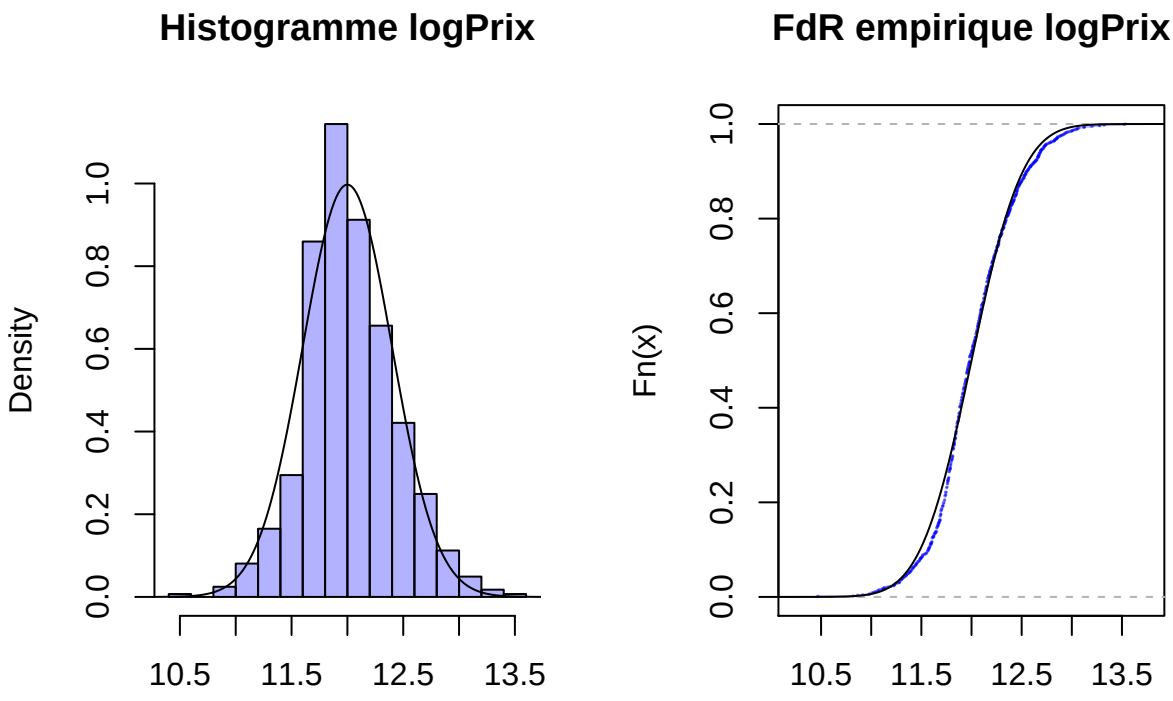
```
mean(database$logPrix)
```

```
## [1] 12.0219
```

```
var(database$logPrix)
```

```
## [1] 0.1662255
```

```
par(mfrow=c(1,2))
with(database,hist(logPrix,probability=TRUE,col=rgb(0,0,1,.3),
                   main="Histogramme logPrix",xlab=""))
u=seq(10,14,length=201)
lines(u,dnorm(u,12,0.4))
with(database,plot(ecdf(logPrix),
                   main="FdR empirique logPrix",xlab="",cex=.1,col=rgb(0,0,1,.3)))
lines(u,pnorm(u,12,0.4))
```



Variable x `Int_Qualite`

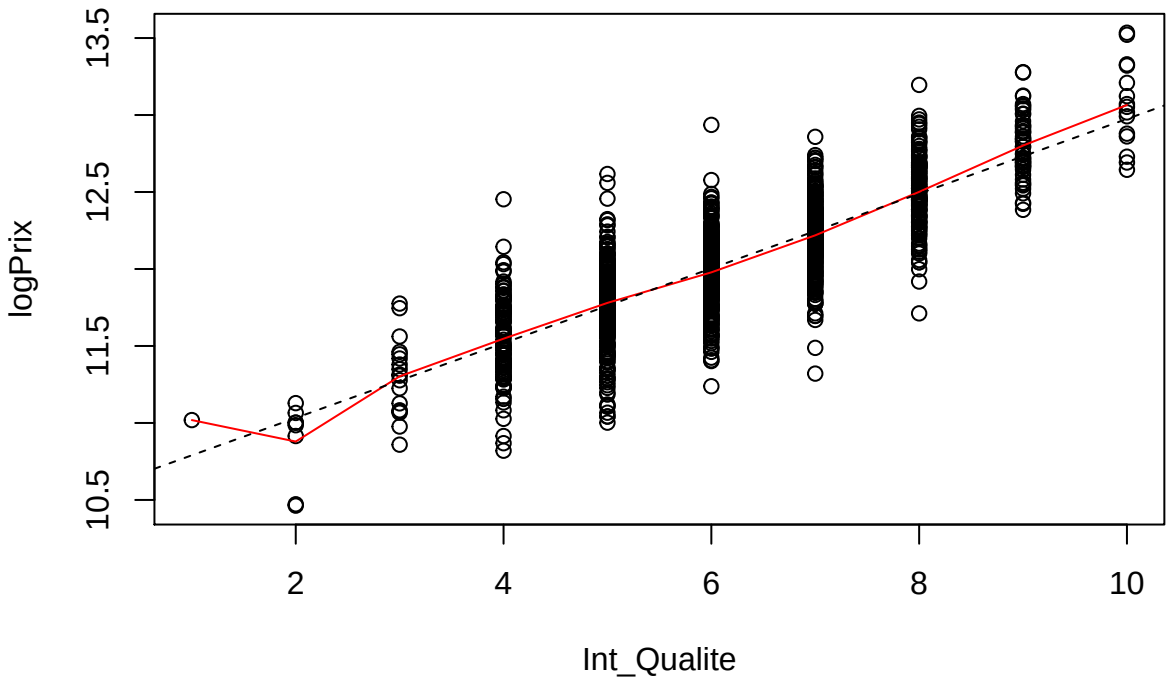
```
table(database$Int_Qualite)
```

```
##
##   1   2   3   4   5   6   7   8   9  10
##   1   8  18 106 416 357 293 160  51  15
```

```
A=with(data = database, aggregate(logPrix,by=list(Int_Qualite),FUN=mean))
A
```

```
##    Group.1        x
## 1        1 11.01863
## 2        2 10.87919
## 3        3 11.30263
## 4        4 11.54726
## 5        5 11.77987
## 6        6 11.97822
## 7        7 12.21854
## 8        8 12.50094
## 9        9 12.80182
## 10      10 13.06505
```

```
with(data = database, plot(Int_Qualite,logPrix))
lines(A[,1],A[,2],col="red")
reg_Int_Qualite=lm(logPrix~Int_Qualite,data=database)
abline(reg_Int_Qualite,lty=2)
```

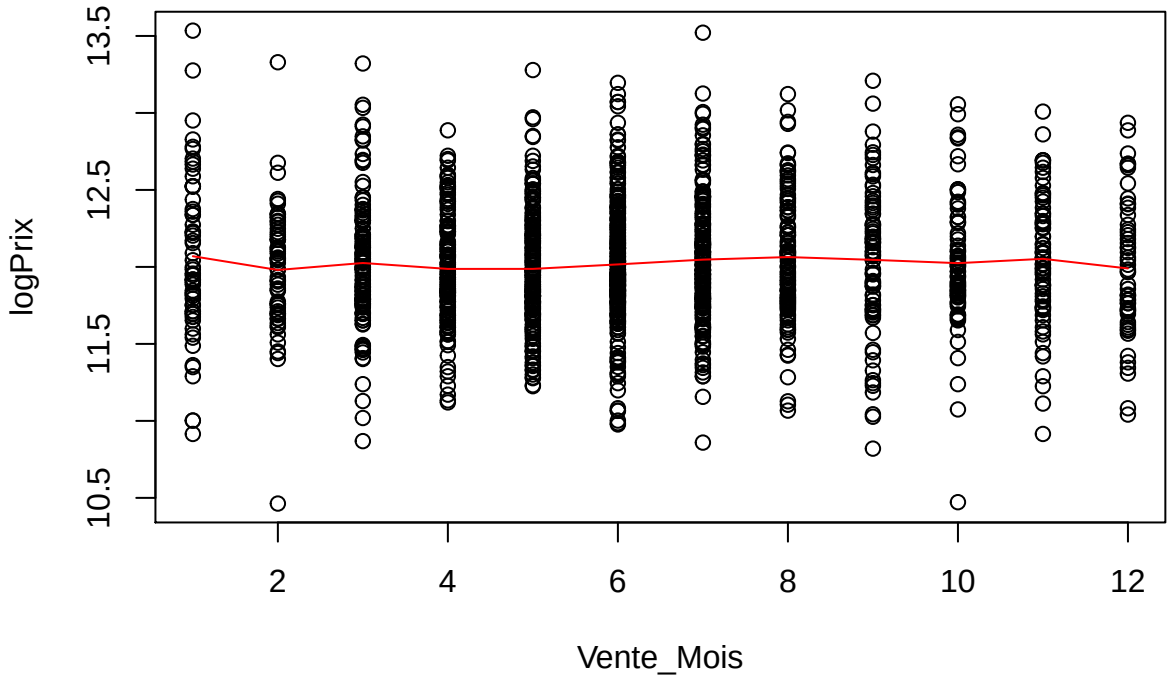


```r
summary(reg_Int_Qualite)
```

```
##
## Call:
## lm(formula = logPrix ~ Int_Qualite, data = database)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.547252   0.026776  393.90   <2e-16 ***
## Int_Qualite  0.242737   0.004295   56.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2264 on 1423 degrees of freedom
## Multiple R-squared:  0.6917, Adjusted R-squared:  0.6915
## F-statistic:  3193 on 1 and 1423 DF,  p-value: < 2.2e-16
```

Variable x `Vente_Mois`

```r
with(data = database, plot(Vente_Mois,logPrix ))
A=with(data = database, aggregate(logPrix,by=list(Vente_Mois),FUN=mean))
lines(A$Group.1,A$x,col="red")
```



```r
summary(lm(logPrix~Vente_Mois,data=database))
```

```
##
## Call:
## lm(formula = logPrix ~ Vente_Mois, data = database)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5615 -0.2508 -0.0415  0.2483  1.5280
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.003419   0.026627 450.800   <2e-16 ***
## Vente_Mois   0.003002   0.003953   0.759    0.448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4078 on 1423 degrees of freedom
## Multiple R-squared:  0.0004051,  Adjusted R-squared:  -0.0002974
## F-statistic: 0.5766 on 1 and 1423 DF,  p-value: 0.4478
```

```r
summary(lm(logPrix~as.factor(Vente_Mois),data=database))
```

```
##
## Call:
## lm(formula = logPrix ~ as.factor(Vente_Mois), data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55254 -0.25089 -0.03578  0.25575  1.47390
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              12.069514   0.052278 230.871   <2e-16 ***
## as.factor(Vente_Mois)2   -0.089363   0.073343  -1.218    0.223
## as.factor(Vente_Mois)3   -0.044045   0.063606  -0.692    0.489
## as.factor(Vente_Mois)4   -0.082557   0.062506  -1.321    0.187
## as.factor(Vente_Mois)5   -0.082382   0.060244  -1.367    0.172
## as.factor(Vente_Mois)6   -0.053233   0.058174  -0.915    0.360
## as.factor(Vente_Mois)7   -0.022278   0.059517  -0.374    0.708
## as.factor(Vente_Mois)8   -0.005624   0.065848  -0.085    0.932
## as.factor(Vente_Mois)9   -0.023625   0.069218  -0.341    0.733
## as.factor(Vente_Mois)10 -0.045021   0.070398  -0.640    0.523
## as.factor(Vente_Mois)11 -0.017040   0.071517  -0.238    0.812
## as.factor(Vente_Mois)12 -0.079561   0.076291  -1.043    0.297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4083 on 1413 degrees of freedom
## Multiple R-squared:  0.004813,   Adjusted R-squared:  -0.002934
## F-statistic: 0.6213 on 11 and 1413 DF,  p-value: 0.812
```

Variable x `I_piscine`

```
database$I_piscine = (database$Piscine_Surface>0)
with(database, mean(logPrix))
```

```
## [1] 12.0219
```

```
with(database, mean(logPrix[I_piscine==TRUE]))
```

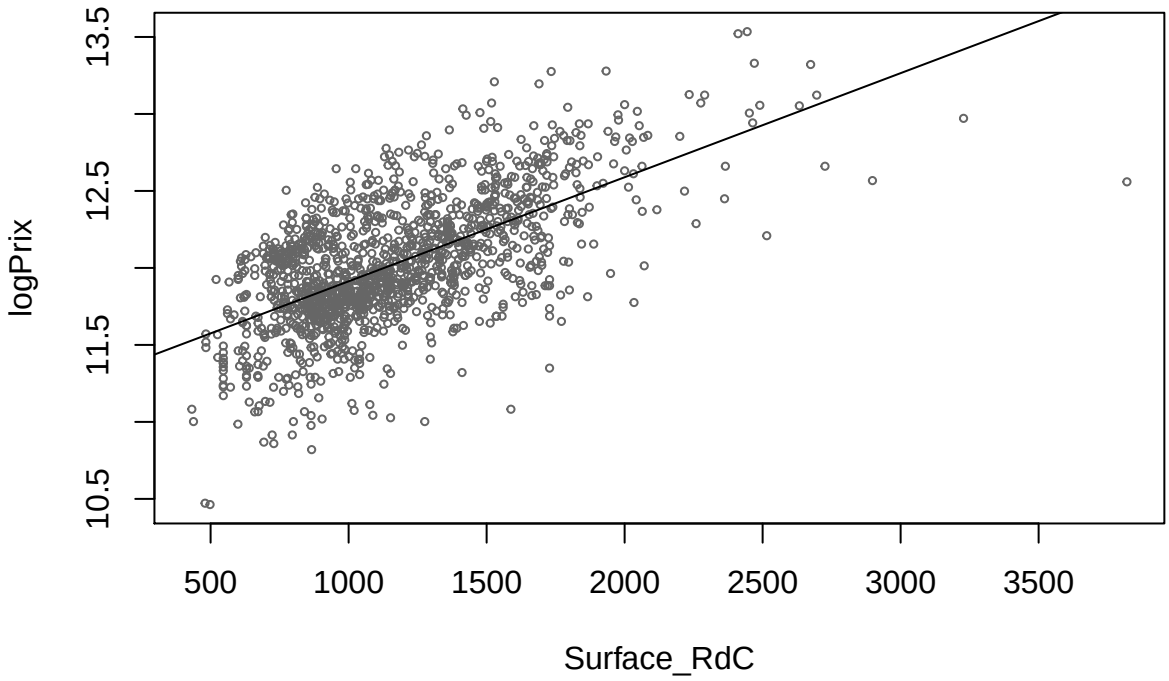
```
## [1] 12.67577
```

```
with(database, mean(logPrix[I_piscine==FALSE]))
```

```
## [1] 12.01913
```

Variable x `Surface_RdC`

```
with(data = database, plot( Surface_RdC,logPrix,cex=.5,col=grey(.4)))
reg_Surface_RdC = lm(logPrix ~ Surface_RdC, data = database)
abline(reg_Surface_RdC)
```

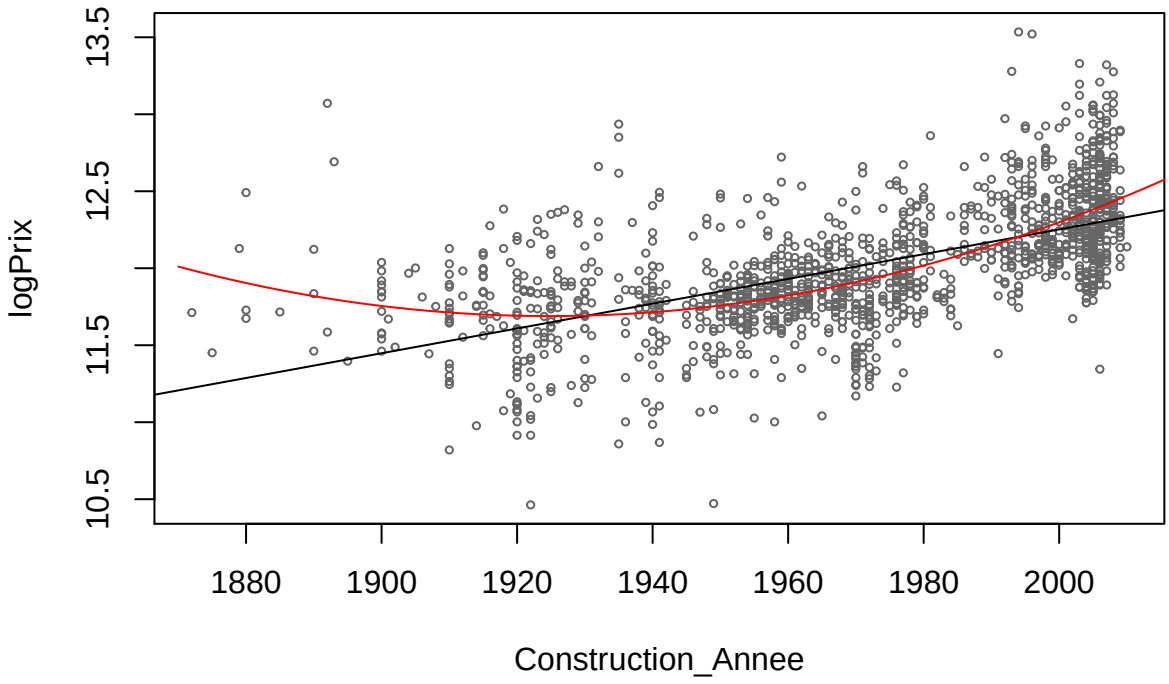



```
summary(reg_Surface_RdC)
```

```
##
## Call:
## lm(formula = logPrix ~ Surface_RdC, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26020 -0.18063 -0.01767  0.21851  0.93906
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.124e+01  2.697e-02  416.62   <2e-16 ***
## Surface_RdC 6.760e-04  2.208e-05   30.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3166 on 1423 degrees of freedom
## Multiple R-squared:  0.3972, Adjusted R-squared:  0.3968
## F-statistic: 937.8 on 1 and 1423 DF,  p-value: < 2.2e-16
```

Variable x `Construction_Annee`

```
with(data = database, plot( Construction_Annee,logPrix,cex=.5,col=grey(.4)))
reg_Construction_Annee = lm(logPrix ~ Construction_Annee, data = database)
abline(reg_Construction_Annee)
u=seq(1870,2020)
reg_Construction_Annee_2 = lm(logPrix ~ Construction_Annee+I(Construction_Annee^2), data = database)
v = predict(reg_Construction_Annee_2,newdata=data.frame(Construction_Annee=u))
lines(u,v,col="red")
```



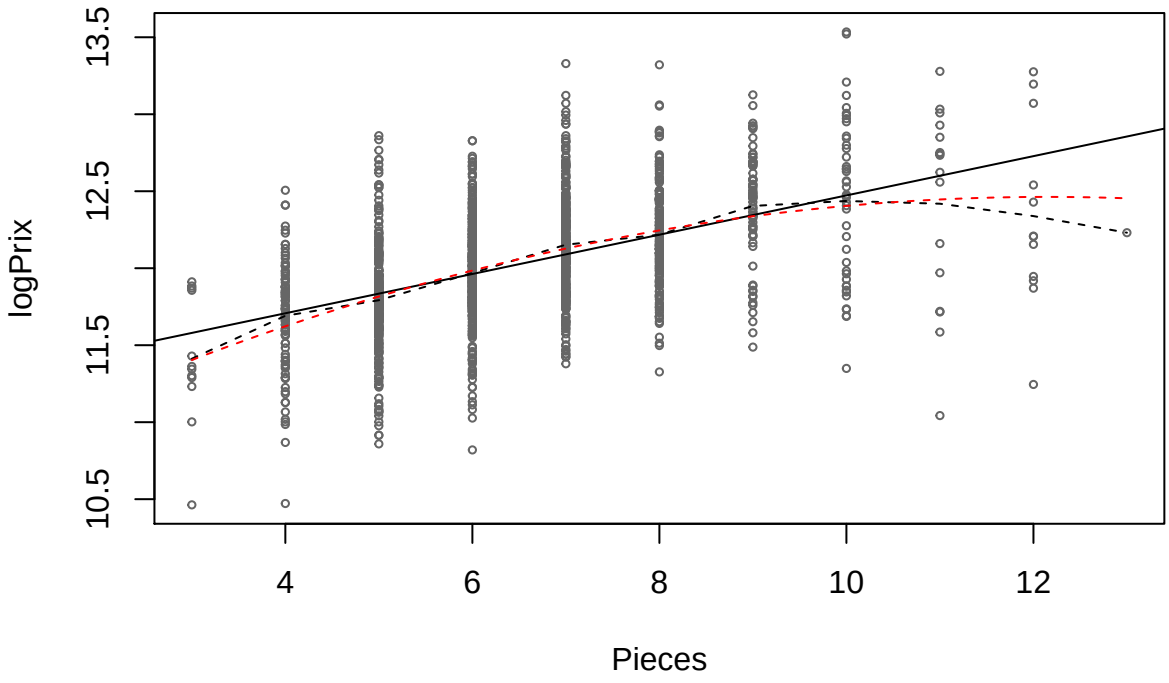

```
summary(reg_Construction_Annee_2)
```

```
##
## Call:
## lm(formula = logPrix ~ Construction_Annee + I(Construction_Annee^2),
##     data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28085 -0.17688 -0.01207  0.17147  1.32904
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               4.097e+02  3.170e+01   12.92   <2e-16 ***
## Construction_Annee       -4.135e-01  3.232e-02  -12.80   <2e-16 ***
## I(Construction_Annee^2)   1.074e-04  8.236e-06   13.04   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3093 on 1422 degrees of freedom
## Multiple R-squared:  0.4252, Adjusted R-squared:  0.4244
## F-statistic: 525.9 on 2 and 1422 DF,  p-value: < 2.2e-16
```

Variable x `Pieces`

```
with(data = database, plot(Pieces,logPrix ,cex=.5,col=grey(.4)))
A = with(data = database, aggregate(logPrix,by=list(Pieces),FUN=mean))
lines(A$Group.1,A$x,col="black",lty=2)
abline(lm(logPrix~Pieces,data=database))
reg_Pieces_2 = lm(logPrix~poly(Pieces,2),data=database)
u=seq(3,13,by=.1)
v=predict(reg_Pieces_2,newdata =data.frame(Pieces = u))
lines(u,v,lty=2,col="red")
```

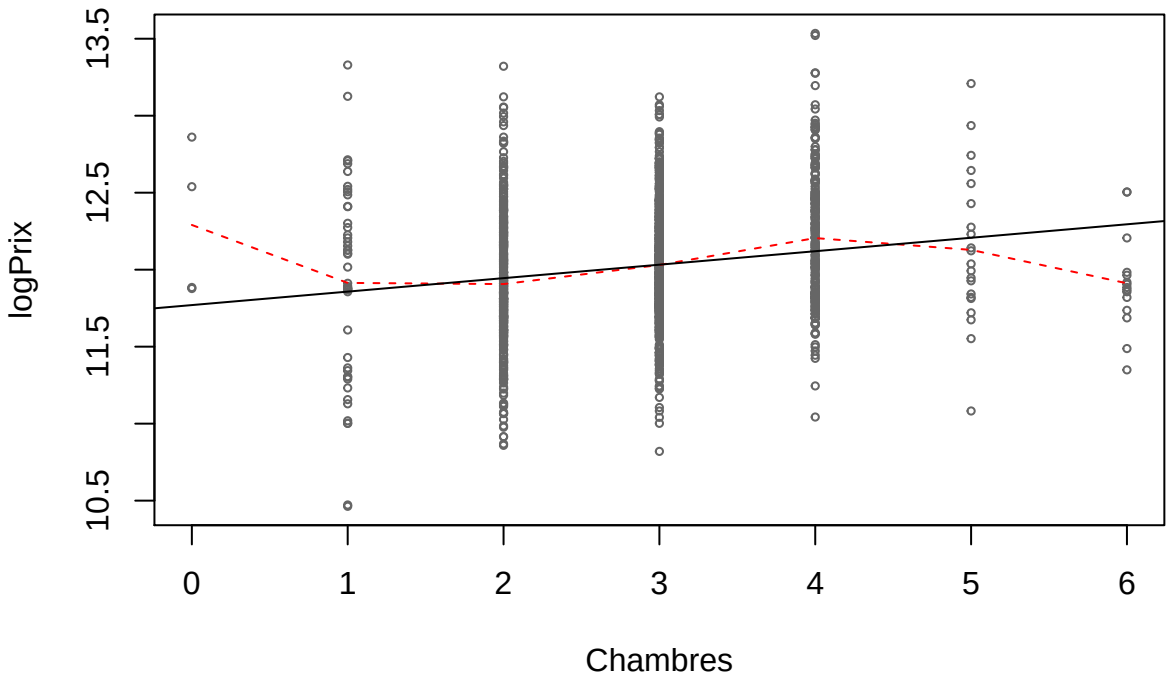


```r
summary(reg_Pieces_2)
```

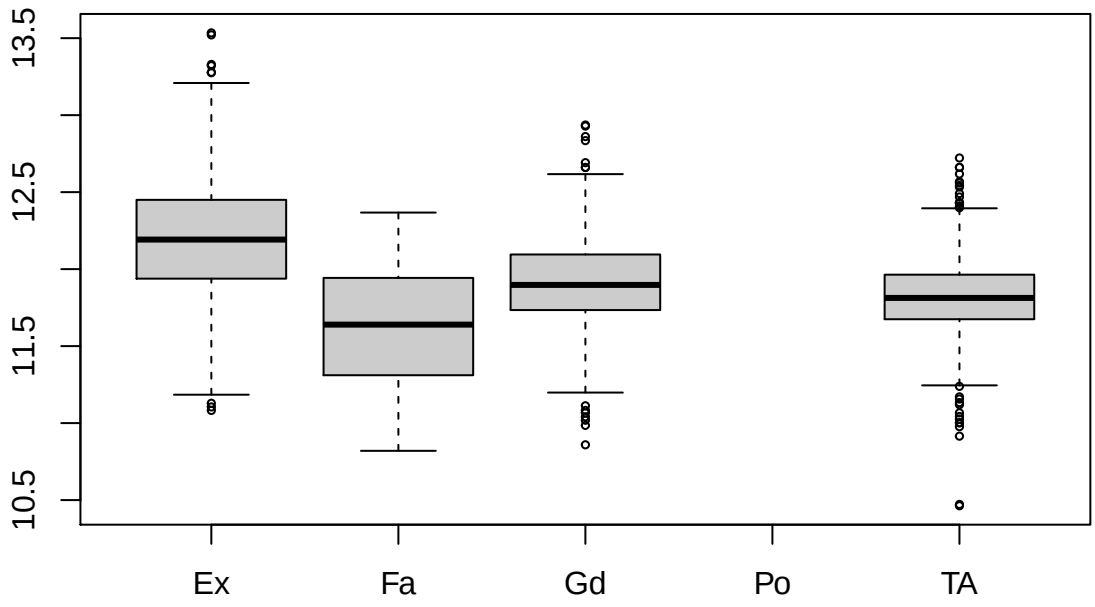
```
##
## Call:
## lm(formula = logPrix ~ poly(Pieces, 2), data = database)
##
## Coefficients:
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      12.02190    0.00926 1298.311  < 2e-16 ***
## poly(Pieces, 2)1  7.71190    0.34954   22.063  < 2e-16 ***
## poly(Pieces, 2)2 -1.86829    0.34954   -5.345 1.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3495 on 1422 degrees of freedom
## Multiple R-squared:  0.266,  Adjusted R-squared:  0.265
## F-statistic: 257.7 on 2 and 1422 DF,  p-value: < 2.2e-16
```

Variable x `Chambres`

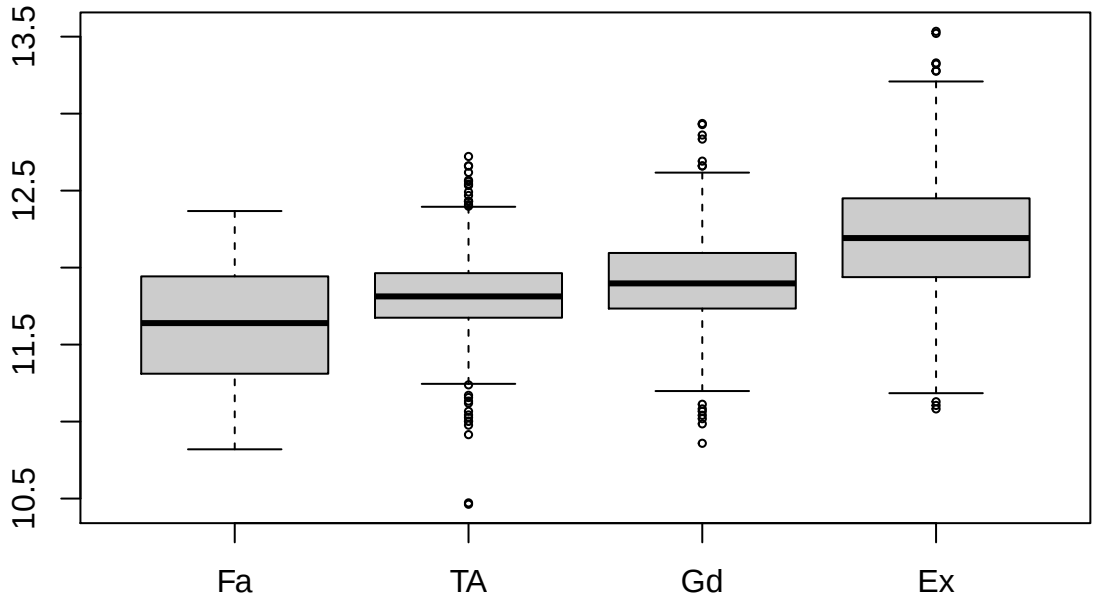
```r
with(data = database, plot(Chambres,logPrix ,cex=.5,col=grey(.4)))
A = with(data = database, aggregate(logPrix,by=list(Chambres),FUN=mean))
lines(A$Group.1,A$x,col="red",lty=2)
abline(lm(logPrix~Chambres,data=database))
```



```r
with(data = database, boxplot(logPrix~Chauff_Qualite,cex=.5,col=grey(.8)))
```




```r
A = with(data = database, aggregate(logPrix,by=list(Chauff_Qualite),FUN=mean))
A = A[order(A$x),]
L = as.character(A$Group.1)
database$Chauff_Qualite = factor(database$Chauff_Qualite, level=L)
with(data = database, boxplot(logPrix~Chauff_Qualite,cex=.5,col=grey(.8)))
```




```
library(car)
```

```
## Loading required package: carData
```

```
reg_Chauff_Qualite = lm(logPrix~Chauff_Qualite, data = database)
summary(reg_Chauff_Qualite)
```

```
##
## Call:
## lm(formula = logPrix ~ Chauff_Qualite, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34381 -0.22010 -0.00356  0.20817  1.32686
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.61399    0.05466 212.482  < 2e-16 ***
## Chauff_QualiteTA  0.19292    0.05739   3.361 0.000797 ***
## Chauff_QualiteGd  0.30052    0.05930   5.068 4.55e-07 ***
## Chauff_QualiteEx  0.59362    0.05627  10.550  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3584 on 1421 degrees of freedom
## Multiple R-squared:  0.2288, Adjusted R-squared:  0.2272
## F-statistic: 140.5 on 3 and 1421 DF,  p-value: < 2.2e-16
```

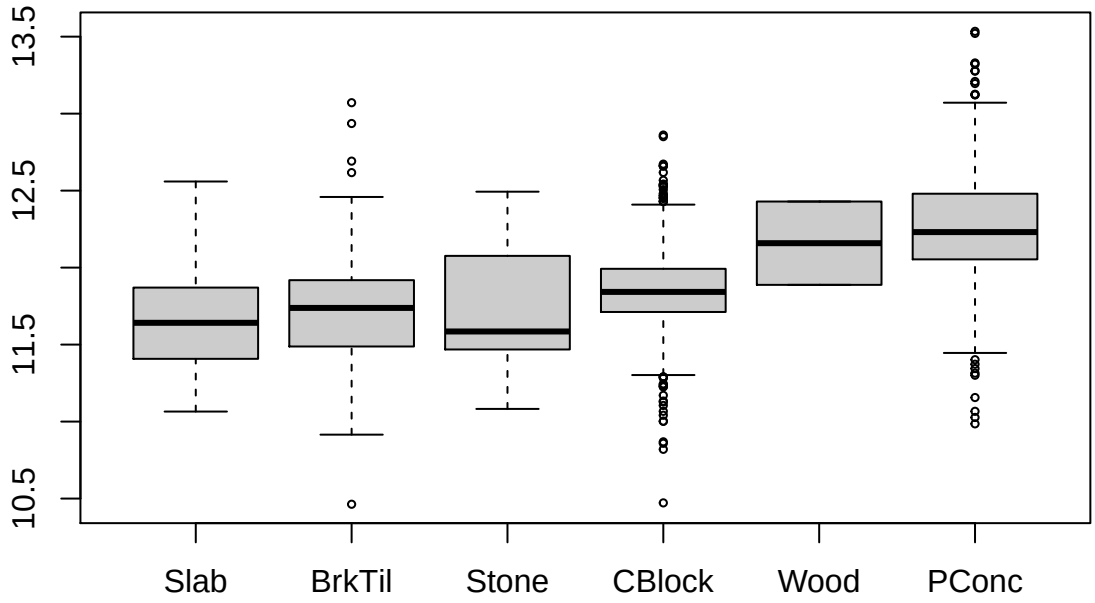
```r
linearHypothesis(reg_Chauff_Qualite, "Chauff_QualiteTA = Chauff_QualiteGd")
```

```
## Linear hypothesis test
##
## Hypothesis:
## Chauff_QualiteTA - Chauff_QualiteGd = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Chauff_Qualite
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1422 184.33
## 2   1421 182.55  1    1.7809 13.863 0.0002043 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
database$Chauff_Qualite= relevel(database$Chauff_Qualite,"TA")
reg_Chauff_Qualite = lm(logPrix~Chauff_Qualite, data = database)
summary(reg_Chauff_Qualite)
```

```
##
## Call:
## lm(formula = logPrix ~ Chauff_Qualite, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34381 -0.22010 -0.00356  0.20817  1.32686
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.80691    0.01751 674.296  < 2e-16 ***
## Chauff_QualiteFa -0.19292    0.05739  -3.361 0.000797 ***
## Chauff_QualiteGd  0.10761    0.02890   3.723 0.000204 ***
## Chauff_QualiteEx  0.40070    0.02202  18.195  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3584 on 1421 degrees of freedom
## Multiple R-squared:  0.2288, Adjusted R-squared:  0.2272
## F-statistic: 140.5 on 3 and 1421 DF,  p-value: < 2.2e-16
```

```r
A = with(data = database, aggregate(logPrix,by=list(Foundation),FUN=mean))
A = A[order(A$x),]
L = as.character(A$Group.1)
database$Foundation = factor(database$Foundation, level=L)
with(data = database, boxplot(logPrix~Foundation,cex=.5,col=grey(.8)))
```




```
reg_Foundation = lm(logPrix~Foundation, data = database)
summary(reg_Foundation)
```

```
##
## Call:
## lm(formula = logPrix ~ Foundation, data = database)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3804 -0.1824 -0.0173  0.1828  1.3516
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.65921    0.07407 157.413  < 2e-16 ***
## FoundationBrkTil  0.06026    0.07908   0.762   0.4462
## FoundationStone   0.09075    0.14814   0.613   0.5402
## FoundationCBlock  0.19310    0.07534   2.563   0.0105 *
## FoundationWood    0.49936    0.25118   1.988   0.0470 *
## FoundationPConc   0.60898    0.07527   8.090 1.27e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3394 on 1419 degrees of freedom
## Multiple R-squared:  0.3094, Adjusted R-squared:  0.3069
## F-statistic: 127.1 on 5 and 1419 DF,  p-value: < 2.2e-16
```

```r
linearHypothesis(reg_Foundation, c("FoundationBrkTil = 0",
                        "FoundationStone = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## FoundationBrkTil = 0
## FoundationStone = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Foundation
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1421 163.55
## 2   1419 163.48  2  0.076547 0.3322 0.7174
```

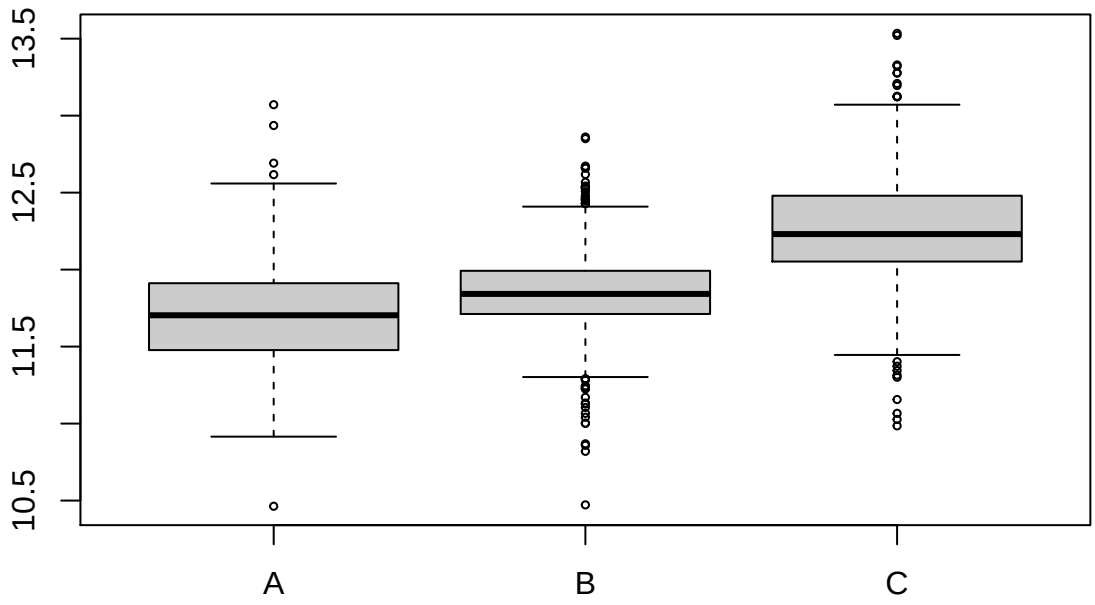
```r
linearHypothesis(reg_Foundation, c("FoundationBrkTil = 0",
                        "FoundationStone = 0",
                        "FoundationCBlock = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## FoundationBrkTil = 0
## FoundationStone = 0
## FoundationCBlock = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Foundation
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1422 166.20
## 2   1419 163.48  3    2.7253 7.8854 3.224e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

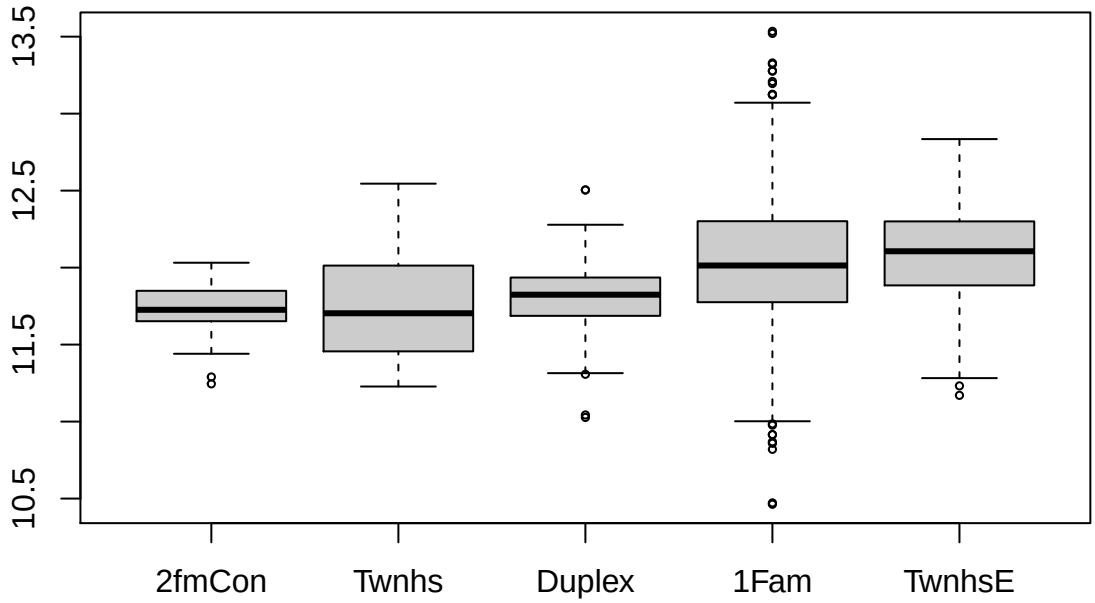
```r
linearHypothesis(reg_Foundation, "FoundationPConc = FoundationWood")
```

```
## Linear hypothesis test
##
## Hypothesis:
## - FoundationWood  + FoundationPConc = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Foundation
##
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1   1420 163.50
## 2   1419 163.48  1  0.023958 0.208 0.6484
```

```r
levels(database$Foundation) = c(rep("A",3),"B",rep("C",2))
with(data = database, boxplot(logPrix~Foundation,cex=.5,col=grey(.8)))
```




```
A = with(data = database, aggregate(logPrix,by=list(Logement),FUN=mean))
A = A[order(A$x),]
L = as.character(A$Group.1)
database$Logement = factor(database$Logement, level=L)
with(data = database, boxplot(logPrix~Logement,cex=.5,col=grey(.8)))
```




```
reg_Logement = lm(logPrix~Logement, data = database)
summary(reg_Logement)
```

```
##
## Call:
## lm(formula = logPrix ~ Logement, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58090 -0.25345 -0.02486  0.23840  1.49047
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.71174    0.07295 160.549  < 2e-16 ***
## LogementTwnhs   0.04037    0.09505   0.425    0.671
## LogementDuplex  0.08956    0.08960   1.000    0.318
## Logement1Fam    0.33226    0.07387   4.498 7.42e-06 ***
## LogementTwnhsE  0.38036    0.08222   4.626 4.06e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3996 on 1420 degrees of freedom
## Multiple R-squared:  0.04229,    Adjusted R-squared:  0.0396
## F-statistic: 15.68 on 4 and 1420 DF,  p-value: 1.467e-12
```

```
linearHypothesis(reg_Logement, c("LogementTwnhs = 0",
                        "LogementDuplex = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## LogementTwnhs = 0
## LogementDuplex = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Logement
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1422 226.86
## 2   1420 226.69  2   0.16999 0.5324 0.5873
```

```
linearHypothesis(reg_Logement, "Logement1Fam = LogementTwnhsE")
```

```
## Linear hypothesis test
##
## Hypothesis:
## Logement1Fam - LogementTwnhsE = 0
##
## Model 1: restricted model
## Model 2: logPrix ~ Logement
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1421 226.93
## 2   1420 226.69  1   0.23473 1.4703 0.2255
```

```
levels(database$Logement) = c(rep("A",3),rep("B",2))
```

```
reg_total_1 = lm(logPrix ~ Surface_Lot + Surface_RdC + Surface_Etage +
          Int_Qualite + Construction_Annee + I(Construction_Annee^2) +
           Pieces+I(Pieces^2) + Foyer + Int_Condition + Logement + Ext_Qualite +
           Cuisine_Qualite + Foundation, data = database)
summary(reg_total_1)
```

```
##
## Call:
## lm(formula = logPrix ~ Surface_Lot + Surface_RdC + Surface_Etage +
##     Int_Qualite + Construction_Annee + I(Construction_Annee^2) +
##     Pieces + I(Pieces^2) + Foyer + Int_Condition + Logement +
##     Ext_Qualite + Cuisine_Qualite + Foundation, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.78636 -0.07063  0.00626  0.07997  0.46040
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.737e+01  2.117e+01   1.766 0.077684 .
## Surface_Lot              6.248e-06  6.649e-07   9.397  < 2e-16 ***
## Surface_RdC              4.189e-04  1.705e-05  24.570  < 2e-16 ***
## Surface_Etage            2.824e-04  1.510e-05  18.704  < 2e-16 ***
## Int_Qualite              8.256e-02  4.791e-03  17.233  < 2e-16 ***
## Construction_Annee      -3.111e-02  2.169e-02  -1.434 0.151682
## I(Construction_Annee^2)  8.859e-06  5.556e-06   1.595 0.111033
## Pieces                   3.113e-02  1.385e-02   2.248 0.024762 *
```

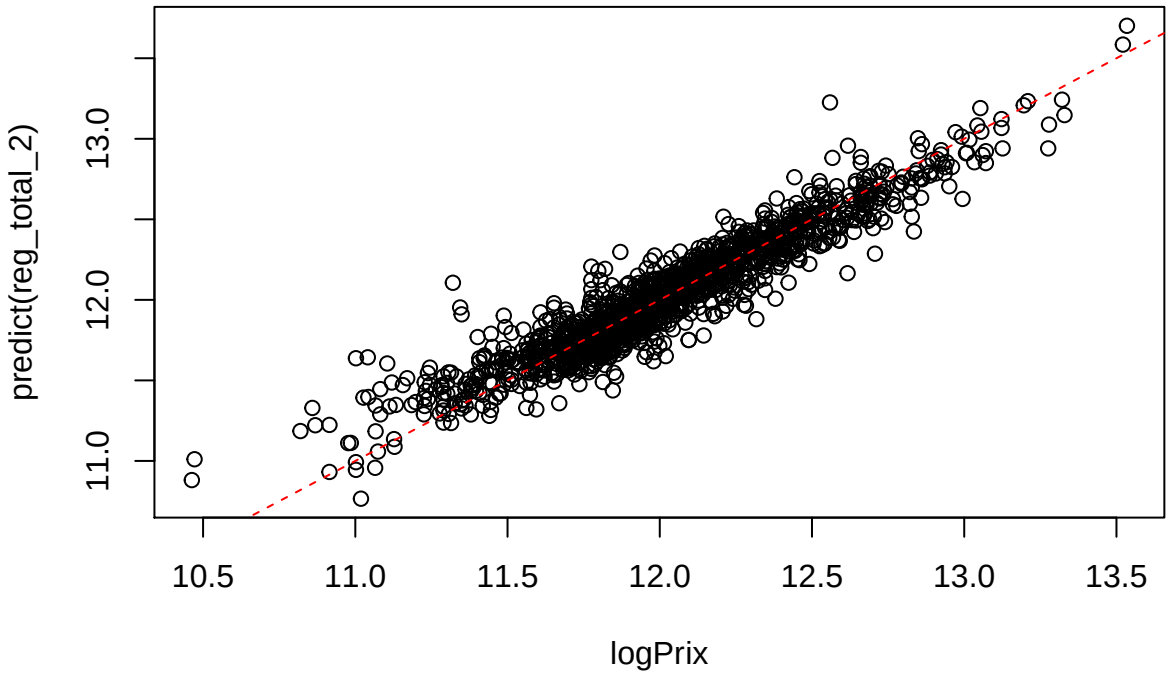
```
## I(Pieces^2)             -2.712e-03  9.522e-04  -2.849 0.004455 **
## Foyer                    4.818e-02  6.647e-03   7.249 6.93e-13 ***
## Int_Condition            5.828e-02  3.737e-03  15.594  < 2e-16 ***
## LogementB                7.874e-02  1.333e-02   5.909 4.31e-09 ***
## Ext_QualiteFa           -1.589e-01  4.929e-02  -3.225 0.001289 **
## Ext_QualiteGd           -6.880e-02  2.391e-02  -2.877 0.004078 **
## Ext_QualiteTA           -9.407e-02  2.697e-02  -3.488 0.000502 ***
## Cuisine_QualiteFa       -1.100e-01  3.134e-02  -3.509 0.000464 ***
## Cuisine_QualiteGd       -5.296e-02  1.805e-02  -2.934 0.003401 **
## Cuisine_QualiteTA       -6.460e-02  1.993e-02  -3.241 0.001219 **
## FoundationB              3.059e-02  1.387e-02   2.205 0.027604 *
## FoundationC              6.242e-02  1.653e-02   3.777 0.000165 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1337 on 1405 degrees of freedom
## Multiple R-squared:  0.8939, Adjusted R-squared:  0.8925
## F-statistic: 623.1 on 19 and 1405 DF,  p-value: < 2.2e-16
```

```
reg_total_2 = lm(logPrix ~ Surface_Lot + Surface_RdC + Surface_Etage +
           Int_Qualite + Construction_Annee  +
           Pieces+I(Pieces^2) + Foyer + Int_Condition + Logement + Ext_Qualite +
           Cuisine_Qualite + Foundation , data = database)
summary(reg_total_2)
```

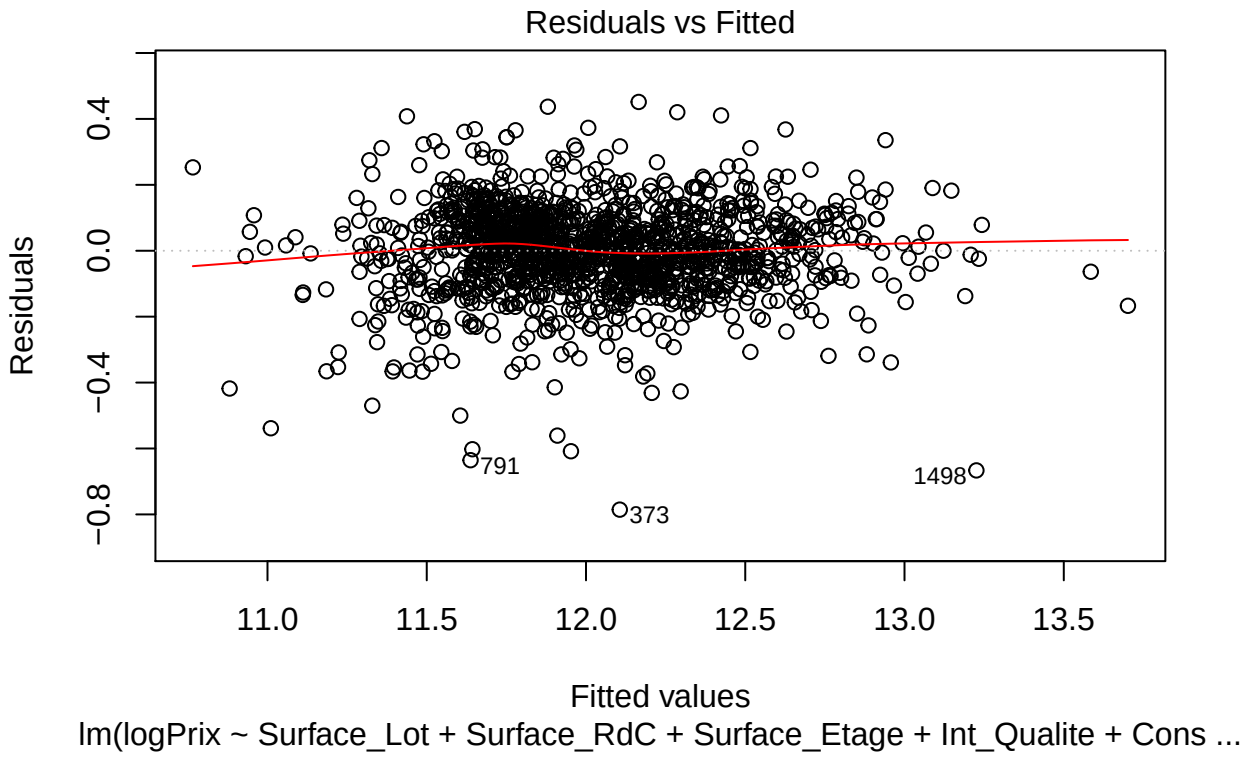
```
##
## Call:
## lm(formula = logPrix ~ Surface_Lot + Surface_RdC + Surface_Etage +
##     Int_Qualite + Construction_Annee + Pieces + I(Pieces^2) +
##     Foyer + Int_Condition + Logement + Ext_Qualite + Cuisine_Qualite +
##     Foundation, data = database)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.78549 -0.07269  0.00567  0.08088  0.45143
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.626e+00  4.178e-01   8.678  < 2e-16 ***
## Surface_Lot          6.215e-06  6.649e-07   9.347  < 2e-16 ***
## Surface_RdC          4.166e-04  1.700e-05  24.509  < 2e-16 ***
## Surface_Etage        2.824e-04  1.511e-05  18.693  < 2e-16 ***
## Int_Qualite          8.451e-02  4.634e-03  18.237  < 2e-16 ***
## Construction_Annee   3.473e-03  2.101e-04  16.526  < 2e-16 ***
## Pieces               3.165e-02  1.386e-02   2.284 0.022531 *
## I(Pieces^2)         -2.729e-03  9.527e-04  -2.865 0.004234 **
## Foyer                4.640e-02  6.556e-03   7.077 2.31e-12 ***
## Int_Condition        5.736e-02  3.694e-03  15.526  < 2e-16 ***
## LogementB            7.821e-02  1.333e-02   5.868 5.50e-09 ***
## Ext_QualiteFa       -1.641e-01  4.920e-02  -3.336 0.000873 ***
## Ext_QualiteGd       -7.024e-02  2.391e-02  -2.938 0.003360 **
## Ext_QualiteTA       -1.002e-01  2.671e-02  -3.751 0.000183 ***
## Cuisine_QualiteFa   -1.129e-01  3.131e-02  -3.606 0.000322 ***
## Cuisine_QualiteGd   -5.349e-02  1.806e-02  -2.962 0.003104 **
## Cuisine_QualiteTA   -6.684e-02  1.989e-02  -3.360 0.000801 ***
```

```
## FoundationB         2.490e-02  1.341e-02   1.856 0.063631 .
## FoundationC         6.621e-02  1.636e-02   4.046 5.49e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1338 on 1406 degrees of freedom
## Multiple R-squared:  0.8937, Adjusted R-squared:  0.8924
## F-statistic: 656.9 on 18 and 1406 DF,  p-value: < 2.2e-16
```

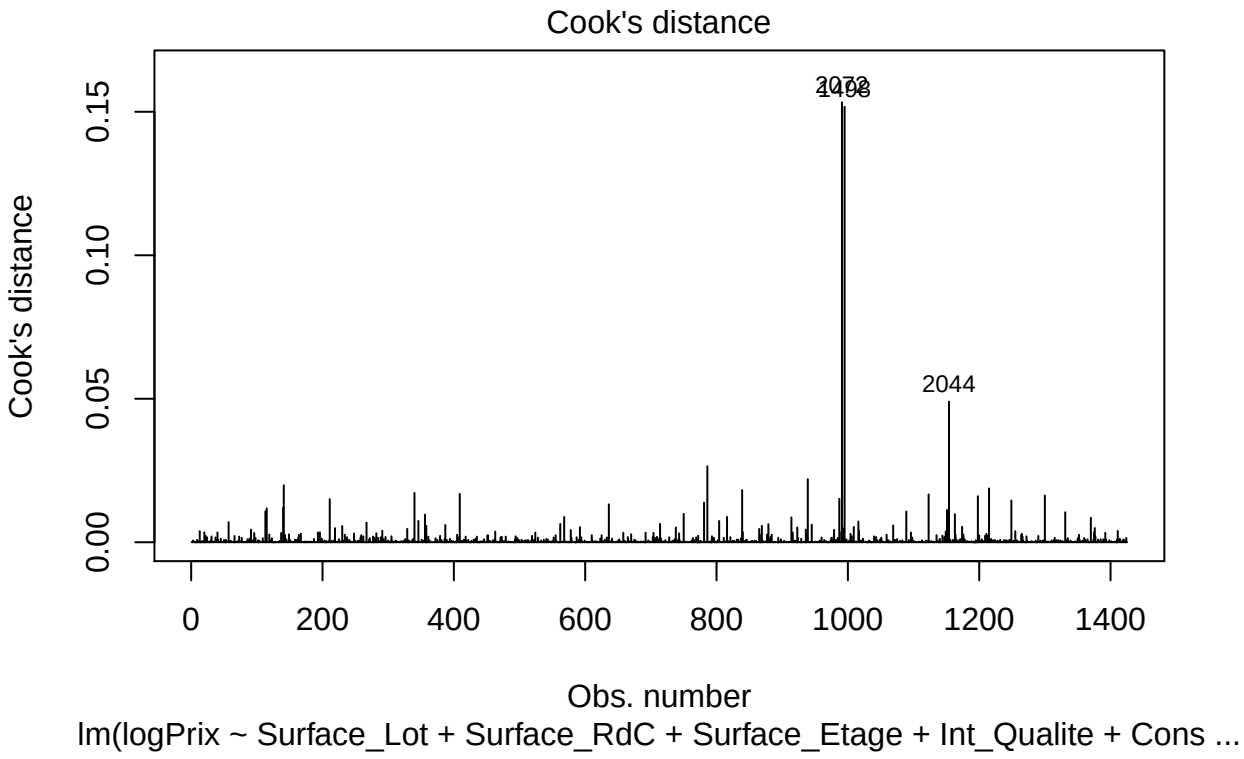
```
with(database,plot(logPrix,predict(reg_total_2)))
abline(a=0,b=1,lty=2,col="red")
```



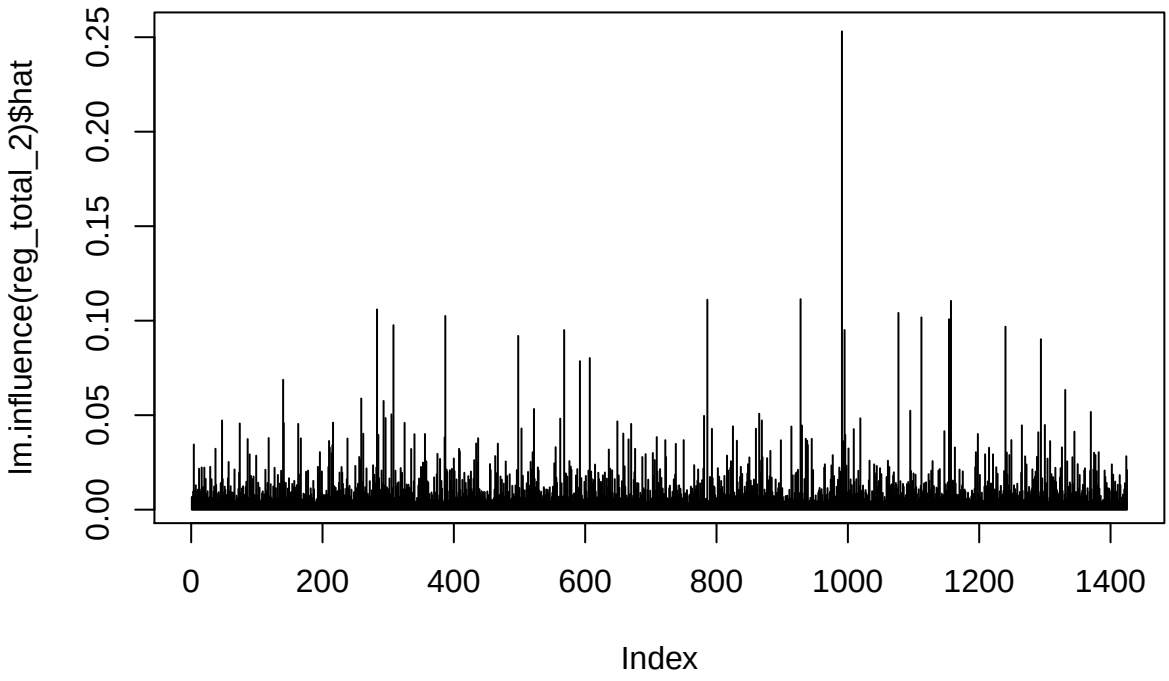

```
plot(reg_total_2,which=1)
```




```
plot(reg_total_2,which=4)
```




```
plot(lm.influence(reg_total_2)$hat,type="h")
```




```
lm.influence(reg_total_2)$hat[which(lm.influence(reg_total_2)$hat>.15)]
```

```
##      2072
## 0.2530952
```

```
database["373",]
```

```
##     Zone Surface_Lot  Rue Forme Utilities Configuration Proxim_1 Proxim_2
## 373   RL       11900 Pave   Reg    AllPub        Inside     Norm     Norm
##     Logement  Style Int_Qualite Int_Condition Construction_Annee
## 373        B 1Story           7             5               1977
##     Renovation_Annee Toit Exterieur Maconnerie Ext_Qualite Ext_Condition
## 373             1977  Hip   Plywood    BrkFace          TA            Gd
##     Foundation Chauffage Chauff_Qualite Electricite Surface_RdC Surface_Etage
## 373          B      GasA             TA       SBrkr        1411             0
##     Surface_Autre Toilettes Chambres Cuisine_Qualite Pieces Note Foyer Allee
## 373             0         2        3              TA      6  Typ     1     Y
##     Garage_Ext_Surface Garage_Int_Surface Piscine_Surface Vente_Mois
## 373                  0                 0               0          4
##     Vente_Annee Vente_Type Vente_Condition  logPrix I_piscine
## 373        2009        WD           Family 11.32055     FALSE
```

```
predict(reg_total_2,newdata=database["373",])
```

```
##      373
## 12.10604
```

```r
database["1498",]
```

```
##      Zone Surface_Lot  Rue Forme Utilities Configuration Proxim_1 Proxim_2
## 1498   RL       47007 Pave   IR1    AllPub        Inside     Norm     Norm
##      Logement  Style Int_Qualite Int_Condition Construction_Annee
## 1498        B 1Story           5             7               1959
##      Renovation_Annee  Toit Exterieur Maconnerie Ext_Qualite Ext_Condition
## 1498             1996 Gable   Plywood       None          TA            TA
##      Foundation Chauffage Chauff_Qualite Electricite Surface_RdC Surface_Etage
## 1498          A      GasA             TA       SBrkr        3820             0
##      Surface_Autre Toilettes Chambres Cuisine_Qualite Pieces Note Foyer Allee
## 1498             0         3        5              Ex     11  Typ     2     Y
##      Garage_Ext_Surface Garage_Int_Surface Piscine_Surface Vente_Mois
## 1498                372                  0               0          7
##      Vente_Annee Vente_Type Vente_Condition  logPrix I_piscine
## 1498        2008        WD           Normal 12.55919     FALSE
```

```r
predict(reg_total_2,newdata=database["1498",])
```

```
##     1498
## 13.22566
```

```r
database["2072",]
```

```
##      Zone Surface_Lot  Rue Forme Utilities Configuration Proxim_1 Proxim_2
## 2072   RL      115149 Pave   IR2    AllPub       CulDSac     Norm     Norm
##      Logement  Style Int_Qualite Int_Condition Construction_Annee
## 2072        B 1Story           7             5               1971
##      Renovation_Annee  Toit Exterieur Maconnerie Ext_Qualite Ext_Condition
## 2072             2002 Gable   Plywood      Stone          TA            TA
##      Foundation Chauffage Chauff_Qualite Electricite Surface_RdC Surface_Etage
## 2072          B      GasA             TA       SBrkr        1824             0
##      Surface_Autre Toilettes Chambres Cuisine_Qualite Pieces Note Foyer Allee
## 2072             0         2        2              Gd      5  Typ     2     Y
##      Garage_Ext_Surface Garage_Int_Surface Piscine_Surface Vente_Mois
## 2072                 48                  0               0          6
##      Vente_Annee Vente_Type Vente_Condition  logPrix I_piscine
## 2072        2007        WD           Normal 12.61818     FALSE
```

```r
predict(reg_total_2,newdata=database["2072",])
```

```
##     2072
## 12.95709
```