

Impact du CO2 sur la température de la planète

Pour de l'aide sur R Markdown on peut aller sur

<http://rmarkdown.rstudio.com>>

<https://lms.fun-mooc.fr/c4x/UPSUD/42001S02/asset/RMarkdown.html>

Introduction

On s'intéresse à la dépendance entre la température de surface de la planète et les émissions de CO2 et leurs évolutions dans le temps. L'analyse de la température de surface GISS ver. 4 (GISTEMP v4) est une estimation du changement global de la température de surface. Les données proviennent de NOAA GHCN v4 (stations météorologiques) et ERSST v5 (zones océaniques). Plus de détails sont accessibles sur ce lien <https://data.giss.nasa.gov/gistemp/>. Concernant le CO2, il est mesuré sur le Mauna Loa (sommet de l'archipel d'Hawai) depuis la fin des années 50. Détails sur <https://gml.noaa.gov/ccgg/trends/>

Le fichier "climat_CO2.txt" contient 3 colonnes :

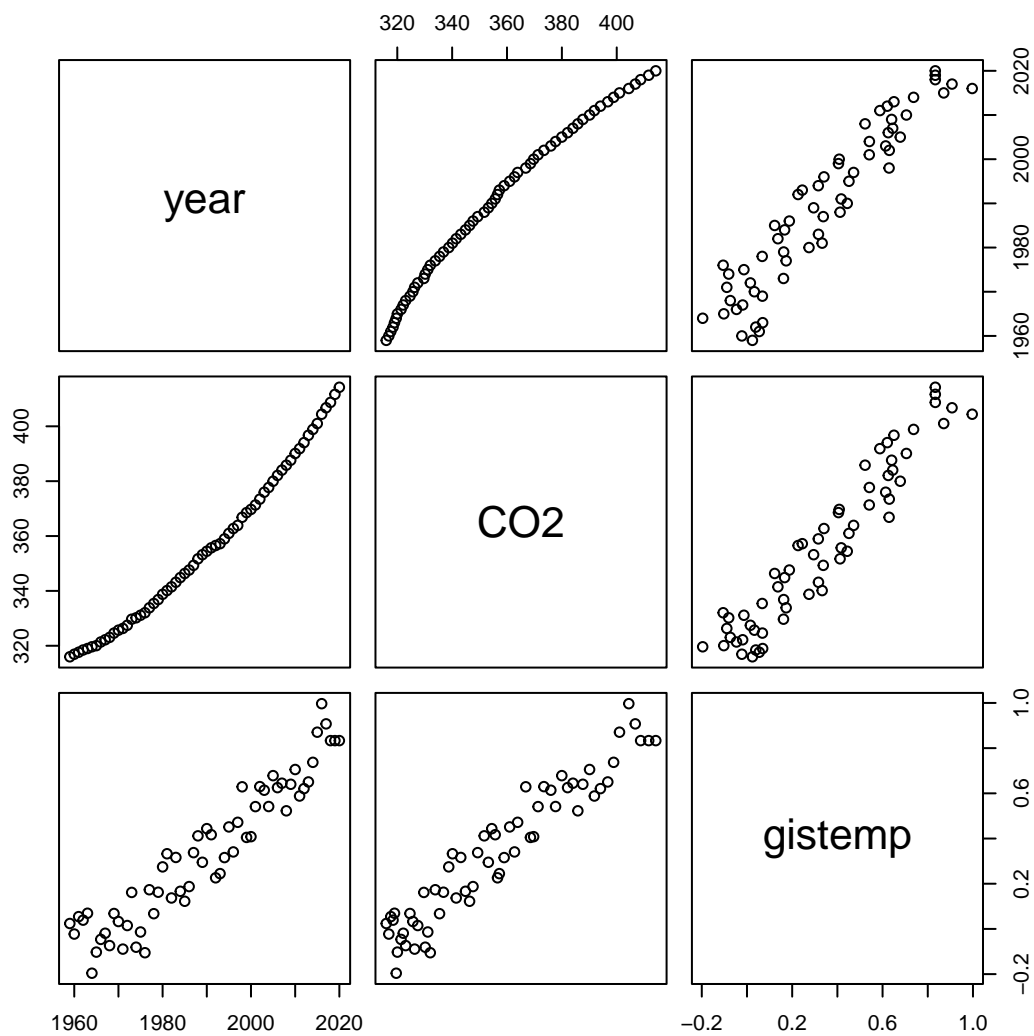
- **year** : années entre 1959 et 2020
- **CO2** : CO2 exprimé en fraction molaire dans l'air sec, micromol/mol, abrégé en ppm. Voir www.esrl.noaa.gov/gmd/ccgg/trends/ pour plus de détails.
- **gistemp** : anomalies de température en degrés Celsius. Il s'agit des écarts entre la température annuelle et la température moyenne de la période 1951-1980.

1 - On commence par lire et visualiser les données du fichier

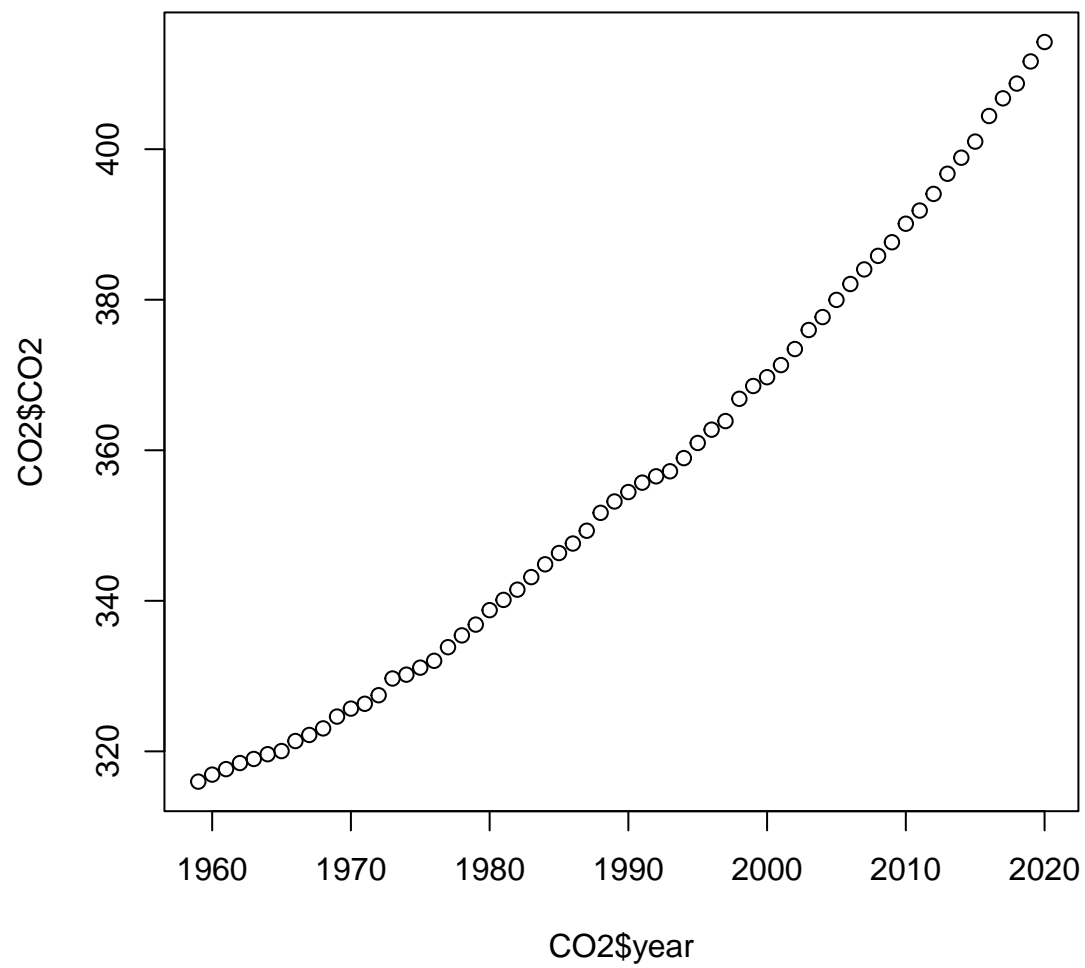
```
C02 = read.table(file = "climat_CO2.txt", header = TRUE)
head(C02)
```

```
##   year    C02 gistemp
## 1 1959 315.98  0.0240
## 2 1960 316.91 -0.0227
## 3 1961 317.64  0.0544
## 4 1962 318.45  0.0389
## 5 1963 318.99  0.0694
## 6 1964 319.62 -0.1960
```

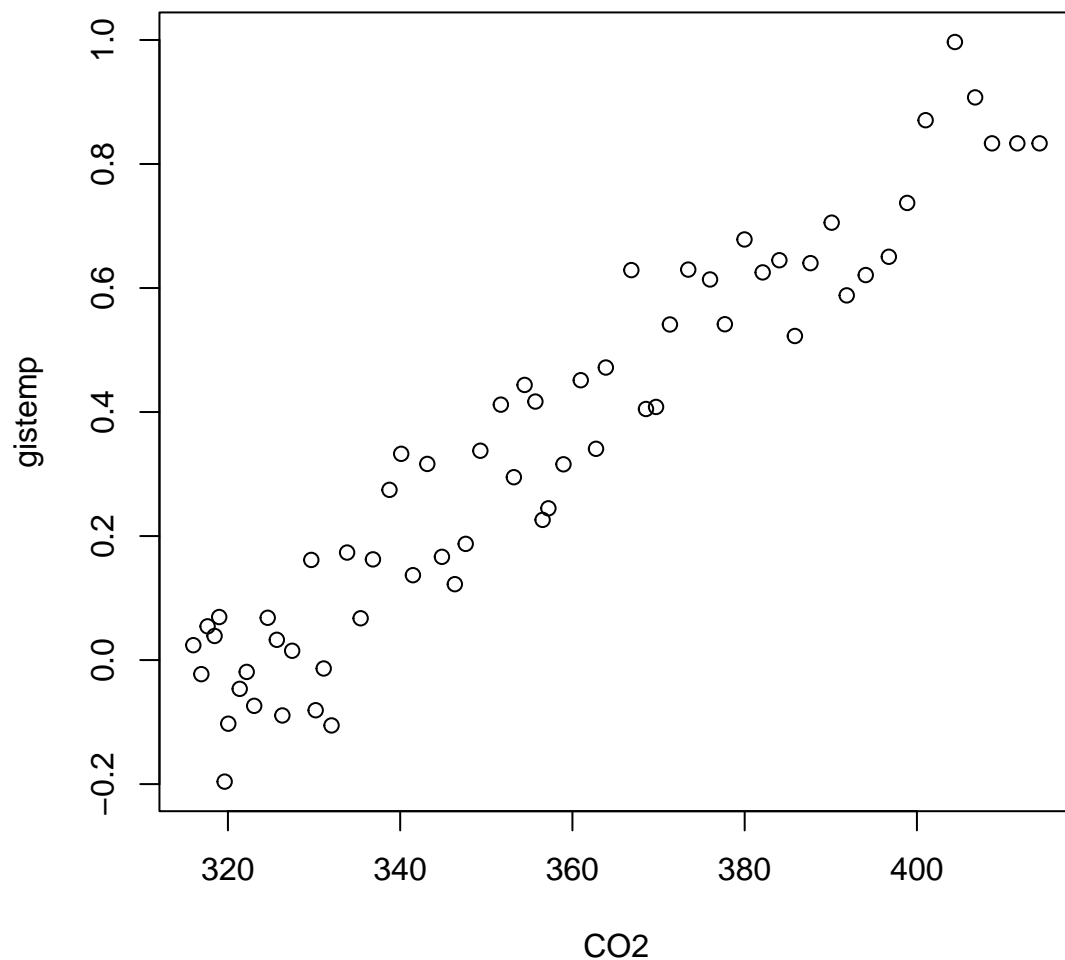
```
pairs(C02)
```



```
plot(CO2$year,CO2$CO2)
```



```
plot(gistemp~CO2, data = CO2)
```



On remarque que l'objet CO2 est une structure de données. L'accès aux différents champs se fait avec le \$. Mais la table CO2 peut aussi être utilisée comme une matrice. Les commandes pour accéder à une colonne, à une ligne, à un set d'indices sont alors les suivantes :

```
CO2[,3]
```

```
## [1] 0.0240 -0.0227 0.0544 0.0389 0.0694 -0.1960 -0.1026 -0.0464 -0.0191
## [10] -0.0738 0.0682 0.0327 -0.0893 0.0150 0.1615 -0.0809 -0.0136 -0.1054
## [19] 0.1734 0.0674 0.1624 0.2745 0.3326 0.1369 0.3163 0.1665 0.1224
## [28] 0.1874 0.3376 0.4119 0.2949 0.4437 0.4169 0.2261 0.2447 0.3157
## [37] 0.4513 0.3407 0.4718 0.6289 0.4049 0.4082 0.5412 0.6297 0.6138
## [46] 0.5416 0.6783 0.6252 0.6448 0.5226 0.6401 0.7054 0.5881 0.6209
## [55] 0.6504 0.7372 0.8706 0.9967 0.9074 0.8333 0.8333 0.8333
```

```
CO2[30,]
```

```
## year CO2 gistemp
```

```
## 30 1988 351.69 0.4119
```

```
indices = C02$gistemp>0.5  
C02[indices,1]
```

```
## [1] 1998 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014  
## [16] 2015 2016 2017 2018 2019 2020
```

```
C02[c(12, 35, 50),]
```

```
##      year      C02 gistemp  
## 12 1970 325.68 0.0327  
## 35 1993 357.21 0.2447  
## 50 2008 385.83 0.5226
```

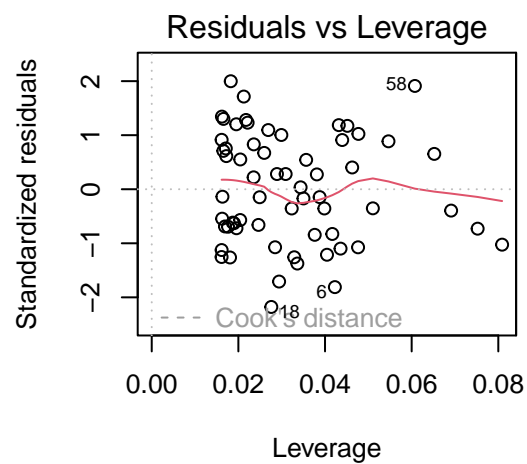
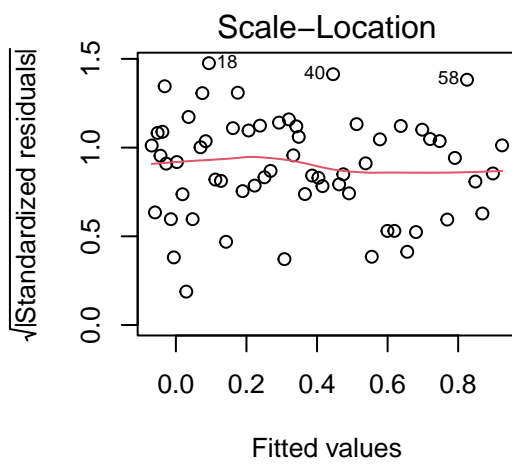
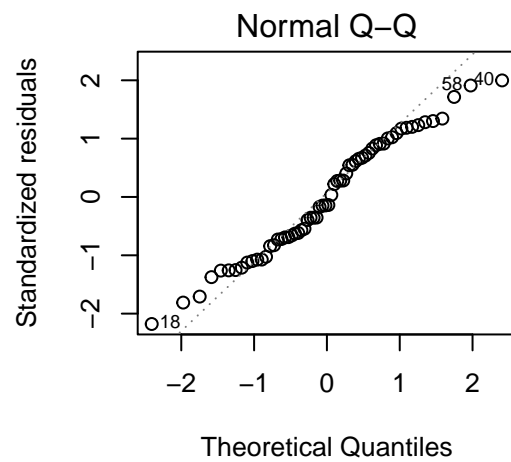
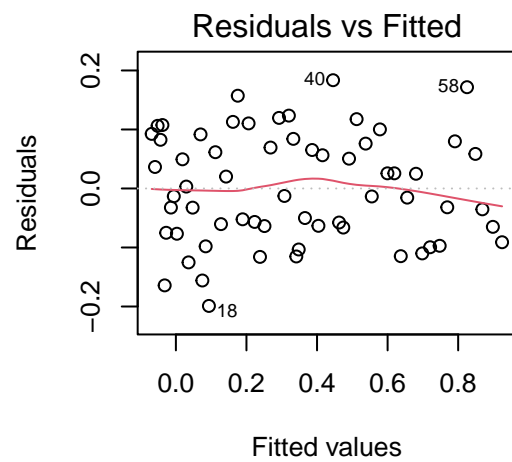
2 - On met en place un modèle de régression de la température en fonction du CO2

La commande `summary` permet d'avoir accès aux estimations des paramètres et aux tests statistiques. La commande `plot` permet l'analyse des résidus.

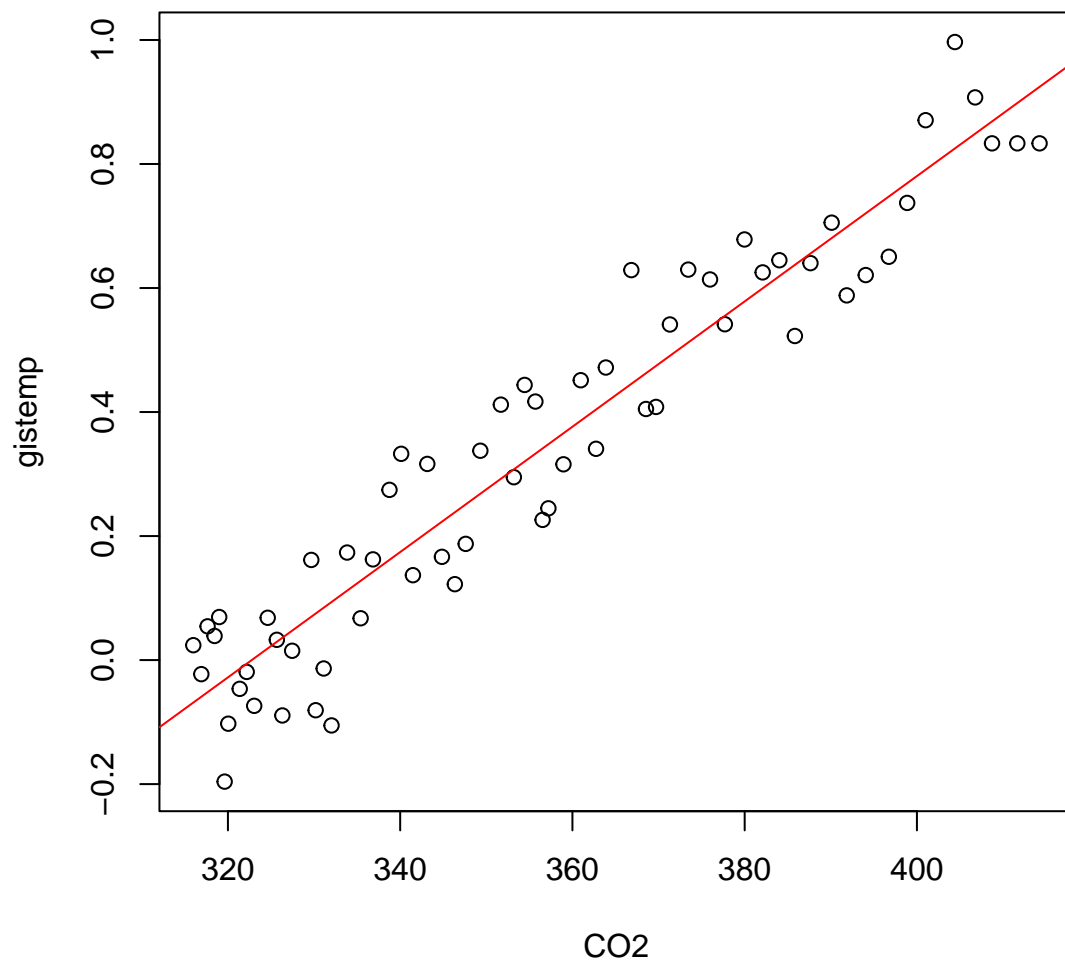
```
mod1 = lm(gistemp~C02, data = C02)  
summary(mod1)
```

```
##  
## Call:  
## lm(formula = gistemp ~ C02, data = C02)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.19904 -0.06591 -0.01294  0.07893  0.18349   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -3.2616369  0.1457453  -22.38  <2e-16 ***  
## C02          0.0101053  0.0004076   24.79  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.09268 on 60 degrees of freedom  
## Multiple R-squared:  0.9111, Adjusted R-squared:  0.9096   
## F-statistic: 614.6 on 1 and 60 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))  
plot(mod1)
```



```
plot(gistemp~C02, data = C02)
abline(mod1$coefficients, col = 'red')
```

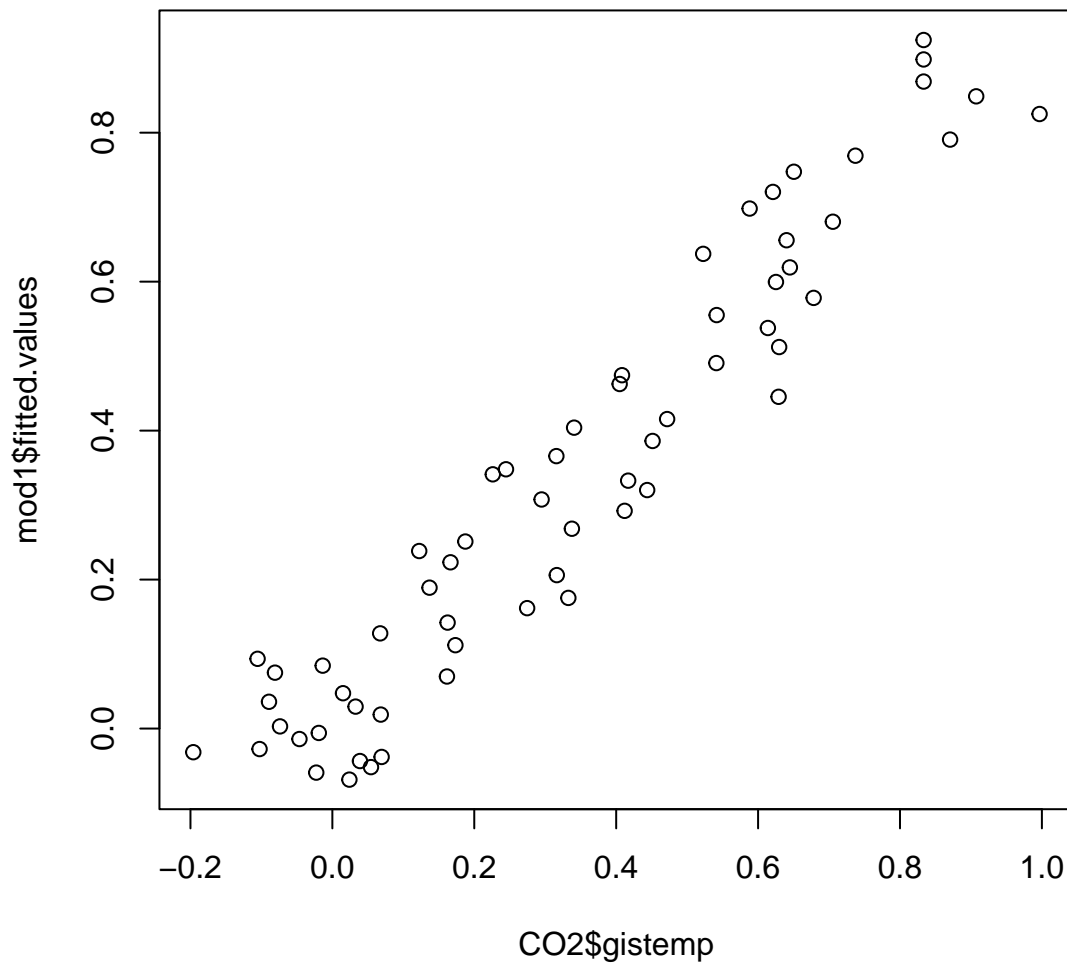


3 - Prédiction du modèle

La prédiction aux points d'observation

Le champ *fitted.values* de l'objet *CO2* contient les valeurs de $\hat{y}_i, \forall i \in \{1, \dots, n\}$.

```
plot(CO2$gistemp ,mod1$fitted.values)
```



On considère que les résidus traduisent un bruit de mesure. Question : quelle est la donnée ayant la plus grande erreur de mesure ?

```
ecarts = abs(CO2$gistemp - mod1$fitted.values)
indice = ecarts == max(ecarts)
CO2[indice,]
```

```
##      year      CO2 gistemp
## 18 1976 332.03 -0.1054
```

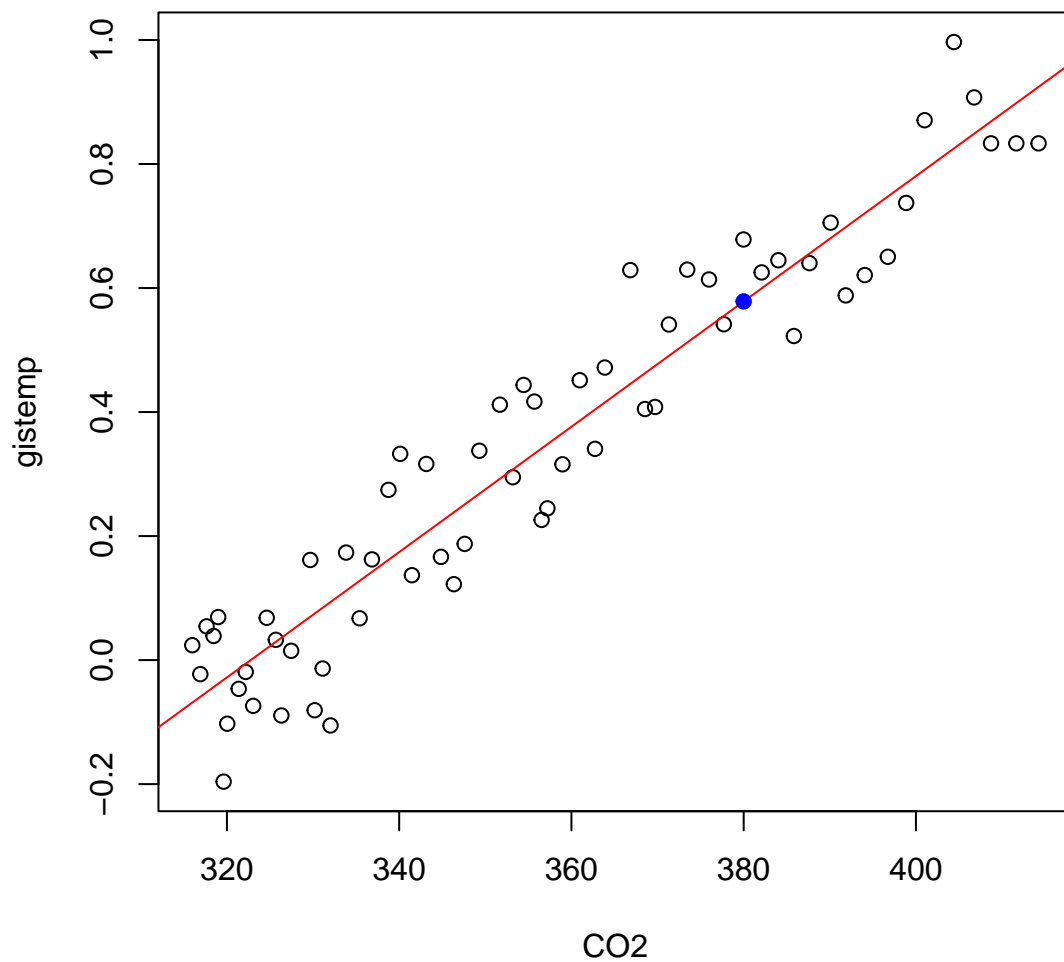
```
c(CO2$gistemp[indice] , mod1$fitted.values[indice])
```

```
##              18
## -0.10540000  0.09363919
```


La prédiction en dehors des points d'observation

On utilise la fonction *predict* sur un seul point :

```
nouvel_individu = data.frame(CO2 = 380)
prediction = predict(mod1, newdata = nouvel_individu)
plot(gistemp~CO2, data = CO2)
abline(mod1$coefficients, col = 'red')
points(nouvel_individu$CO2, prediction, col = "blue", pch = 19)
```



On peut aussi obtenir les intervalles de confiance et de prédiction :

```
prediction_IC = data.frame(predict(mod1, newdata = nouvel_individu, interval = 'confidence', level = 0.95))
prediction_IC
```

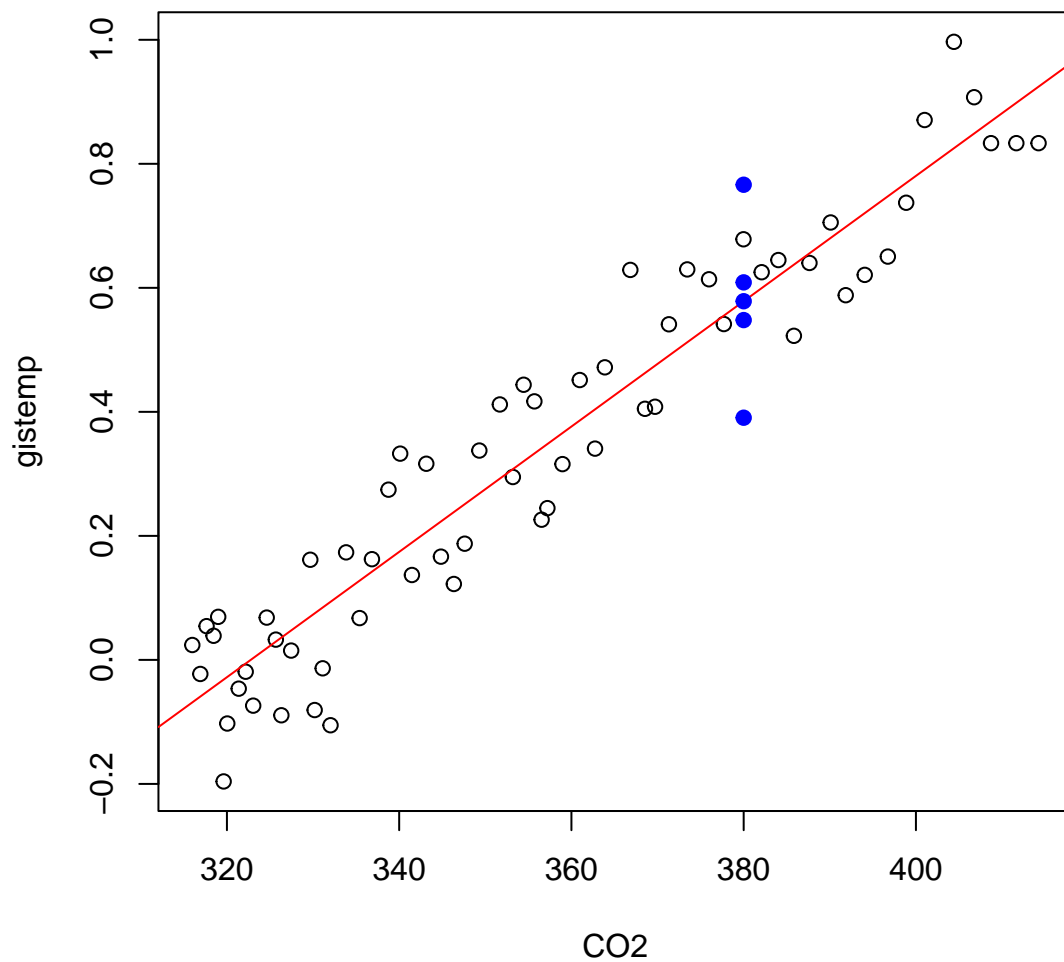
```
##          fit          lwr          upr
## 1 0.5783924 0.5479782 0.6088065
```

```

plot(gistemp~CO2, data = C02)
abline(mod1$coefficients, col = 'red')
points(nouvel_individu$CO2,prediction_IC$fit,col ="blue", pch = 19)
points(nouvel_individu$CO2,prediction_IC$lwr,col ="blue", pch = 19)
points(nouvel_individu$CO2,prediction_IC$upr,col ="blue", pch = 19)

prediction_IP = data.frame(predict(mod1, newdata = nouvel_individu, interval = 'prediction', level = 0.95))
points(nouvel_individu$CO2,prediction_IP$lwr,col ="blue", pch = 19)
points(nouvel_individu$CO2,prediction_IP$upr,col ="blue", pch = 19)

```



hhhhh