Régressions 2018

```
library(tidyverse)
```

DONNEES

Population totale

```
setwd("~/Documents/2A/Statapp/données")
library(readr)
nat2018us <- read_csv("W:/Documents/2A/Statapp/données/nat2018us.csv", col_ty
pes = cols_only(apgar5r = col_guess(),
                                                                dbwt = col_guess()
,
                                                                rf_inftr = col_gue
ss(),
                                                                rf_fedrg = col_gue
ss(),
                                                                rf_artec = col_gue
ss(),
                                                                mager = col_guess(
),
                                                                mrace6 = col_guess
(),
                                                                dmar = col_guess()
,
                                                                meduc = col_guess(
),
                                                                fagerec11 = col_gu
ess(),
                                                                frace6 = col_guess
(),
                                                               feduc = col_guess()
,
                                                               priorlive = col_gue
ss(),
                                                               dplural = col_guess
()))

head.matrix(nat2018us)

## # A tibble: 6 x 14
##   mager mrace6  dmar meduc fagerec11 frace6 feduc priorlive rf_inftr
```

```
##   <dbl>  <dbl> <dbl> <dbl>     <dbl>  <dbl> <dbl>     <dbl> <chr>
## 1    30     10     1     6         5      1     3         1 N
## 2    35     30     2     9         6      3     4         2 N
## 3    28     10     1     6         5      1     4         1 N
## 4    23     30     2     2         4      3     2         2 N
## 5    37     10     1     4         6      2     3         1 N
## 6    26     10     1     6         4      1     6         1 N
## # ... with 5 more variables: rf_fedrg <chr>, rf_artec <chr>,
## #   apgar5r <dbl>, dplural <dbl>, dbwt <dbl>
```

```
nat2018us <- subset(nat2018us,(nat2018us$mrace6 != 6 & nat2018us$meduc != 9 &
nat2018us$fagerec11 != 11 & nat2018us$frace6 != 9 & nat2018us$frace6 != 6 & n
at2018us$feduc != 9 & nat2018us$apgar5r != 5 & nat2018us$priorlive != 99))
```

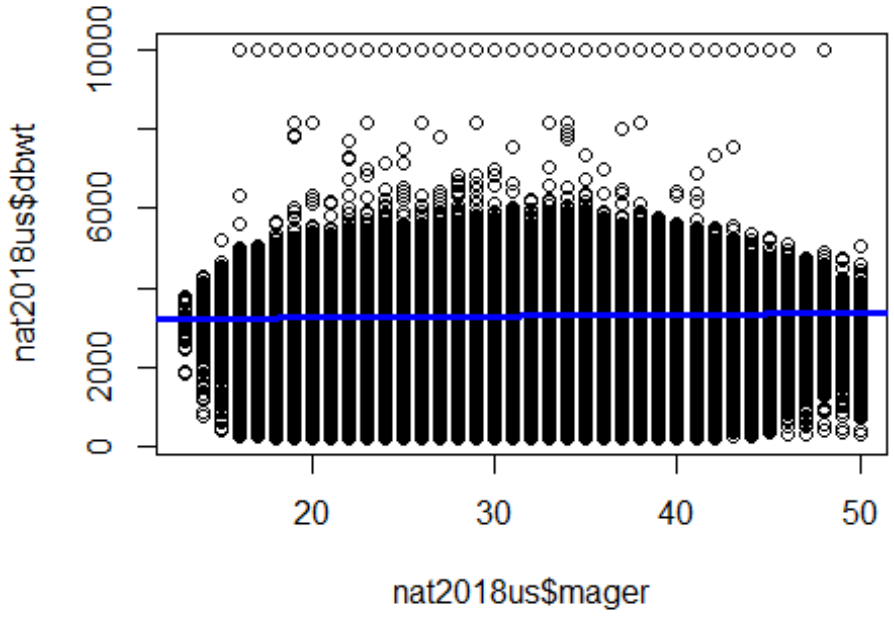
Population cible

```
pop = subset(nat2018us,rf_inftr == "Y")
```

REGRESSIONS

Test régression simple du poids à la naissance sur l'âge de la mère

```
LinReg=lm(nat2018us$dbwt~ nat2018us$mager)
plot(nat2018us$mager,nat2018us$dbwt,bg="red")
abline(LinReg,lwd=3,col="blue")
```




```
summary(LinReg)
```

```
##
## Call:
## lm(formula = nat2018us$dbwt ~ nat2018us$mager)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3102.5  -301.3    34.5   358.0  6756.7
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.186e+03  1.852e+00 1720.35   <2e-16 ***
## nat2018us$mager 3.547e+00  6.174e-02   57.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 596.2 on 2973575 degrees of freedom
## Multiple R-squared:  0.001109,   Adjusted R-squared:  0.001109
## F-statistic:  3301 on 1 and 2973575 DF,  p-value: < 2.2e-16
```

Régressions multiples

```
age2=nat2018us$mager**2
MR1=lm(dbwt~ mager + age2 + meduc + mrace6 + fagerec11 + feduc + frace6 + pri
orlive + dmar + dplural + rf_inftr + rf_artec + rf_fedrg,data=nat2018us)
summary(MR1)
```

```
##
## Call:
## lm(formula = dbwt ~ mager + age2 + meduc + mrace6 + fagerec11 +
##     feduc + frace6 + priorlive + dmar + dplural + rf_inftr +
##     rf_artec + rf_fedrg, data = nat2018us)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3244.9  -295.3    22.1   335.7  9340.9
##
## Coefficients: (3 not defined because of singularities)
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  3.970e+03  8.791e+00  451.635  < 2e-16 ***
## mager        2.564e+01  5.762e-01   44.502  < 2e-16 ***
## age2        -4.578e-01  9.377e-03  -48.822  < 2e-16 ***
## meduc        1.680e+01  3.061e-01   54.894  < 2e-16 ***
## mrace6      -1.572e+00  4.421e-02  -35.558  < 2e-16 ***
## fagerec11   -2.779e+00  3.775e-01   -7.363  1.8e-13 ***
## feduc        1.184e+01  2.873e-01   41.202  < 2e-16 ***
## frace6      -6.421e+01  5.781e-01 -111.061  < 2e-16 ***
## priorlive    3.108e+01  3.154e-01   98.532  < 2e-16 ***
## dmar        -7.756e+01  8.804e-01  -88.096  < 2e-16 ***
## dplural     -9.299e+02  1.874e+00 -496.239  < 2e-16 ***
## rf_inftrU   -5.959e+01  2.092e+01   -2.849  0.00439 **
## rf_inftrY   -8.168e+01  9.715e+00   -8.408  < 2e-16 ***
## rf_artecU    2.873e+01  1.310e+01    2.194  0.02826 *
## rf_artecX           NA         NA       NA       NA
## rf_artecY    4.624e+01  9.132e+00    5.064  4.1e-07 ***
## rf_fedrgU           NA         NA       NA       NA
## rf_fedrgX           NA         NA       NA       NA
## rf_fedrgY    2.321e+01  8.877e+00    2.615  0.00893 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 566.5 on 2610221 degrees of freedom
##   (363340 observations deleted due to missingness)
## Multiple R-squared:  0.1113, Adjusted R-squared:  0.1113
## F-statistic: 2.18e+04 on 15 and 2610221 DF,  p-value: < 2.2e-16
```

Maintenant, penchons nous sur les caractéristiques de la population ayant recours au traitement

```
age3 = pop$mager**2
MR2 = lm(dbwt~ mager + meduc + mrace6 + age3 + fagerec11 + feduc + frace6 + p
```

```
riorlive + dmar + dplural + rf_artec + rf_fedrg, data=pop)
summary(MR2)
```

```
##
## Call:
## lm(formula = dbwt ~ mager + meduc + mrace6 + age3 + fagerec11 +
##     feduc + frace6 + priorlive + dmar + dplural + rf_artec +
##     rf_fedrg, data = pop)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3143.4  -310.1    45.2   373.3  8593.0
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 3239.42635   91.22275   35.511  < 2e-16 ***
## mager         52.12933    5.18231   10.059  < 2e-16 ***
## meduc         11.23704    2.30296    4.879 1.07e-06 ***
## mrace6        -1.82768    0.32505   -5.623 1.89e-08 ***
## age3          -0.80851    0.07288  -11.093  < 2e-16 ***
## fagerec11      3.65313    2.82039    1.295 0.195237
## feduc         12.55481    2.03597    6.167 7.03e-10 ***
## frace6       -56.65300    3.97909  -14.238  < 2e-16 ***
## priorlive     56.96312    3.03730   18.755  < 2e-16 ***
## dmar         -22.16927   13.68519   -1.620 0.105249
## dplural     -867.83735    5.83578 -148.710  < 2e-16 ***
## rf_artecU     28.86176   15.10362    1.911 0.056020 .
## rf_artecY     36.51745   10.61572    3.440 0.000582 ***
## rf_fedrgU           NA         NA       NA       NA
## rf_fedrgY     26.97725   10.24052    2.634 0.008432 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 652.7 on 60063 degrees of freedom
##   (5844 observations deleted due to missingness)
## Multiple R-squared:  0.2802, Adjusted R-squared:   0.28
## F-statistic:  1798 on 13 and 60063 DF,  p-value: < 2.2e-16
```

Encore une fois, le coefficient de l'éducation du père n'est pas significatif. Priorlive a un coefficient positif plus important que précédemment

Annexe

Corélation entre le recours à médicaments et la gémélité

```
gemelite = lm(dplural~ rf_fedrg, data=nat2018us)
summary(gemelite)
```

```
##
## Call:
## lm(formula = dplural ~ rf_fedrg, data = nat2018us)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2607 -0.0300 -0.0300 -0.0300  3.9700
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.260654   0.001003 1257.23   <2e-16 ***
## rf_fedrgU   -0.108362   0.002737  -39.59   <2e-16 ***
## rf_fedrgX   -0.230683   0.001009 -228.70   <2e-16 ***
## rf_fedrgY   -0.025784   0.001508  -17.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1862 on 2973573 degrees of freedom
## Multiple R-squared:  0.02822,    Adjusted R-squared:  0.02822
## F-statistic: 2.879e+04 on 3 and 2973573 DF,  p-value: < 2.2e-16
```

```
MR3=lm(dbwt~ mager + age2 + meduc + mrace6 + fagerec11 + feduc + frace6 + pri
orlive + dmar+ rf_inftr + rf_artec + rf_fedrg,data=nat2018us)
summary(MR3)
```

```
##
## Call:
## lm(formula = dbwt ~ mager + age2 + meduc + mrace6 + fagerec11 +
##     feduc + frace6 + priorlive + dmar + rf_inftr + rf_artec +
##     rf_fedrg, data = nat2018us)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3214.6  -297.8    33.5   353.5  7209.0
##
## Coefficients: (3 not defined because of singularities)
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  3.049e+03  8.989e+00  339.211  < 2e-16 ***
## mager        2.503e+01  6.028e-01   41.521  < 2e-16 ***
## age2        -4.427e-01  9.809e-03  -45.135  < 2e-16 ***
## meduc        1.410e+01  3.202e-01   44.039  < 2e-16 ***
## mrace6      -1.535e+00  4.625e-02  -33.192  < 2e-16 ***
## fagerec11   -3.383e+00  3.949e-01   -8.567  < 2e-16 ***
## feduc        1.144e+01  3.006e-01   38.055  < 2e-16 ***
## frace6      -6.418e+01  6.048e-01 -106.111  < 2e-16 ***
## priorlive    2.096e+01  3.293e-01   63.653  < 2e-16 ***
## dmar        -8.101e+01  9.210e-01  -87.967  < 2e-16 ***
## rf_inftrU   -9.345e+01  2.188e+01   -4.271 1.95e-05 ***
## rf_inftrY   -2.719e+02  1.015e+01  -26.777  < 2e-16 ***
## rf_artecU    6.809e+01  1.370e+01    4.969 6.71e-07 ***
## rf_artecX           NA         NA       NA       NA
## rf_artecY    2.212e+01  9.552e+00    2.315  0.02060 *
## rf_fedrgU           NA         NA       NA       NA
## rf_fedrgX           NA         NA       NA       NA
## rf_fedrgY    2.580e+01  9.286e+00    2.779  0.00545 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 592.6 on 2610222 degrees of freedom
##   (363340 observations deleted due to missingness)
## Multiple R-squared:  0.0275, Adjusted R-squared:  0.0275
## F-statistic:  5273 on 14 and 2610222 DF,  p-value: < 2.2e-16
```

```
MR4 = lm(dbwt~ mager + meduc + mrace6 + age3 + fagerec11 + feduc + frace6 + p
riorlive + dmar + rf_artec + rf_fedrg, data=pop)
summary(MR4)
```

```
##
## Call:
## lm(formula = dbwt ~ mager + meduc + mrace6 + age3 + fagerec11 +
##     feduc + frace6 + priorlive + dmar + rf_artec + rf_fedrg,
##     data = pop)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2955.6  -393.5    88.1   481.0  7273.9
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1888.33340  106.17152  17.786  < 2e-16 ***
## mager         63.71680    6.06100  10.513  < 2e-16 ***
## meduc          9.93303    2.69372   3.687 0.000227 ***
## mrace6        -2.05048    0.38020  -5.393 6.95e-08 ***
## age3          -0.89154    0.08525 -10.458  < 2e-16 ***
## fagerec11     -0.18743    3.29884  -0.057 0.954691
## feduc         22.32442    2.38020   9.379  < 2e-16 ***
## frace6       -60.71055    4.65418 -13.044  < 2e-16 ***
## priorlive    -36.97629    3.47500 -10.641  < 2e-16 ***
## dmar          16.52051   16.00449   1.032 0.301963
## rf_artecU     63.09014   17.66445   3.572 0.000355 ***
## rf_artecY     -1.73700   12.41342  -0.140 0.888717
## rf_fedrgU           NA         NA      NA       NA
## rf_fedrgY     23.80367   11.97817   1.987 0.046899 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 763.4 on 60064 degrees of freedom
##   (5844 observations deleted due to missingness)
## Multiple R-squared:  0.01516,    Adjusted R-squared:  0.01497
## F-statistic: 77.06 on 12 and 60064 DF,  p-value: < 2.2e-16
```