

# Stocks de Carbon

On s'intéresse au taux de mortalité en fonction de facteurs économiques et environnementaux.

On cherche à comprendre l'influence des caractéristiques sociologiques et environnementales ci-dessous sur le taux de mortalité. Pour cela on utilise la commande ci-dessous.

```
death_lm = lm(death_rate ~ Precipitation+January_temperature+July_temperature+
               percent_65_or_older+household_size+schooling_over_22+full_kitchens+
               urban_population_density+nonwhite_population+office_workers+
               poor_families+hydrocarbons+oxides_of_Nitrogen+Sulfur_Dioxide+
               humidity,data=death_data)
```

1. Ecrire le modèle correspondant aux instructions R données ci dessous. Rappelez ses hypothèses.

```
summary(death_lm)
```

```
##
## Call:
## lm(formula = death_rate ~ Precipitation + January_temperature +
##     July_temperature + percent_65_or_older + household_size +
##     schooling_over_22 + full_kitchens + urban_population_density +
##     nonwhite_population + office_workers + poor_families + hydrocarbons +
##     oxides_of_Nitrogen + Sulfur_Dioxide + humidity, data = death_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -75.285 -14.640   0.694  14.790  75.586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.863e+03  4.108e+02   4.535  4.4e-05 ***
## Precipitation    2.072e+00  8.418e-01   2.462  0.01781 *
## January_temperature -2.178e+00  6.752e-01  -3.225  0.00238 **
## July_temperature  -2.834e+00  1.771e+00  -1.600  0.11670
## percent_65_or_older -1.404e+01  7.746e+00  -1.813  0.07670 .
## household_size    -1.154e+02  6.200e+01  -1.862  0.06933 .
## schooling_over_22  -2.425e+01  1.121e+01  -2.163  0.03605 *
## full_kitchens     -1.146e+00  1.467e+00  -0.781  0.43871
## urban_population_density 1.004e-02  4.123e-03   2.435  0.01899 *
## nonwhite_population  3.533e+00  1.282e+00   2.755  0.00850 **
## office_workers     5.229e-01  1.551e+00   0.337  0.73760
## poor_families     2.671e-01  2.565e+00   0.104  0.91755
## hydrocarbons      -8.890e-01  4.524e-01  -1.965  0.05574 .
## oxides_of_Nitrogen  1.866e+00  9.345e-01   1.997  0.05201 .
## Sulfur_Dioxide    -3.447e-02  1.423e-01  -0.242  0.80968
## humidity          5.331e-01  1.052e+00   0.507  0.61474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.33 on 44 degrees of freedom
```

```
## Multiple R-squared:  0.7985, Adjusted R-squared:  0.7298
## F-statistic: 11.63 on 15 and 44 DF,  p-value: 9.56e-11
```

2. On cherche à tester le modèle. Rappelez la définition de ce test. Rappelez l'expression de la statistique de test ainsi que sa loi. Retrouvez les degrés de libertés donnés dans les sorties R précédentes.

3. On fournit aussi les sorties suivantes:

```
death_lm_0 = lm(death_rate ~ 1, data=death_data)
anova(death_lm_0, death_lm)

## Analysis of Variance Table
##
## Model 1: death_rate ~ 1
## Model 2: death_rate ~ Precipitation + January_temperature + July_temperature +
##   percent_65_or_older + household_size + schooling_over_22 +
##   full_kitchens + urban_population_density + nonwhite_population +
##   office_workers + poor_families + hydrocarbons + oxides_of_Nitrogen +
##   Sulfur_Dioxide + humidity
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      59 228308
## 2      44  46001 15   182307 11.625 9.56e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A partir des sorties précédentes, donnez la valeur des sommes des carrées *SCT*, *SCR* et *SCM*, après avoir rappelé leurs définitions mathématiques.

4. Donnez l'estimation de  $\sigma^2$  pour le modèle contenant toutes les covariables.

5. On ajoute une covariable au modèle appelée *add* et on ré-estime les paramètres du modèle.

```
summary(death_lm_add)

##
## Call:
## lm(formula = death_rate ~ add + Precipitation + January_temperature +
##   July_temperature + percent_65_or_older + household_size +
##   schooling_over_22 + full_kitchens + urban_population_density +
##   nonwhite_population + office_workers + poor_families + hydrocarbons +
##   oxides_of_Nitrogen + Sulfur_Dioxide + humidity, data = death_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -73.321 -15.734   1.795  15.682  68.796
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.755e+03  4.162e+02   4.218 0.000125 ***
## add           6.669e+00  5.159e+00   1.293 0.202991
## Precipitation  2.245e+00  8.460e-01   2.654 0.011112 *
## January_temperature -2.299e+00  6.766e-01  -3.397 0.001476 **
## July_temperature -2.565e+00  1.770e+00  -1.449 0.154568
## percent_65_or_older -1.553e+01  7.774e+00  -1.998 0.052043 .
## household_size  -1.131e+02  6.156e+01  -1.837 0.073139 .
## schooling_over_22  -2.581e+01  1.119e+01  -2.306 0.025999 *
## full_kitchens    -8.305e-01  1.476e+00  -0.563 0.576531
## urban_population_density 1.137e-02  4.218e-03   2.694 0.010019 *
```

```
## nonwhite_population      3.128e+00  1.311e+00  2.387 0.021473 *
## office_workers          6.977e-01  1.545e+00  0.451 0.653900
## poor_families           6.688e-01  2.565e+00  0.261 0.795522
## hydrocarbons            -1.022e+00  4.606e-01  -2.219 0.031826 *
## oxides_of_Nitrogen       2.176e+00  9.578e-01  2.271 0.028181 *
## Sulfur_Dioxide          -8.819e-02  1.472e-01  -0.599 0.552231
## humidity                 6.052e-01  1.045e+00  0.579 0.565598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.09 on 43 degrees of freedom
## Multiple R-squared:  0.8061, Adjusted R-squared:  0.7339
## F-statistic: 11.17 on 16 and 43 DF,  p-value: 1.574e-10
```

Comparez les sorties des deux modèles (estimation des paramètres, test de la nullité des paramètres...). Observez maintenant les  $R^2$  du modèle  $\mathcal{M}$  et de celui contenant la variable additionnelle. La variable ajoutée a en fait été simulée complètement au hasard. Commentez.

6. On revient aux tests de paramètres  $\beta_k$ . Quelles sont les variables qui vous semblent pertinentes pour expliquer le taux de mortalité? En particulier, les comparer aux corrélations représentées ci-dessous?

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.5.3
```

```
## corrplot 0.84 loaded
```

```
corrplot(cor(death_data),method = "ellipse")
```

