Project Bayesian

SEFFANE Asmaa

2023-10-14

Table des matières

INTRODUCTION:

In this project, I will study a dataset that provides the number of points required to obtain a job transfer in the high schools of the Versailles Academy in 2012.

DATASET EXPLORATION:

Let's start with downloading the dataset and calling the packages I'll need all along the project:

```
library(MCMCpack)
library(BMS)
library(corrplot)
d = read.csv("mutations2.csv")
#Deplacer la column "barre" dansla premiere column
d = d[, c(6, 1:21)]
View(d)
summary(d)
```

```
##      Barre        code_etablissement    ville           etablissement
##  Min.   :  21.0   Length:516         Length:516         Length:516
##  1st Qu.: 111.0   Class :character   Class :character   Class :character
##  Median : 196.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 321.9
##  3rd Qu.: 292.0
##  Max.   :2056.0
##     commune        Matiere             Barre.1       effectif_presents_ser
ie_l
##  Min.   :78005   Length:516         Min.   :  21.0   Min.   :  6.00
##  1st Qu.:91027   Class :character   1st Qu.: 111.0   1st Qu.: 18.00
##  Median :92012   Mode  :character   Median : 196.0   Median : 30.00
##  Mean   :89739                      Mean   : 321.9   Mean   : 34.24
##  3rd Qu.:95018                      3rd Qu.: 292.0   3rd Qu.: 47.00
##  Max.   :95637                      Max.   :2056.0   Max.   :133.00
##  effectif_presents_serie_es effectif_presents_serie_s
##  Min.   : 10.00             Min.   : 13.0
##  1st Qu.: 53.00             1st Qu.: 64.0
##  Median : 69.00             Median :100.0
##  Mean   : 74.42             Mean   :106.1
##  3rd Qu.: 99.00             3rd Qu.:140.0
##  Max.   :192.00             Max.   :328.0
##  taux_brut_de_reussite_serie_l taux_brut_de_reussite_serie_es
##  Min.   : 36.00                Min.   : 51.0
##  1st Qu.: 82.00                1st Qu.: 81.0
##  Median : 89.00                Median : 88.0
##  Mean   : 86.35                Mean   : 86.4
##  3rd Qu.: 94.00                3rd Qu.: 94.0
##  Max.   :100.00                Max.   :100.0
##  taux_brut_de_reussite_serie_s taux_reussite_attendu_serie_l
```

```
##  Min.   :50.00                Min.   :65.00
##  1st Qu.:81.00                1st Qu.:84.00
##  Median :88.00                Median :89.00
##  Mean   :86.23                Mean   :86.91
##  3rd Qu.:93.00                3rd Qu.:92.00
##  Max.   :99.00                Max.   :98.00
##  taux_reussite_attendu_serie_es taux_reussite_attendu_serie_s
##  Min.   :61.00                  Min.   :61.00
##  1st Qu.:86.00                  1st Qu.:86.00
##  Median :90.00                  Median :89.00
##  Mean   :87.97                  Mean   :87.39
##  3rd Qu.:94.00                  3rd Qu.:94.00
##  Max.   :98.00                  Max.   :98.00
##  effectif_de_seconde effectif_de_premiere taux_acces_brut_seconde_bac
##  Min.   : 36.0       Min.   : 36.0        Min.   :49.00
##  1st Qu.:268.0       1st Qu.:226.5        1st Qu.:64.00
##  Median :336.0       Median :289.0        Median :71.00
##  Mean   :351.6       Mean   :307.7        Mean   :69.61
##  3rd Qu.:415.0       3rd Qu.:364.0        3rd Qu.:76.00
##  Max.   :764.0       Max.   :691.0        Max.   :87.00
##  taux_acces_attendu_seconde_bac taux_acces_brut_premiere_bac
##  Min.   :50.00                  Min.   :65.00
##  1st Qu.:64.00                  1st Qu.:82.00
##  Median :69.00                  Median :85.00
##  Mean   :68.47                  Mean   :84.53
##  3rd Qu.:73.00                  3rd Qu.:89.25
##  Max.   :83.00                  Max.   :97.00
##  taux_acces_attendu_premiere_bac
##  Min.   :70.00
##  1st Qu.:81.00
##  Median :85.00
##  Mean   :84.19
##  3rd Qu.:89.00
##  Max.   :94.00
```

BAYESIAN LINEAR REGRESSION:

The purpose is to explain the covariate of interest "Barre" by using the other covariates, for that I use Bayesian regression by choosing only the numeric column.

I will use MCMC method first, then the second method by using BMS function.

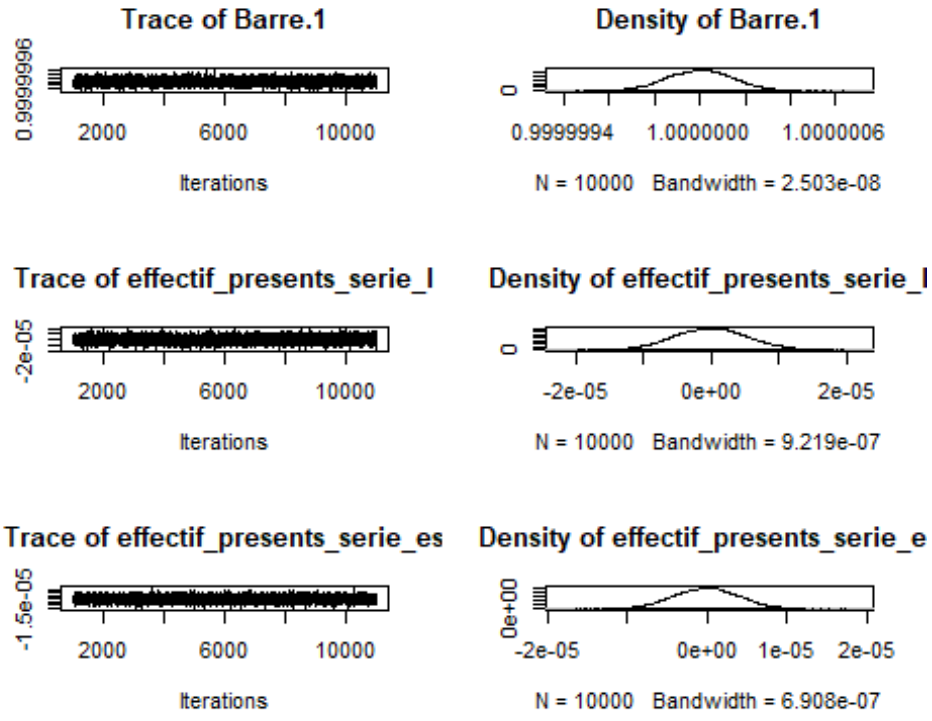
Basic function for Bayesian regression MCMC:

```
numeric_d <- d[sapply(d, is.numeric)]
reg = MCMCregress(Barre ~ ., data = numeric_d)
summary(reg)

##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                                     Mean        SD  Naive SE Time-series
SE
## (Intercept)                      5.854e-06 2.468e-03 2.468e-05      2.446e
-05
## commune                         -1.957e-10 1.240e-08 1.240e-10      0.000e
+00
## Barre.1                          1.000e+00 1.490e-07 1.490e-09      1.490e
-09
## effectif_presents_serie_l       -6.084e-08 5.487e-06 5.487e-08      5.322e
-08
## effectif_presents_serie_es      -8.660e-08 4.112e-06 4.112e-08      4.112e
-08
## effectif_presents_serie_s       -1.390e-08 3.457e-06 3.457e-08      3.375e
-08
## taux_brut_de_reussite_serie_l    1.002e-08 7.477e-06 7.477e-08      7.477e
-08
## taux_brut_de_reussite_serie_es  -7.125e-08 1.044e-05 1.044e-07      1.044e
-07
## taux_brut_de_reussite_serie_s   -4.077e-08 1.612e-05 1.612e-07      1.642e
-07
## taux_reussite_attendu_serie_l   -5.504e-08 1.859e-05 1.859e-07      1.920e
-07
## taux_reussite_attendu_serie_es   1.321e-07 2.368e-05 2.368e-07      2.368e
-07
## taux_reussite_attendu_serie_s    1.308e-08 2.451e-05 2.451e-07      2.451e
-07
```

```
## effectif_de_seconde               6.222e-09 2.010e-06 2.010e-08      2.010e
-08
## effectif_de_premiere              3.312e-08 2.337e-06 2.337e-08      2.337e
-08
## taux_acces_brut_seconde_bac      -1.120e-07 1.898e-05 1.898e-07      1.898e
-07
## taux_acces_attendu_seconde_bac    5.908e-07 3.056e-05 3.056e-07      3.056e
-07
## taux_acces_brut_premiere_bac      1.087e-07 3.304e-05 3.304e-07      3.304e
-07
## taux_acces_attendu_premiere_bac -3.800e-07 5.420e-05 5.420e-07      5.420e
-07
## sigma2                            2.017e-06 1.291e-07 1.291e-09      1.366e
-09
##
## 2. Quantiles for each variable:
##
##                                        2.5%         25%         50%         75%
## (Intercept)                      -4.781e-03 -1.661e-03 -1.475e-05 1.607e-03
## commune                          -2.440e-08 -8.624e-09 -6.221e-11 8.175e-09
## Barre.1                           1.000e+00  1.000e+00  1.000e+00 1.000e+00
## effectif_presents_serie_l        -1.072e-05 -3.823e-06 -8.985e-08 3.594e-06
## effectif_presents_serie_es       -8.184e-06 -2.901e-06 -6.735e-08 2.697e-06
## effectif_presents_serie_s        -6.729e-06 -2.322e-06 -1.060e-08 2.315e-06
## taux_brut_de_reussite_serie_l    -1.455e-05 -5.111e-06  7.916e-08 5.157e-06
## taux_brut_de_reussite_serie_es   -2.044e-05 -7.107e-06 -2.577e-08 7.040e-06
## taux_brut_de_reussite_serie_s    -3.150e-05 -1.080e-05 -7.811e-08 1.077e-05
## taux_reussite_attendu_serie_l    -3.687e-05 -1.264e-05  7.716e-08 1.240e-05
## taux_reussite_attendu_serie_es   -4.717e-05 -1.566e-05  2.827e-07 1.569e-05
## taux_reussite_attendu_serie_s    -4.842e-05 -1.604e-05  1.473e-07 1.673e-05
## effectif_de_seconde              -3.929e-06 -1.338e-06  1.042e-08 1.341e-06
## effectif_de_premiere             -4.575e-06 -1.533e-06  2.278e-08 1.613e-06
## taux_acces_brut_seconde_bac      -3.767e-05 -1.286e-05 -7.438e-08 1.275e-05
## taux_acces_attendu_seconde_bac   -5.985e-05 -2.023e-05  3.850e-07 2.120e-05
## taux_acces_brut_premiere_bac     -6.541e-05 -2.159e-05  9.316e-08 2.228e-05
## taux_acces_attendu_premiere_bac -1.085e-04 -3.555e-05 -6.105e-07 3.490e-05
## sigma2                           1.779e-06  1.928e-06  2.012e-06 2.099e-06
##                                      97.5%
## (Intercept)                      4.961e-03
## commune                          2.399e-08
## Barre.1                          1.000e+00
## effectif_presents_serie_l        1.074e-05
## effectif_presents_serie_es       7.977e-06
## effectif_presents_serie_s        6.804e-06
## taux_brut_de_reussite_serie_l    1.443e-05
## taux_brut_de_reussite_serie_es   2.031e-05
## taux_brut_de_reussite_serie_s    3.216e-05
## taux_reussite_attendu_serie_l    3.651e-05
## taux_reussite_attendu_serie_es   4.670e-05
## taux_reussite_attendu_serie_s    4.781e-05
```

```
## effectif_de_seconde              3.963e-06
## effectif_de_premiere             4.581e-06
## taux_acces_brut_seconde_bac      3.749e-05
## taux_acces_attendu_seconde_bac   5.946e-05
## taux_acces_brut_premiere_bac     6.535e-05
## taux_acces_attendu_premiere_bac 1.074e-04
```

- There is β negative which means their covariate have an opposite effect on the model like “commune, effective present serie s …”, and a positive β which means their covariate have a positive effect on the model like “Barre, taux_brut_de_reussite_serie_l, ...”.
- As we can notice, all the covariate’s interval contains 0 what make them less significant except for “Barre.1”.



- On this graph, we can see the zigzag on the trace for the three covariates easily, which means that we have a convergence.

```
raftery.diag(reg)

##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
```

```
## Probability (s) = 0.95
##
##                                  Burn-in  Total Lower bound  Dependence
##                                  (M)      (N)   (Nmin)       factor (I)
##  (Intercept)                     2        3897  3746         1.040
##  commune                         2        3867  3746         1.030
##  Barre.1                         2        3834  3746         1.020
##  effectif_presents_serie_l       2        3710  3746         0.990
##  effectif_presents_serie_es      2        3741  3746         0.999
##  effectif_presents_serie_s       2        3741  3746         0.999
##  taux_brut_de_reussite_serie_l   2        3680  3746         0.982
##  taux_brut_de_reussite_serie_es  2        3680  3746         0.982
##  taux_brut_de_reussite_serie_s   2        3741  3746         0.999
##  taux_reussite_attendu_serie_l   2        3620  3746         0.966
##  taux_reussite_attendu_serie_es  2        3929  3746         1.050
##  taux_reussite_attendu_serie_s   2        3650  3746         0.974
##  effectif_de_seconde             2        3851  3746         1.030
##  effectif_de_premiere            2        3741  3746         0.999
##  taux_acces_brut_seconde_bac     2        3994  3746         1.070
##  taux_acces_attendu_seconde_bac  2        3680  3746         0.982
##  taux_acces_brut_premiere_bac    2        3650  3746         0.974
##  taux_acces_attendu_premiere_bac 2        3802  3746         1.010
##  sigma2                          2        3680  3746         0.982
```

- In this result we look at the Dependence factor, it refers to a factor that quantifies the dependence between successive samples in the MCMC chain.

- It’s a hight which means that all covariates are correlate between them

```
effectiveSize(reg)
```

```
##                     (Intercept)                         commune
##                       10185.315                           0.000
##                         Barre.1       effectif_presents_serie_l
##                        10000.000                       10630.361
##      effectif_presents_serie_es       effectif_presents_serie_s
##                        10000.000                       10496.730
##   taux_brut_de_reussite_serie_l  taux_brut_de_reussite_serie_es
##                        10000.000                       10000.000
##   taux_brut_de_reussite_serie_s   taux_reussite_attendu_serie_l
##                         9635.508                        9374.749
##  taux_reussite_attendu_serie_es   taux_reussite_attendu_serie_s
##                        10000.000                       10000.000
##             effectif_de_seconde            effectif_de_premiere
##                        10000.000                       10000.000
##     taux_acces_brut_seconde_bac  taux_acces_attendu_seconde_bac
##                        10000.000                       10000.000
##    taux_acces_brut_premiere_bac taux_acces_attendu_premiere_bac
##                        10000.000                       10000.000
```

```
##                             sigma2
##                           8936.514
```

- The number of iterations is 11000, here we have two covariates with highest effective sample size, this makes them the best ones.
- Then we have the other covariates with effective sample size equal 10000, that make them good enough and better that the ones with an effective sample size under 10000.

CONCLUSION:

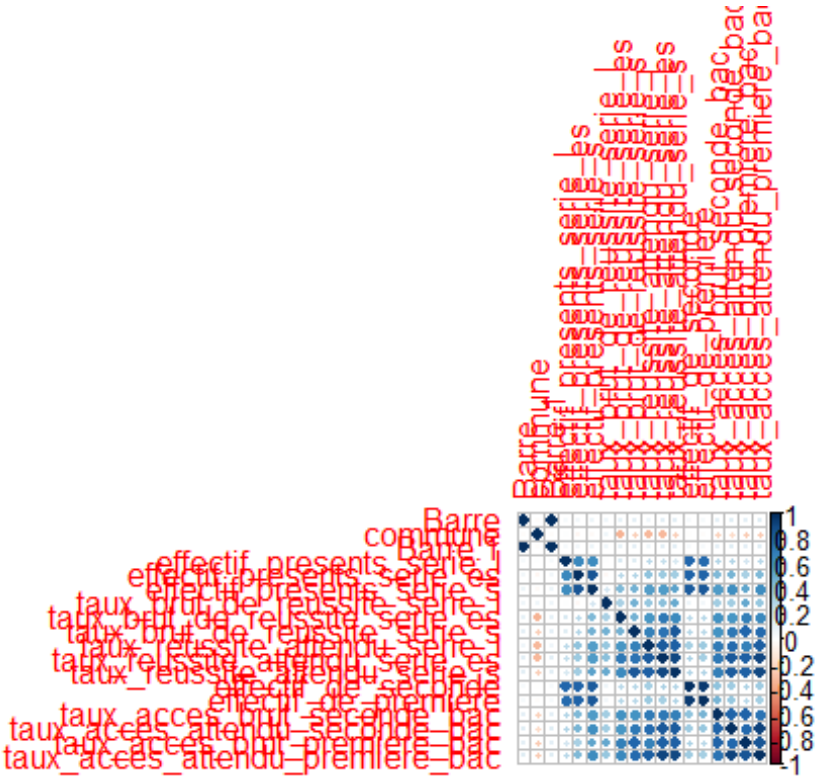
We can conclude that the best model by using MCMC method is the one with the covariates below:

"effectif_presents_serie_l + effectif_presents_serie_s + taux_acces_brut_premiere_bac + Barre.1 + taux_acces_attendu_seconde_bac"

Second method BMS:

I'll check the correlation first:

```
corrplot(cor(numeric_d))
```



Let's look on the covariates with an important correlation:

```
cor(numeric_d$Barre, numeric_d$Barre.1)
```

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

```
cor(numeric_d$effectif_presents_serie_es, numeric_d$effectif_presents_serie_s
)
```

```
## [1] 0.8578568
```

```
cor(numeric_d$taux_reussite_attendu_serie_es, numeric_d$taux_reussite_attendu
_serie_s)
```

```
## [1] 0.909138
```

```
cor(numeric_d$effectif_de_seconde, numeric_d$effectif_de_premiere)
```

```
## [1] 0.9537516
```

```
cor(numeric_d$taux_acces_attendu_seconde_bac, numeric_d$taux_acces_attendu_pr
emiere_bac)
```

```
## [1] 0.9282386
```

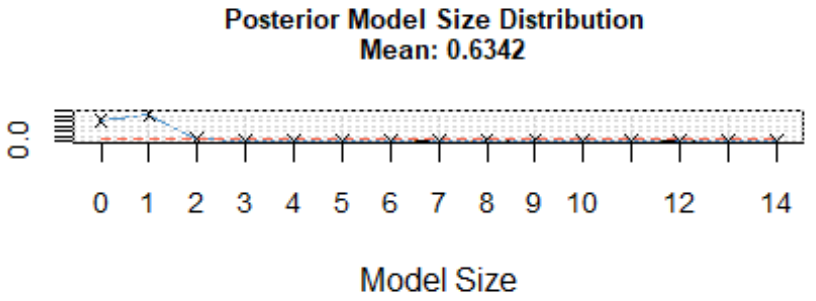
The huge correlation is obvious between those covariables. I will remove the one with high corr equal 1:

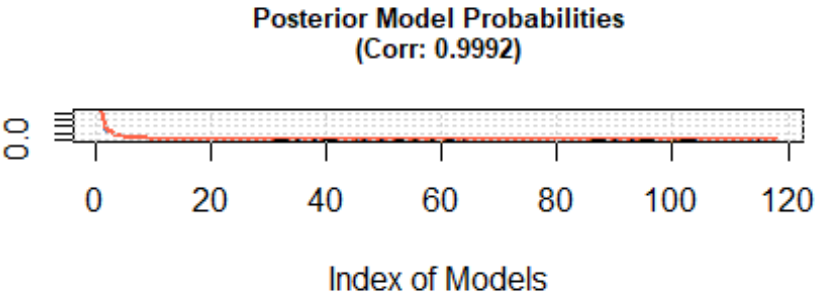
```
numeric_d1 <- subset(numeric_d, select = -c(Barre.1))
reg_2 = bms(numeric_d1, burn = 1e4, iter = 5e4)
```

```
##                                    PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.17450  1.769032e+00 4.1599412048
## taux_acces_attendu_seconde_bac  0.12946  9.999266e-01 2.7774896702
## taux_reussite_attendu_serie_s   0.06136  3.173368e-01 1.3936509367
## taux_reussite_attendu_serie_es  0.05978  3.633699e-01 1.5718834647
## taux_acces_brut_seconde_bac     0.04170  2.102891e-01 1.1201780128
## taux_brut_de_reussite_serie_s   0.03850  1.879680e-01 1.0490965994
## taux_acces_brut_premiere_bac    0.03710  2.407977e-01 1.3894774011
## taux_brut_de_reussite_serie_es  0.03324  1.441572e-01 0.8725019735
## taux_reussite_attendu_serie_l   0.01696 -5.942475e-02 0.9470483507
## effectif_presents_serie_s       0.01022  3.799607e-03 0.0588524494
## taux_brut_de_reussite_serie_l   0.00888  1.522661e-02 0.2397726036
## effectif_presents_serie_l       0.00606  2.066950e-03 0.0805422109
## effectif_de_seconde             0.00494 -1.213643e-05 0.0106528256
## commune                         0.00476  4.022003e-06 0.0002278852
## effectif_presents_serie_es      0.00448  1.854033e-03 0.0515302124
## effectif_de_premiere            0.00226  5.219593e-07 0.0081124018
##                                 Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_s       0.9778357  10
## taux_reussite_attendu_serie_es      0.9926397   9
```

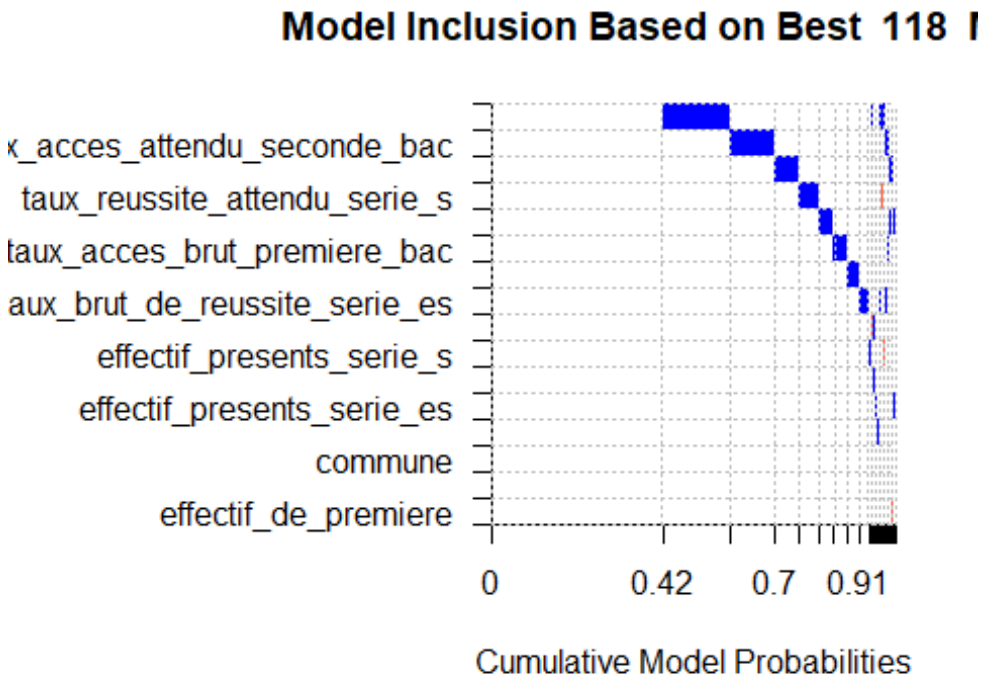
```
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_acces_brut_premiere_bac        0.9660377  15
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_reussite_attendu_serie_l       0.2959906   8
## effectif_presents_serie_s           0.7436399   4
## taux_brut_de_reussite_serie_l       1.0000000   5
## effectif_presents_serie_l           0.6336634   2
## effectif_de_seconde                 0.4777328  11
## commune                             0.7983193   1
## effectif_presents_serie_es          0.8660714   3
## effectif_de_premiere                0.2477876  12
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.6342"             "50000"             "10000"     "2.060342
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "4763"             "65536"               "7.3"
"100"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "0.9992"               "516"        "random / 8"
"UIP"
##     Shrinkage-Stats
##         "Av=0.9981"
##
## Time difference of 2.060342 secs
```

- The first think we got from this is that the PIP is low, which means that the model will contain a few covariates, for example we have taux_acces_attendu_premiere_bac with PIP = 0,017, that means that this covariate is present 17% in the model, and that is the higher PIP we've got, but at least this covariate has a positive effect on the model.





- As we can see in the graph up, the model will have 2 covariates.



- The image below shows us that the best model regarding to BMS will have a one covariate.
- when we passed the 90%, then the model contains more than one covariate, but it is not good anymore.

The following function give us the best model according to “bms”:

```
topmodels.bma(reg_2)[, 1:5]
```

```
##                                    00000     00001     00004      00080
## commune                          0.000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_l        0.000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_es       0.000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_s        0.000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_l    0.000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_es   0.000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_s    0.000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_l    0.000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_es   0.000000 0.0000000 0.0000000 1.00000000
## taux_reussite_attendu_serie_s    0.000000 0.0000000 0.0000000 0.00000000
## effectif_de_seconde              0.000000 0.0000000 0.0000000 0.00000000
## effectif_de_premiere             0.000000 0.0000000 0.0000000 0.00000000
## taux_acces_brut_seconde_bac      0.000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_seconde_bac   0.000000 0.0000000 1.0000000 0.00000000
## taux_acces_brut_premiere_bac     0.000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_premiere_bac 0.000000 1.0000000 0.0000000 0.00000000
## PMP (Exact)                      0.422035 0.1681806 0.1107765 0.05862268
## PMP (MCMC)                       0.419620 0.1525400 0.1179600 0.05430000
##                                    00040
## commune                          0.00000000
## effectif_presents_serie_l        0.00000000
## effectif_presents_serie_es       0.00000000
## effectif_presents_serie_s        0.00000000
## taux_brut_de_reussite_serie_l    0.00000000
## taux_brut_de_reussite_serie_es   0.00000000
## taux_brut_de_reussite_serie_s    0.00000000
## taux_reussite_attendu_serie_l    0.00000000
## taux_reussite_attendu_serie_es   0.00000000
## taux_reussite_attendu_serie_s    1.00000000
## effectif_de_seconde              0.00000000
## effectif_de_premiere             0.00000000
## taux_acces_brut_seconde_bac      0.00000000
## taux_acces_attendu_seconde_bac   0.00000000
## taux_acces_brut_premiere_bac     0.00000000
## taux_acces_attendu_premiere_bac 0.00000000
## PMP (Exact)                      0.04993973
## PMP (MCMC)                       0.05376000
```

CONCLUSION

All the models contain just one covariate.

STATISTIQUE FREQUENTISTE:

To have a different sight of the effects of the covariates on the covariate of interest, I'll use a different process like a linear regression:

```
reg.f1 = lm(Barre ~ ., data = numeric_d1)
summary(reg.f1)

##
## Call:
## lm(formula = Barre ~ ., data = numeric_d1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -433.3 -203.7 -125.3  -11.3 1636.8
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -5.819e+02  7.299e+02  -0.797  0.42563
## commune                        1.465e-03  3.649e-03   0.402  0.68816
## effectif_presents_serie_l      7.413e-01  1.616e+00   0.459  0.64653
## effectif_presents_serie_es     2.719e-01  1.232e+00   0.221  0.82538
## effectif_presents_serie_s     -9.359e-02  1.023e+00  -0.092  0.92712
## taux_brut_de_reussite_serie_l  2.532e+00  2.237e+00   1.132  0.25829
## taux_brut_de_reussite_serie_es 4.281e+00  3.082e+00   1.389  0.16544
## taux_brut_de_reussite_serie_s  7.744e+00  4.769e+00   1.624  0.10503
## taux_reussite_attendu_serie_l -1.456e+01  5.511e+00  -2.642  0.00849 **
## taux_reussite_attendu_serie_es 4.189e+00  7.077e+00   0.592  0.55414
## taux_reussite_attendu_serie_s -5.133e+00  7.324e+00  -0.701  0.48370
## effectif_de_seconde            1.635e-01  5.891e-01   0.278  0.78148
## effectif_de_premiere          -4.426e-01  6.970e-01  -0.635  0.52570
## taux_acces_brut_seconde_bac    1.147e+01  5.589e+00   2.052  0.04065 *
## taux_acces_attendu_seconde_bac -7.306e+00  9.144e+00  -0.799  0.42468
## taux_acces_brut_premiere_bac  -2.267e+01  9.718e+00  -2.333  0.02006 *
## taux_acces_attendu_premiere_bac 2.996e+01  1.613e+01   1.858  0.06380 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 422 on 499 degrees of freedom
## Multiple R-squared:  0.0403, Adjusted R-squared:  0.009529
## F-statistic:  1.31 on 16 and 499 DF,  p-value: 0.1858

reg2.f = step(reg.f1)
```

```
## Start:  AIC=6255.23
## Barre ~ commune + effectif_presents_serie_l + effectif_presents_serie_es +
##     effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##     taux_reussite_attendu_serie_s + effectif_de_seconde + effectif_de_prem
iere +
##     taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - effectif_presents_serie_s         1      1492 88877341 6253.2
## - effectif_presents_serie_es        1      8679 88884528 6253.3
## - effectif_de_seconde               1     13719 88889569 6253.3
## - commune                           1     28723 88904573 6253.4
## - effectif_presents_serie_l         1     37503 88913352 6253.4
## - taux_reussite_attendu_serie_es    1     62414 88938264 6253.6
## - effectif_de_premiere              1     71826 88947675 6253.6
## - taux_reussite_attendu_serie_s     1     87493 88963342 6253.7
## - taux_acces_attendu_seconde_bac    1    113699 88989549 6253.9
## - taux_brut_de_reussite_serie_l     1    228129 89103979 6254.6
## - taux_brut_de_reussite_serie_es    1    343655 89219505 6255.2
## <none>                                         88875850 6255.2
## - taux_brut_de_reussite_serie_s     1    469671 89345521 6255.9
## - taux_acces_attendu_premiere_bac   1    614643 89490492 6256.8
## - taux_acces_brut_seconde_bac       1    750278 89626128 6257.6
## - taux_acces_brut_premiere_bac      1    969180 89845030 6258.8
## - taux_reussite_attendu_serie_l     1   1243525 90119374 6260.4
##
## Step:  AIC=6253.24
## Barre ~ commune + effectif_presents_serie_l + effectif_presents_serie_es +
##     taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##     taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##     taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##     effectif_de_seconde + effectif_de_premiere + taux_acces_brut_seconde_b
ac +
##     taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - effectif_presents_serie_es        1      7251 88884592 6251.3
## - effectif_de_seconde               1     13239 88890580 6251.3
## - commune                           1     27244 88904585 6251.4
## - effectif_presents_serie_l         1     43891 88921232 6251.5
## - taux_reussite_attendu_serie_es    1     61126 88938467 6251.6
## - taux_reussite_attendu_serie_s     1     87081 88964422 6251.7
## - effectif_de_premiere              1     89288 88966629 6251.8
## - taux_acces_attendu_seconde_bac    1    112805 88990146 6251.9
## - taux_brut_de_reussite_serie_l     1    227981 89105322 6252.6
## - taux_brut_de_reussite_serie_es    1    342529 89219870 6253.2
```

```
## <none>                                             88877341 6253.2
## - taux_brut_de_reussite_serie_s    1    468181 89345522 6253.9
## - taux_acces_attendu_premiere_bac  1    656600 89533941 6255.0
## - taux_acces_brut_seconde_bac      1    766516 89643857 6255.7
## - taux_acces_brut_premiere_bac     1    970845 89848186 6256.8
## - taux_reussite_attendu_serie_l    1   1243595 90120936 6258.4
##
## Step:  AIC=6251.28
## Barre ~ commune + effectif_presents_serie_l + taux_brut_de_reussite_serie_
l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##     taux_reussite_attendu_serie_s + effectif_de_seconde + effectif_de_prem
iere +
##     taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - effectif_de_seconde              1     17235 88901827 6249.4
## - commune                          1     27350 88911942 6249.4
## - effectif_presents_serie_l        1     40860 88925452 6249.5
## - taux_reussite_attendu_serie_es   1     56321 88940913 6249.6
## - effectif_de_premiere             1     82633 88967225 6249.8
## - taux_reussite_attendu_serie_s    1     88960 88973552 6249.8
## - taux_acces_attendu_seconde_bac   1    110249 88994841 6249.9
## - taux_brut_de_reussite_serie_l    1    220911 89105503 6250.6
## - taux_brut_de_reussite_serie_es   1    339253 89223845 6251.2
## <none>                                        88884592 6251.3
## - taux_brut_de_reussite_serie_s    1    463502 89348094 6252.0
## - taux_acces_attendu_premiere_bac  1    710207 89594799 6253.4
## - taux_acces_brut_seconde_bac      1    763345 89647937 6253.7
## - taux_acces_brut_premiere_bac     1    971501 89856093 6254.9
## - taux_reussite_attendu_serie_l    1   1236859 90121451 6256.4
##
## Step:  AIC=6249.38
## Barre ~ commune + effectif_presents_serie_l + taux_brut_de_reussite_serie_
l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##     taux_reussite_attendu_serie_s + effectif_de_premiere + taux_acces_brut
_seconde_bac +
##     taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - commune                          1     27088 88928915 6247.5
## - taux_reussite_attendu_serie_es   1     50562 88952389 6247.7
## - effectif_presents_serie_l        1     57325 88959151 6247.7
## - taux_reussite_attendu_serie_s    1     92622 88994449 6247.9
## - taux_acces_attendu_seconde_bac   1    131674 89033500 6248.1
```

```
## - effectif_de_premiere             1    155346 89057173 6248.3
## - taux_brut_de_reussite_serie_l    1    218043 89119870 6248.6
## - taux_brut_de_reussite_serie_es   1    327386 89229213 6249.3
## <none>                                         88901827 6249.4
## - taux_brut_de_reussite_serie_s    1    451531 89353358 6250.0
## - taux_acces_brut_seconde_bac      1    765572 89667399 6251.8
## - taux_acces_attendu_premiere_bac  1    782867 89684694 6251.9
## - taux_acces_brut_premiere_bac     1    999124 89900951 6253.1
## - taux_reussite_attendu_serie_l    1   1234209 90136036 6254.5
##
## Step:  AIC=6247.54
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##     taux_reussite_attendu_serie_s + effectif_de_premiere + taux_acces_brut
_seconde_bac +
##     taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - taux_reussite_attendu_serie_es   1     39689 88968604 6245.8
## - effectif_presents_serie_l        1     53494 88982410 6245.8
## - taux_reussite_attendu_serie_s    1     81161 89010076 6246.0
## - taux_acces_attendu_seconde_bac   1    115207 89044122 6246.2
## - effectif_de_premiere             1    138710 89067625 6246.3
## - taux_brut_de_reussite_serie_l    1    280620 89209535 6247.2
## - taux_brut_de_reussite_serie_es   1    303405 89232320 6247.3
## <none>                                         88928915 6247.5
## - taux_brut_de_reussite_serie_s    1    425726 89354641 6248.0
## - taux_acces_brut_seconde_bac      1    743871 89672786 6249.8
## - taux_acces_attendu_premiere_bac  1    780359 89709274 6250.0
## - taux_acces_brut_premiere_bac     1    990405 89919320 6251.3
## - taux_reussite_attendu_serie_l    1   1337342 90266257 6253.2
##
## Step:  AIC=6245.77
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_s +
##     effectif_de_premiere + taux_acces_brut_seconde_bac + taux_acces_attend
u_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - taux_reussite_attendu_serie_s    1     49937 89018541 6244.1
## - effectif_presents_serie_l        1     55155 89023759 6244.1
## - taux_acces_attendu_seconde_bac   1    103426 89072030 6244.4
## - effectif_de_premiere             1    138459 89107063 6244.6
## - taux_brut_de_reussite_serie_l    1    264043 89232647 6245.3
## <none>                                         88968604 6245.8
## - taux_brut_de_reussite_serie_s    1    395947 89364551 6246.1
```

```
## - taux_brut_de_reussite_serie_es   1    401287 89369891 6246.1
## - taux_acces_brut_seconde_bac      1    721630 89690234 6247.9
## - taux_acces_attendu_premiere_bac  1    834232 89802836 6248.6
## - taux_acces_brut_premiere_bac     1    955967 89924572 6249.3
## - taux_reussite_attendu_serie_l    1   1307447 90276051 6251.3
##
## Step:  AIC=6244.06
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##     taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + effectif_de_premiere + taux_acces_brut
_seconde_bac +
##     taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - effectif_presents_serie_l        1     78545 89097086 6242.5
## - taux_acces_attendu_seconde_bac   1     87271 89105812 6242.6
## - effectif_de_premiere             1    154152 89172692 6242.9
## - taux_brut_de_reussite_serie_l    1    219425 89237966 6243.3
## <none>                                       89018541 6244.1
## - taux_brut_de_reussite_serie_s    1    354147 89372688 6244.1
## - taux_brut_de_reussite_serie_es   1    359406 89377947 6244.1
## - taux_acces_brut_seconde_bac      1    695976 89714517 6246.1
## - taux_acces_attendu_premiere_bac  1    853688 89872229 6247.0
## - taux_acces_brut_premiere_bac     1    907272 89925813 6247.3
## - taux_reussite_attendu_serie_l    1   1418883 90437424 6250.2
##
## Step:  AIC=6242.51
## Barre ~ taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##     taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##     effectif_de_premiere + taux_acces_brut_seconde_bac + taux_acces_attend
u_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - effectif_de_premiere             1     76781 89173867 6241.0
## - taux_acces_attendu_seconde_bac   1     82820 89179906 6241.0
## - taux_brut_de_reussite_serie_l    1    218004 89315090 6241.8
## - taux_brut_de_reussite_serie_s    1    294818 89391904 6242.2
## - taux_brut_de_reussite_serie_es   1    317631 89414717 6242.3
## <none>                                       89097086 6242.5
## - taux_acces_brut_seconde_bac      1    640439 89737525 6244.2
## - taux_acces_brut_premiere_bac     1    859668 89956754 6245.5
## - taux_acces_attendu_premiere_bac  1    910795 90007881 6245.8
## - taux_reussite_attendu_serie_l    1   1410562 90507648 6248.6
##
## Step:  AIC=6240.96
## Barre ~ taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##     taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##     taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
```

```
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - taux_acces_attendu_seconde_bac    1     76775 89250642 6239.4
## - taux_brut_de_reussite_serie_l     1    234055 89407922 6240.3
## - taux_brut_de_reussite_serie_s     1    259466 89433332 6240.5
## - taux_brut_de_reussite_serie_es    1    291185 89465052 6240.6
## <none>                                         89173867 6241.0
## - taux_acces_brut_seconde_bac       1    563976 89737842 6242.2
## - taux_acces_brut_premiere_bac      1    788549 89962415 6243.5
## - taux_acces_attendu_premiere_bac   1    880326 90054193 6244.0
## - taux_reussite_attendu_serie_l     1   1407485 90581352 6247.0
##
## Step:  AIC=6239.4
## Barre ~ taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##     taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##     taux_acces_brut_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - taux_brut_de_reussite_serie_l     1    221048 89471690 6238.7
## - taux_brut_de_reussite_serie_s     1    221840 89472482 6238.7
## - taux_brut_de_reussite_serie_es    1    307546 89558188 6239.2
## <none>                                         89250642 6239.4
## - taux_acces_brut_seconde_bac       1    534092 89784734 6240.5
## - taux_acces_brut_premiere_bac      1    713231 89963873 6241.5
## - taux_acces_attendu_premiere_bac   1   1306414 90557056 6244.9
## - taux_reussite_attendu_serie_l     1   1333142 90583784 6245.1
##
## Step:  AIC=6238.68
## Barre ~ taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##     taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - taux_brut_de_reussite_serie_s     1    267456 89739145 6238.2
## - taux_brut_de_reussite_serie_es    1    287265 89758955 6238.3
## <none>                                         89471690 6238.7
## - taux_acces_brut_seconde_bac       1    437670 89909360 6239.2
## - taux_acces_brut_premiere_bac      1    528829 90000519 6239.7
## - taux_acces_attendu_premiere_bac   1   1125625 90597315 6243.1
## - taux_reussite_attendu_serie_l     1   1162829 90634519 6243.3
##
## Step:  AIC=6238.22
## Barre ~ taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_l +
##     taux_acces_brut_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac
##
##                                    Df Sum of Sq      RSS    AIC
## - taux_brut_de_reussite_serie_es    1    178959 89918104 6237.2
```

```
## - taux_acces_brut_premiere_bac     1    286408 90025554 6237.9
## - taux_acces_brut_seconde_bac      1    301845 90040990 6237.9
## <none>                                        89739145 6238.2
## - taux_reussite_attendu_serie_l    1    944393 90683538 6241.6
## - taux_acces_attendu_premiere_bac  1   1274207 91013352 6243.5
##
## Step:  AIC=6237.24
## Barre ~ taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##     taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - taux_acces_brut_premiere_bac     1    184758 90102863 6236.3
## - taux_acces_brut_seconde_bac      1    318211 90236315 6237.1
## <none>                                        89918104 6237.2
## - taux_reussite_attendu_serie_l    1    840123 90758227 6240.0
## - taux_acces_attendu_premiere_bac  1   1283071 91201175 6242.6
##
## Step:  AIC=6236.3
## Barre ~ taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##     taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## - taux_acces_brut_seconde_bac      1    154234 90257096 6235.2
## <none>                                        90102863 6236.3
## - taux_reussite_attendu_serie_l    1    693957 90796819 6238.3
## - taux_acces_attendu_premiere_bac  1   1254445 91357307 6241.4
##
## Step:  AIC=6235.19
## Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_premiere_bac
##
##                                   Df Sum of Sq      RSS    AIC
## <none>                                        90257096 6235.2
## - taux_reussite_attendu_serie_l    1    574942 90832038 6236.5
## - taux_acces_attendu_premiere_bac  1   1913707 92170804 6244.0
```

```
summary(reg2.f)
```

```
##
## Call:
## lm(formula = Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_pr
emiere_bac,
##     data = numeric_d1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -387.32 -196.56 -130.83  -14.95 1696.20
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -494.324    260.593  -1.897  0.05840 .
```

```
## taux_reussite_attendu_serie_l     -7.882      4.360  -1.808  0.07124 .
## taux_acces_attendu_premiere_bac   17.833      5.407   3.298  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 419.5 on 513 degrees of freedom
## Multiple R-squared:  0.02539,    Adjusted R-squared:  0.02159
## F-statistic: 6.681 on 2 and 513 DF,  p-value: 0.001366
```

- The best model according to the AIC contain 2 covariates, that approve the previous result that we got from the "bms".
- The model is good with a highest AIC, a p-valeur equal to 0.02159 and it's a good value.
- Both covariates are significant in the model, one has a positive effect, the other has a negative effect.

PREDICTION:

I remove some lines from my dataset:

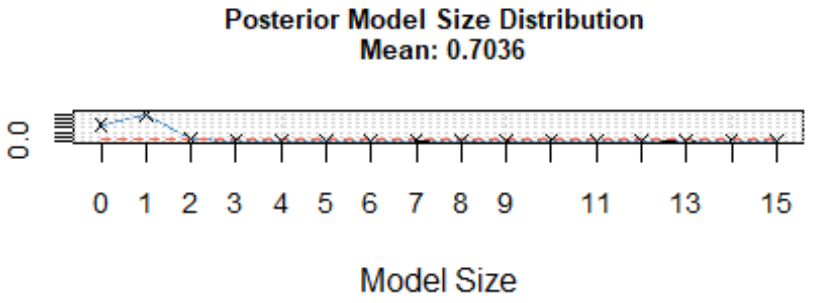
```
idx = 21:26
d2 = numeric_d1[-idx, ]
d3 = numeric_d1[idx, ]
```

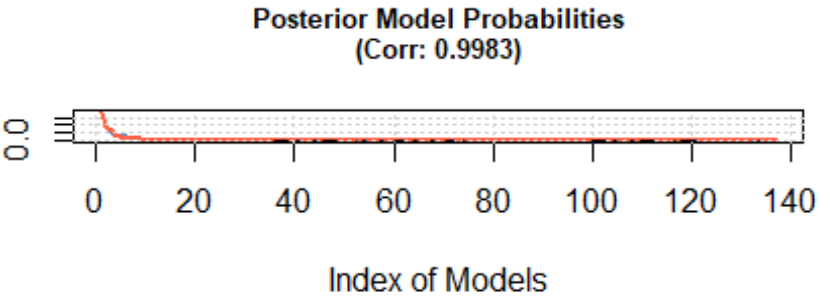
I start with the prediction for Bayesian regression using the model selected by "bms":

```
pred.f = predict(reg2.f, d3, se.fit=T)

reg2.b = bms(X.data= d2, burn=1e4, iter=5e4, nmodel=2e3)

##                                     PIP     Post Mean    Post SD Cond.Pos.
Sign
## taux_acces_attendu_premiere_bac 0.21548  2.251385e+00 4.63962269     1.000
0000
## taux_acces_attendu_seconde_bac  0.12196  9.769072e-01 2.80975881     1.000
0000
## taux_reussite_attendu_serie_es  0.08366  5.183368e-01 1.88033866     0.988
0469
## taux_reussite_attendu_serie_s   0.06406  3.386821e-01 1.45527462     0.967
8427
## taux_acces_brut_seconde_bac     0.05260  2.709801e-01 1.27095749     1.000
0000
## taux_brut_de_reussite_serie_s   0.04240  2.087074e-01 1.12008987     1.000
0000
## taux_acces_brut_premiere_bac    0.03282  2.137234e-01 1.32861228     0.957
9525
```

```
## taux_brut_de_reussite_serie_es  0.03230  1.390383e-01 0.86470432     1.000
0000
## taux_reussite_attendu_serie_l   0.01588 -6.527452e-02 0.94610243     0.249
3703
## taux_brut_de_reussite_serie_l   0.00868  1.607641e-02 0.24689241     1.000
0000
## effectif_presents_serie_s       0.00832  3.073043e-03 0.05413936     0.774
0385
## effectif_presents_serie_es      0.00754  1.504282e-03 0.06284801     0.631
2997
## commune                         0.00582  2.077893e-06 0.00025303     0.745
7045
## effectif_de_premiere            0.00474  6.656115e-06 0.01171295     0.409
2827
## effectif_de_seconde             0.00446  8.062104e-05 0.01140076     0.520
1794
## effectif_presents_serie_l       0.00284  1.196394e-03 0.05666880     0.704
2253
##                                Idx
## taux_acces_attendu_premiere_bac  16
## taux_acces_attendu_seconde_bac   14
## taux_reussite_attendu_serie_es    9
## taux_reussite_attendu_serie_s    10
## taux_acces_brut_seconde_bac      13
## taux_brut_de_reussite_serie_s     7
## taux_acces_brut_premiere_bac     15
## taux_brut_de_reussite_serie_es    6
## taux_reussite_attendu_serie_l     8
## taux_brut_de_reussite_serie_l     5
## effectif_presents_serie_s         4
## effectif_presents_serie_es        3
## commune                           1
## effectif_de_premiere             12
## effectif_de_seconde              11
## effectif_presents_serie_l         2
##
## Mean no. regressors               Draws             Burnins
Time
##           "0.7036"             "50000"             "10000"     "2.066983
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "5260"             "65536"                 "8"
"100"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "0.9983"               "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
```

```
##
## Time difference of 2.066983 secs
```





I choose one model between the best five ones:

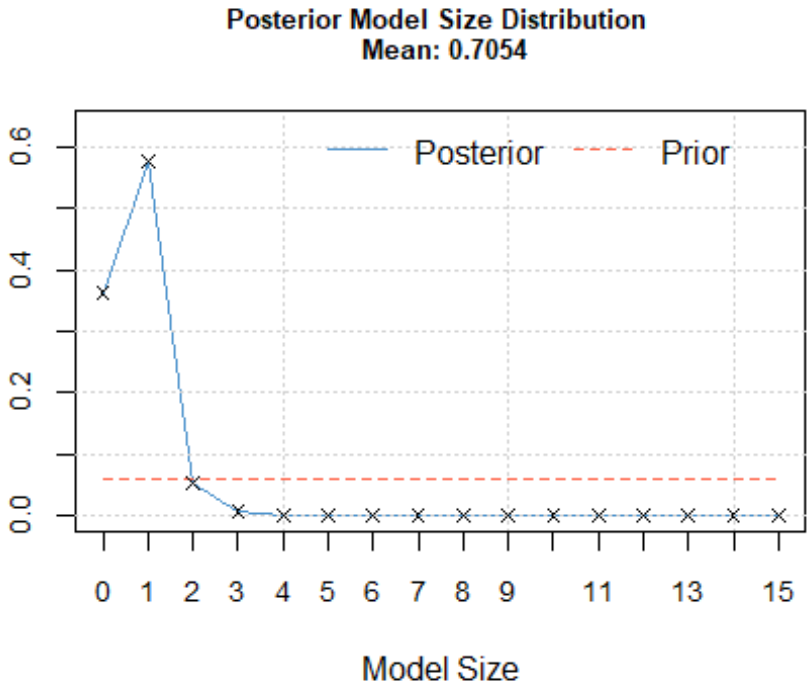
```
topmodels.bma(reg2.b)[, 1]
```

```
##                          commune       effectif_presents_serie_l
##                        0.0000000                       0.0000000
##      effectif_presents_serie_es       effectif_presents_serie_s
##                        0.0000000                       0.0000000
##   taux_brut_de_reussite_serie_l  taux_brut_de_reussite_serie_es
##                        0.0000000                       0.0000000
##   taux_brut_de_reussite_serie_s   taux_reussite_attendu_serie_l
##                        0.0000000                       0.0000000
##  taux_reussite_attendu_serie_es   taux_reussite_attendu_serie_s
##                        0.0000000                       0.0000000
##             effectif_de_seconde            effectif_de_premiere
##                        0.0000000                       0.0000000
##     taux_acces_brut_seconde_bac  taux_acces_attendu_seconde_bac
##                        0.0000000                       0.0000000
##    taux_acces_brut_premiere_bac taux_acces_attendu_premiere_bac
##                        0.0000000                       0.0000000
##                      PMP (Exact)                     PMP (MCMC)
##                        0.3725923                       0.3571600
```

```
reg2.b1 = bms(X.data= d2, burn=1e4, iter=5e4, nmodel = 2)
```

```
##                                      PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.20990  2.189029e+00 4.5802698025
## taux_acces_attendu_seconde_bac  0.14160  1.143730e+00 3.0142737328
## taux_reussite_attendu_serie_s   0.07022  3.577934e-01 1.4985841484
## taux_reussite_attendu_serie_es  0.06516  4.025166e-01 1.6888983926
## taux_acces_brut_seconde_bac     0.04850  2.465020e-01 1.2137534331
## taux_acces_brut_premiere_bac    0.04036  2.590207e-01 1.4701000244
## taux_brut_de_reussite_serie_s   0.03236  1.602162e-01 0.9834864193
## taux_brut_de_reussite_serie_es  0.03062  1.319401e-01 0.8429046815
## taux_reussite_attendu_serie_l   0.01660 -6.499292e-02 0.9438623309
## taux_brut_de_reussite_serie_l   0.01120  1.828524e-02 0.2645344755
## effectif_presents_serie_s       0.00986  4.302011e-03 0.0627890180
## effectif_presents_serie_es      0.00926  2.008808e-03 0.0725057694
## effectif_de_seconde             0.00678  4.974369e-05 0.0137265086
## effectif_presents_serie_l       0.00538  2.075999e-03 0.0790683916
## commune                         0.00462  1.464535e-06 0.0002258049
## effectif_de_premiere            0.00296 -1.140080e-05 0.0103027413
##                                 Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_s       0.9649673  10
## taux_reussite_attendu_serie_es      0.9782075   9
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_acces_brut_premiere_bac        0.9479683  15
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_reussite_attendu_serie_l       0.2493976   8
## taux_brut_de_reussite_serie_l       1.0000000   5
## effectif_presents_serie_s           0.8539554   4
## effectif_presents_serie_es          0.5831534   3
## effectif_de_seconde                 0.4719764  11
## effectif_presents_serie_l           0.8475836   2
## commune                             0.7445887   1
## effectif_de_premiere                0.2972973  12
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.7054"             "50000"             "10000"     "1.797352
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "5236"             "65536"                 "8"
"55"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "1.0000"              "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
```

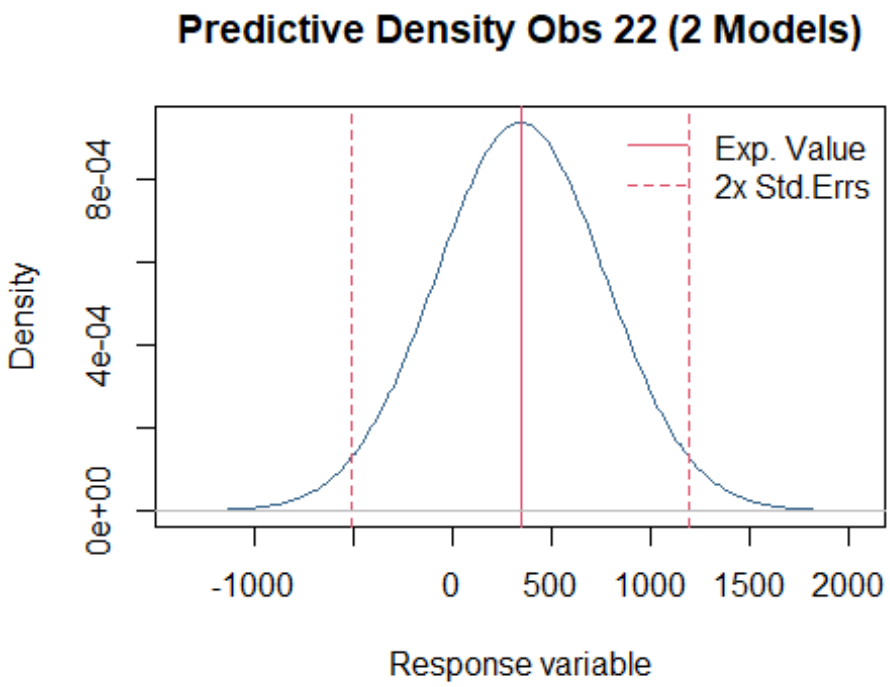
```
##
## Time difference of 1.797352 secs
```



Then I do the predictions for this model:

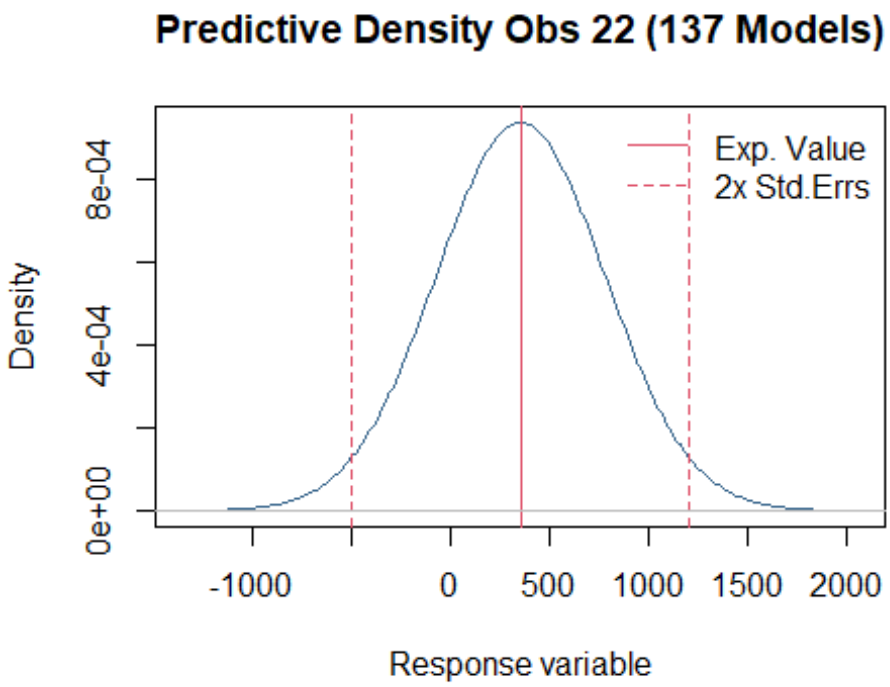
```
pdens1 = pred.density(reg2.b1, d3)
pdens1
```

```
## Call:
## pred.density(reg2.b1, d3)
##
## Densities for conditional forecast(s)
## 300 data points, based on 2 models;
##    Exp.Val. Std.Err.
## 21 347.2493 426.2786
## 22 347.2493 426.2786
## 23 343.7713 426.2284
## 24 343.7713 426.2284
## 25 343.7713 426.2284
## 26 343.7713 426.2284
```



The graph is perfect; it's concentrated in approximately 300, which is around the mean of "Barre": 321.9.

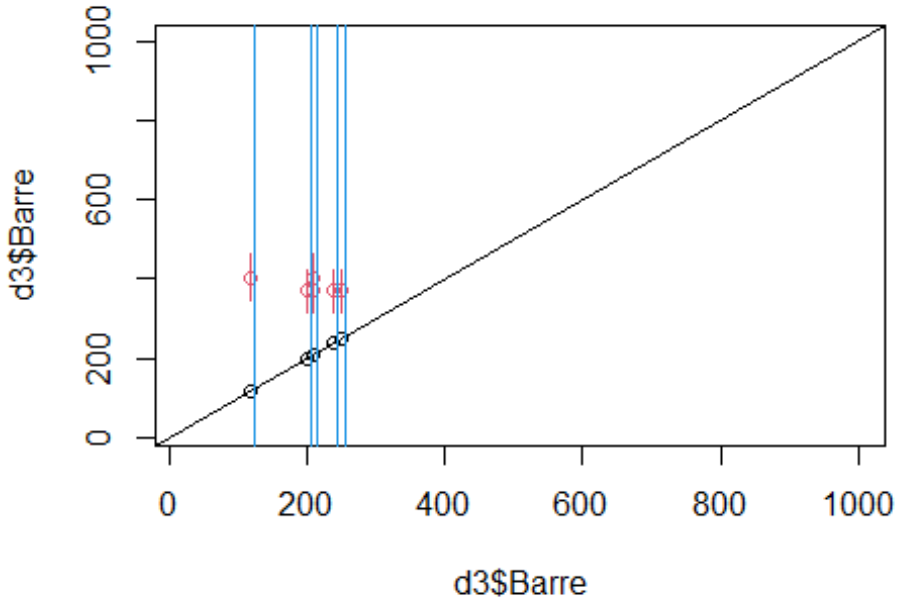
```
pdens.all = pred.density(reg2.b, d3)
pdens.all

## Call:
## pred.density(reg2.b, d3)
##
## Densities for conditional forecast(s)
## 300 data points, based on 137 models;
##    Exp.Val. Std.Err.
## 21 355.9830 425.5251
## 22 355.9830 425.5251
## 23 354.6769 425.5052
## 24 354.6769 425.5052
## 25 354.6769 425.5052
## 26 354.6769 425.5052
```



Also, the graph is perfect and concentrated around the mean of "Barre".

Now, let's assemble all the prediction we did calculate and compare them with the real value:

```
plot(d3$Barre, d3$Barre,col=1, xlim=c(20, 1000), ylim=c(20, 1000))
abline(0, 1)
points(d3$Barre, pred.f$fit, col=2)
for(i in 1:6){
  lines(c(d3$Barre[i], d3$Barre[i]), pred.f$fit[i]+c(-2,2)*pred.f$se.fit[i],
col=2)
}
for(i in 1:6){
  lines(c(d3$Barre[i], d3$Barre[i])+6, quantile(pdens.all, c(.025, .975))[i,
], col=4)
}
```



- In red, we have the prediction with linear regression model, as we can see it is far from the real value and don't reach the values.
- In blue, we have the prediction with the Bayesian progression model, reach the real values what makes our model good one.

APPLICATION:

Let's extract the data only where matière is "mathematique"; then "anglais" and see if the covariates act in the same way for both "mathematique" and "anglaise":

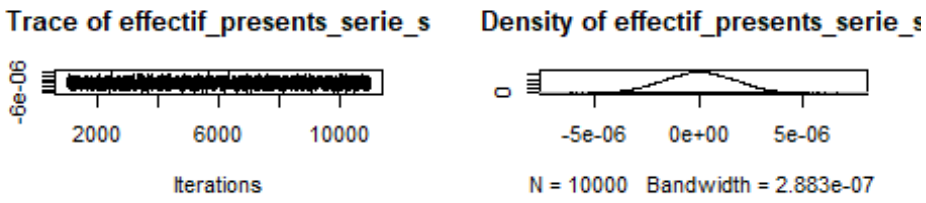
```
d_Ma <- d[d$Matiere == "MATHS" , ]
d_AN <- d[d$Matiere == "ANGLAIS" , ]
```

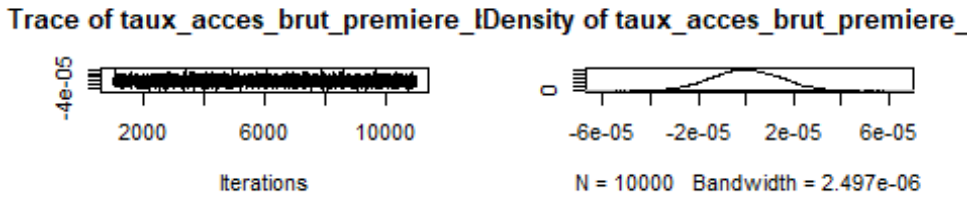
MCMC method:

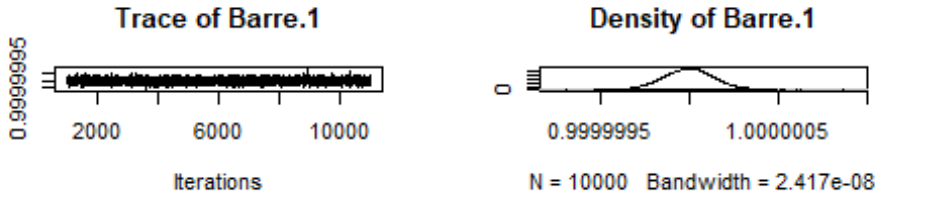
Now, I extract just the numeric covariates then do the Bayesian regression using the previous chosen model: "mathematique"

```
numerci_MA <- d[sapply(d_Ma, is.numeric)]
reg_MA = MCMCregress(Barre ~ effectif_presents_serie_l + effectif_presents_se
rie_s + taux_acces_brut_premiere_bac + Barre.1 + taux_acces_attendu_seconde_b
ac, data=numerci_MA)
summary(reg_MA)

##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                                     Mean        SD  Naive SE Time-series
SE
## (Intercept)                     -1.692e-06 8.765e-04 8.765e-06      8.765e-
06
## effectif_presents_serie_l       -2.545e-08 3.875e-06 3.875e-08      3.875e-
08
## effectif_presents_serie_s        7.993e-09 1.716e-06 1.716e-08      1.716e-
08
## taux_acces_brut_premiere_bac     2.760e-08 1.487e-05 1.487e-07      1.487e-
07
## Barre.1                          1.000e+00 1.466e-07 1.466e-09      1.466e-
09
## taux_acces_attendu_seconde_bac -1.004e-08 1.528e-05 1.528e-07      1.528e-
07
## sigma2                           1.969e-06 1.235e-07 1.235e-09      1.235e-
09
##
## 2. Quantiles for each variable:
##
##                                      2.5%        25%        50%        75%
## (Intercept)                    -1.709e-03 -6.043e-04  5.348e-06 5.958e-04
## effectif_presents_serie_l      -7.739e-06 -2.654e-06 -4.080e-08 2.542e-06
```

```
## effectif_presents_serie_s       -3.366e-06 -1.147e-06  5.547e-09 1.187e-06
## taux_acces_brut_premiere_bac    -2.919e-05 -9.962e-06 -1.261e-07 1.013e-05
## Barre.1                          1.000e+00  1.000e+00  1.000e+00 1.000e+00
## taux_acces_attendu_seconde_bac -2.980e-05 -1.037e-05  1.290e-07 1.021e-05
## sigma2                           1.741e-06  1.883e-06  1.963e-06 2.049e-06
##                                     97.5%
## (Intercept)                     1.701e-03
## effectif_presents_serie_l       7.607e-06
## effectif_presents_serie_s       3.329e-06
## taux_acces_brut_premiere_bac    2.942e-05
## Barre.1                         1.000e+00
## taux_acces_attendu_seconde_bac 2.963e-05
## sigma2                          2.227e-06
```








```
raftery.diag(reg_MA)
```

```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
##                                 Burn-in  Total Lower bound  Dependence
##                                 (M)      (N)   (Nmin)       factor (I)
##  (Intercept)                    2        3834  3746         1.020
##  effectif_presents_serie_l      2        3680  3746         0.982
##  effectif_presents_serie_s      2        3590  3746         0.958
##  taux_acces_brut_premiere_bac   2        3741  3746         0.999
```

```
##  Barre.1                          2        3710  3746         0.990
##  taux_acces_attendu_seconde_bac 2        3771  3746         1.010
##  sigma2                           2        3710  3746         0.990
```

```
effectiveSize(reg_MA)
```

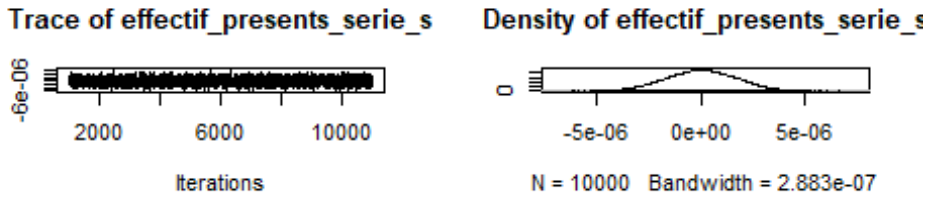
```
##                     (Intercept)      effectif_presents_serie_l
##                           10000                          10000
##      effectif_presents_serie_s   taux_acces_brut_premiere_bac
##                           10000                          10000
##                         Barre.1 taux_acces_attendu_seconde_bac
##                           10000                          10000
##                          sigma2
##                           10000
```

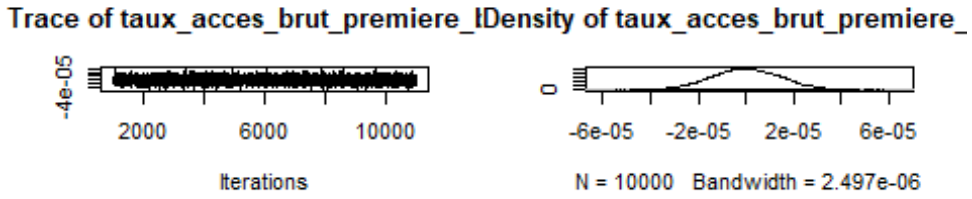
Now it's the "anglais" turn, let's do the Bayesian regression and see if it is the same:

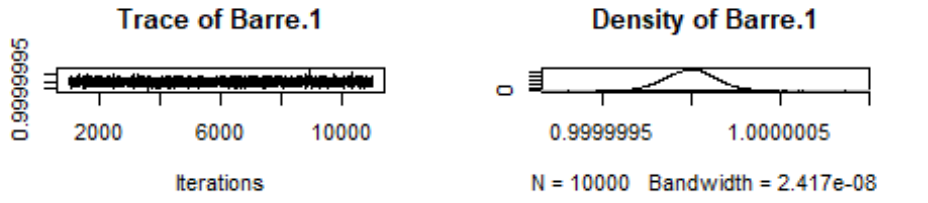
```
numerci_AN <- d[sapply(d_AN, is.numeric)]
reg_AN = MCMCregress(Barre ~ effectif_presents_serie_l + effectif_presents_se
rie_s + taux_acces_brut_premiere_bac + Barre.1 + taux_acces_attendu_seconde_b
ac, data=numerci_AN)
summary(reg_AN)
```

```
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                                      Mean        SD  Naive SE Time-series
SE
## (Intercept)                    -1.692e-06 8.765e-04 8.765e-06      8.765e-
06
## effectif_presents_serie_l      -2.545e-08 3.875e-06 3.875e-08      3.875e-
08
## effectif_presents_serie_s       7.993e-09 1.716e-06 1.716e-08      1.716e-
08
## taux_acces_brut_premiere_bac    2.760e-08 1.487e-05 1.487e-07      1.487e-
07
## Barre.1                         1.000e+00 1.466e-07 1.466e-09      1.466e-
09
## taux_acces_attendu_seconde_bac -1.004e-08 1.528e-05 1.528e-07      1.528e-
07
## sigma2                          1.969e-06 1.235e-07 1.235e-09      1.235e-
```

```
09
##
## 2. Quantiles for each variable:
##
##                                    2.5%        25%         50%       75%
## (Intercept)                   -1.709e-03 -6.043e-04  5.348e-06 5.958e-04
## effectif_presents_serie_l     -7.739e-06 -2.654e-06 -4.080e-08 2.542e-06
## effectif_presents_serie_s     -3.366e-06 -1.147e-06  5.547e-09 1.187e-06
## taux_acces_brut_premiere_bac  -2.919e-05 -9.962e-06 -1.261e-07 1.013e-05
## Barre.1                        1.000e+00  1.000e+00  1.000e+00 1.000e+00
## taux_acces_attendu_seconde_bac -2.980e-05 -1.037e-05  1.290e-07 1.021e-05
## sigma2                         1.741e-06  1.883e-06  1.963e-06 2.049e-06
##                                   97.5%
## (Intercept)                   1.701e-03
## effectif_presents_serie_l     7.607e-06
## effectif_presents_serie_s     3.329e-06
## taux_acces_brut_premiere_bac  2.942e-05
## Barre.1                       1.000e+00
## taux_acces_attendu_seconde_bac 2.963e-05
## sigma2                        2.227e-06
```








```
raftery.diag(reg_AN)

##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
```

```
##
##                                Burn-in  Total Lower bound  Dependence
##                                (M)      (N)   (Nmin)       factor (I)
##  (Intercept)                   2        3834  3746         1.020
##  effectif_presents_serie_l     2        3680  3746         0.982
##  effectif_presents_serie_s     2        3590  3746         0.958
##  taux_acces_brut_premiere_bac  2        3741  3746         0.999
##  Barre.1                       2        3710  3746         0.990
##  taux_acces_attendu_seconde_bac 2       3771  3746         1.010
##  sigma2                        2        3710  3746         0.990
```

```
effectiveSize(reg_AN)
```

```
##                     (Intercept)      effectif_presents_serie_l
##                           10000                          10000
##      effectif_presents_serie_s   taux_acces_brut_premiere_bac
##                           10000                          10000
##                         Barre.1 taux_acces_attendu_seconde_bac
##                           10000                          10000
##                          sigma2
##                           10000
```

We notice that we got the same result even if I use same model on different dataset, we got the same β, same intervals of β and same dependence factor.

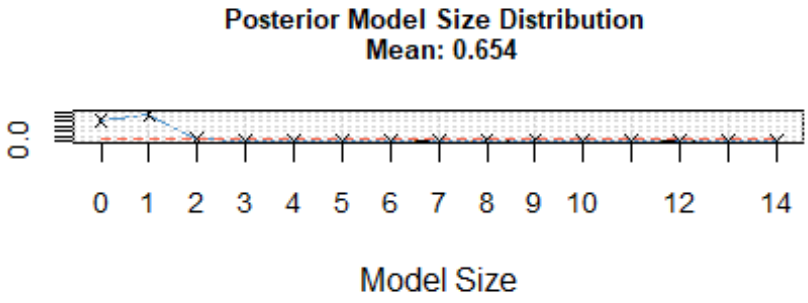
That give as the conclusion that the covariates act the same way on both different dataset.

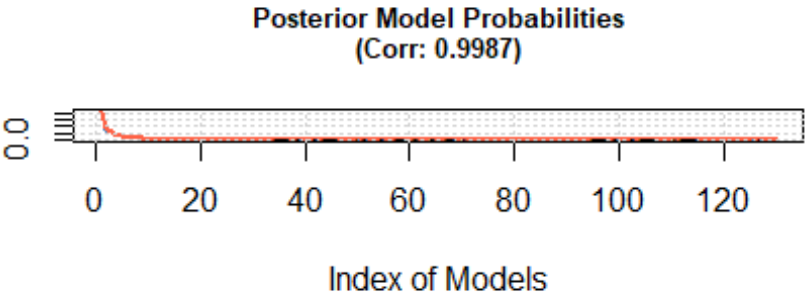
BMS:

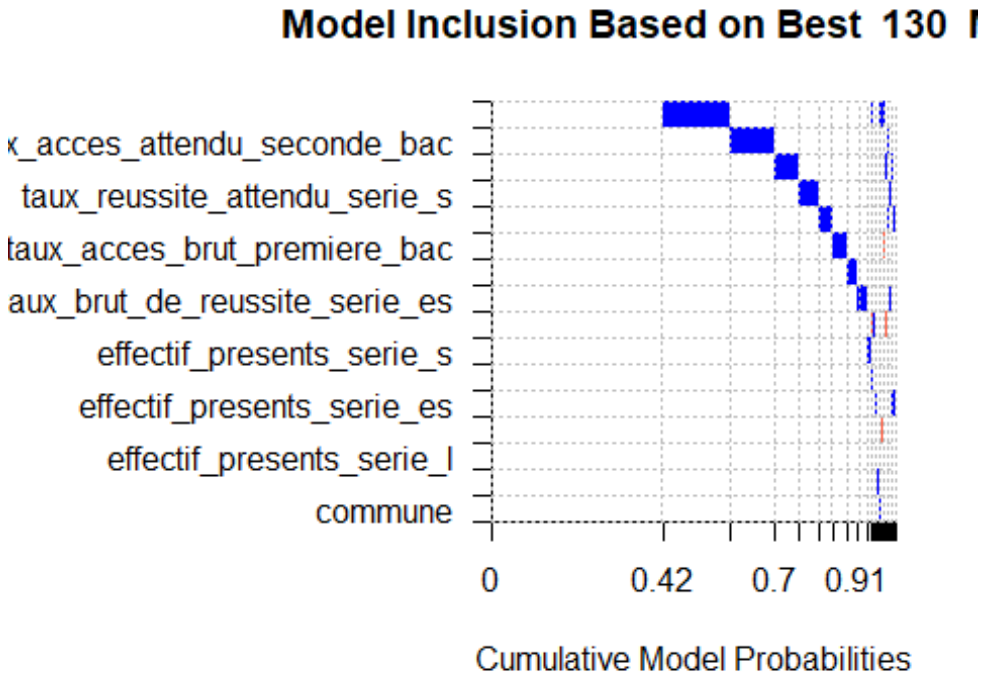
```
numeriC_MA <- d[sapply(d_Ma, is.numeric)]
numeric_MA1 <- subset(numeriC_MA, select = -c(Barre.1))
regMA_1 = bms(numeric_MA1, burn = 1e4, iter = 5e4)
```

```
##                                    PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.17748  1.781566e+00 4.1360904812
## taux_acces_attendu_seconde_bac  0.12948  1.001788e+00 2.7873756353
## taux_reussite_attendu_serie_es  0.07768  4.737795e-01 1.7699304249
## taux_reussite_attendu_serie_s   0.06272  3.299846e-01 1.4318950804
## taux_brut_de_reussite_serie_s   0.04282  2.056797e-01 1.0959286784
## taux_acces_brut_seconde_bac     0.04242  2.136991e-01 1.1318189222
## taux_brut_de_reussite_serie_es  0.03268  1.420783e-01 0.8650132492
## taux_acces_brut_premiere_bac    0.02730  1.681231e-01 1.1761724891
## taux_reussite_attendu_serie_l   0.01770 -5.658843e-02 0.9208280228
## effectif_presents_serie_s       0.00944  4.096329e-03 0.0646272248
## effectif_de_seconde             0.00820 -8.337727e-05 0.0175204159
## effectif_presents_serie_es      0.00800  2.170699e-03 0.0628616247
## taux_brut_de_reussite_serie_l   0.00692  1.111740e-02 0.2069508924
## effectif_presents_serie_l       0.00414  1.376095e-03 0.0657934404
## effectif_de_premiere            0.00388  1.039846e-04 0.0111618508
```

```
## commune                          0.00312  3.463901e-06 0.0001891729
##                                  Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_es      0.9974253   9
## taux_reussite_attendu_serie_s       0.9799107  10
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_acces_brut_premiere_bac        0.9597070  15
## taux_reussite_attendu_serie_l       0.2858757   8
## effectif_presents_serie_s           0.8029661   4
## effectif_de_seconde                 0.4731707  11
## effectif_presents_serie_es          0.8025000   3
## taux_brut_de_reussite_serie_l       1.0000000   5
## effectif_presents_serie_l           0.6763285   2
## effectif_de_premiere                0.4690722  12
## commune                             0.8717949   1
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.6540"             "50000"             "10000"     "7.019122
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "4706"             "65536"               "7.2"
"100"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "0.9987"               "516"        "random / 8"
"UIP"
##     Shrinkage-Stats
##         "Av=0.9981"
##
## Time difference of 7.019122 secs
```








```
topmodels.bma(regMA_1)[, 1:5]
```

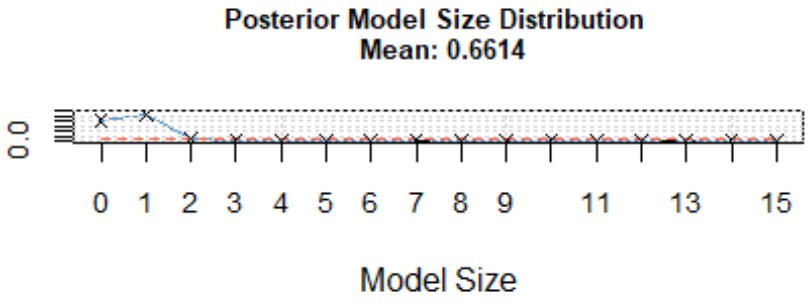
```
##                                    00000     00001     00004      00080
## commune                          0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_l        0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_es       0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_s        0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_l    0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_es   0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_s    0.0000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_l    0.0000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_es   0.0000000 0.0000000 0.0000000 1.00000000
## taux_reussite_attendu_serie_s    0.0000000 0.0000000 0.0000000 0.00000000
## effectif_de_seconde              0.0000000 0.0000000 0.0000000 0.00000000
## effectif_de_premiere             0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_brut_seconde_bac      0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_seconde_bac   0.0000000 0.0000000 1.0000000 0.00000000
## taux_acces_brut_premiere_bac     0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_premiere_bac 0.0000000 1.0000000 0.0000000 0.00000000
## PMP (Exact)                      0.4216642 0.1680328 0.1106792 0.05857117
## PMP (MCMC)                       0.4054400 0.1572000 0.1180000 0.07034000
##                                    00040
## commune                          0.00000000
## effectif_presents_serie_l        0.00000000
## effectif_presents_serie_es       0.00000000
## effectif_presents_serie_s        0.00000000
## taux_brut_de_reussite_serie_l    0.00000000
## taux_brut_de_reussite_serie_es   0.00000000
## taux_brut_de_reussite_serie_s    0.00000000
## taux_reussite_attendu_serie_l    0.00000000
## taux_reussite_attendu_serie_es   0.00000000
## taux_reussite_attendu_serie_s    1.00000000
## effectif_de_seconde              0.00000000
## effectif_de_premiere             0.00000000
## taux_acces_brut_seconde_bac      0.00000000
## taux_acces_attendu_seconde_bac   0.00000000
## taux_acces_brut_premiere_bac     0.00000000
## taux_acces_attendu_premiere_bac 0.00000000
## PMP (Exact)                      0.04989585
## PMP (MCMC)                       0.05418000
```

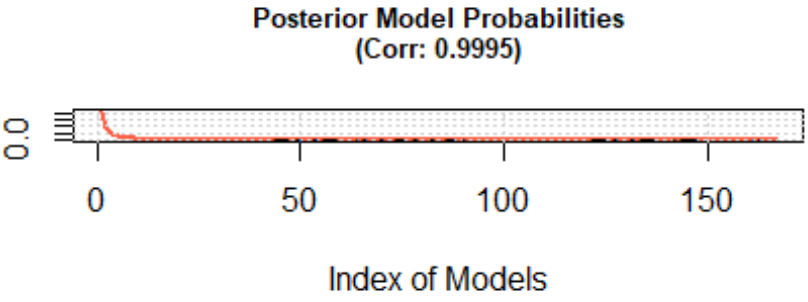
Now we apply the same model on the dataset of “anglaise”:

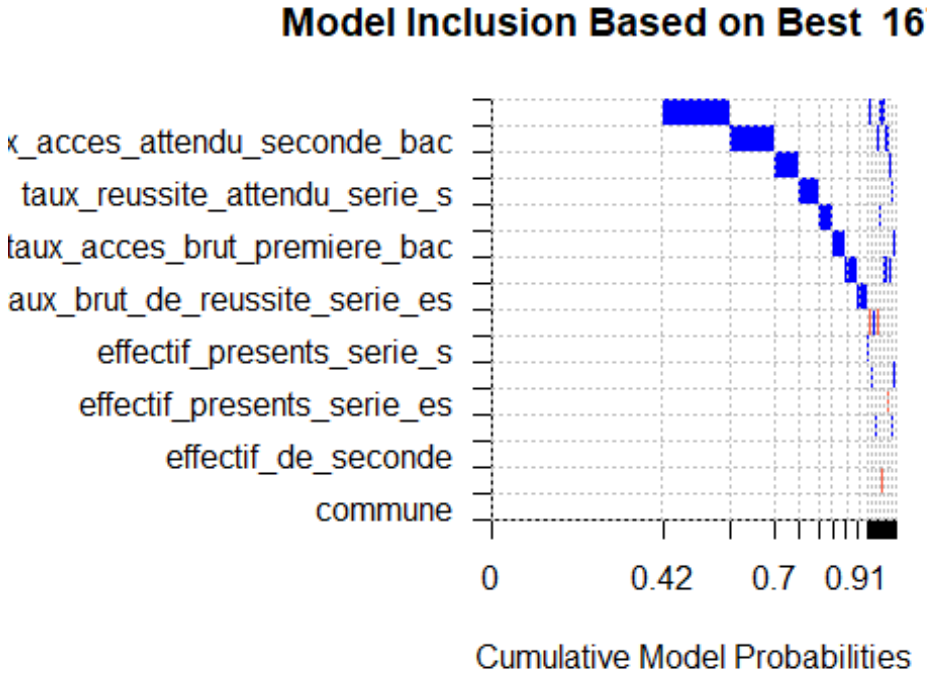
```
numeriC_AN <- d[sapply(d_AN, is.numeric)]
numeric_AN1 <- subset(numeriC_AN, select = -c(Barre.1))
regAN_1 = bms(numeric_AN1, burn = 1e4, iter = 5e4)
```

```
##                                    PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.18864  1.904126e+00 4.2801945741
## taux_acces_attendu_seconde_bac  0.12544  9.671609e-01 2.7459459373
## taux_reussite_attendu_serie_s   0.06800  3.475527e-01 1.4638885647
```

```
## taux_reussite_attendu_serie_es  0.06056  3.591252e-01 1.5581401131
## taux_acces_brut_seconde_bac     0.04196  2.051290e-01 1.1077243859
## taux_acces_brut_premiere_bac    0.04126  2.717579e-01 1.4671756415
## taux_brut_de_reussite_serie_es  0.03594  1.532911e-01 0.8956419895
## taux_brut_de_reussite_serie_s   0.03270  1.486387e-01 0.9385512373
## taux_reussite_attendu_serie_l   0.01576 -5.611362e-02 0.9097073956
## effectif_presents_serie_s       0.01032  4.002359e-03 0.0633844210
## effectif_presents_serie_l       0.01026  3.449042e-03 0.1051625006
## taux_brut_de_reussite_serie_l   0.00902  1.584715e-02 0.2440136435
## commune                         0.00688  5.561854e-06 0.0002758105
## effectif_de_premiere            0.00544 -5.884449e-05 0.0138192306
## effectif_de_seconde             0.00492  1.701388e-06 0.0131884618
## effectif_presents_serie_es      0.00432  1.061693e-03 0.0472402437
##                                Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_s       0.9644118  10
## taux_reussite_attendu_serie_es      0.9778732   9
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_acces_brut_premiere_bac        0.9815802  15
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_reussite_attendu_serie_l       0.3032995   8
## effectif_presents_serie_s           0.7771318   4
## effectif_presents_serie_l           0.7115010   2
## taux_brut_de_reussite_serie_l       1.0000000   5
## commune                             0.7587209   1
## effectif_de_premiere                0.2720588  12
## effectif_de_seconde                 0.4471545  11
## effectif_presents_serie_es          0.6111111   3
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.6614"             "50000"             "10000"     "4.196476
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "4849"             "65536"               "7.4"
"100"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "0.9995"               "516"        "random / 8"
"UIP"
##     Shrinkage-Stats
##         "Av=0.9981"
##
## Time difference of 4.196476 secs
```








```
topmodels.bma(regAN_1)[, 1:5]
```

```
##                                    00000     00001     00004      00080
## commune                          0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_l        0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_es       0.0000000 0.0000000 0.0000000 0.00000000
## effectif_presents_serie_s        0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_l    0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_es   0.0000000 0.0000000 0.0000000 0.00000000
## taux_brut_de_reussite_serie_s    0.0000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_l    0.0000000 0.0000000 0.0000000 0.00000000
## taux_reussite_attendu_serie_es   0.0000000 0.0000000 0.0000000 1.00000000
## taux_reussite_attendu_serie_s    0.0000000 0.0000000 0.0000000 0.00000000
## effectif_de_seconde              0.0000000 0.0000000 0.0000000 0.00000000
## effectif_de_premiere             0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_brut_seconde_bac      0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_seconde_bac   0.0000000 0.0000000 1.0000000 0.00000000
## taux_acces_brut_premiere_bac     0.0000000 0.0000000 0.0000000 0.00000000
## taux_acces_attendu_premiere_bac 0.0000000 1.0000000 0.0000000 0.00000000
## PMP (Exact)                      0.4211673 0.1678348 0.1105488 0.05850214
## PMP (MCMC)                       0.4065800 0.1636600 0.1110200 0.05540000
##                                     00040
## commune                          0.00000000
## effectif_presents_serie_l        0.00000000
## effectif_presents_serie_es       0.00000000
## effectif_presents_serie_s        0.00000000
## taux_brut_de_reussite_serie_l    0.00000000
## taux_brut_de_reussite_serie_es   0.00000000
## taux_brut_de_reussite_serie_s    0.00000000
## taux_reussite_attendu_serie_l    0.00000000
## taux_reussite_attendu_serie_es   0.00000000
## taux_reussite_attendu_serie_s    1.00000000
## effectif_de_seconde              0.00000000
## effectif_de_premiere             0.00000000
## taux_acces_brut_seconde_bac      0.00000000
## taux_acces_attendu_seconde_bac   0.00000000
## taux_acces_brut_premiere_bac     0.00000000
## taux_acces_attendu_premiere_bac 0.00000000
## PMP (Exact)                      0.04983705
## PMP (MCMC)                       0.05872000
```

There is a small difference between the values of the two models, it is under 0.01 even more, even in the best five model, we got the same model with PMP (exact) and PMP(mcmc), even the both images about models are the same.

So, I can say that the covariates are acting in the same way .

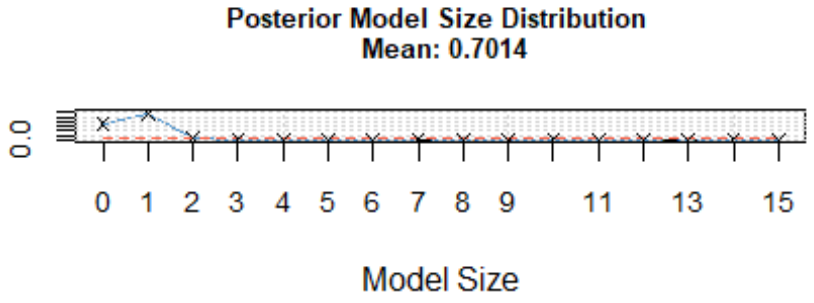
PREDICTIONS:

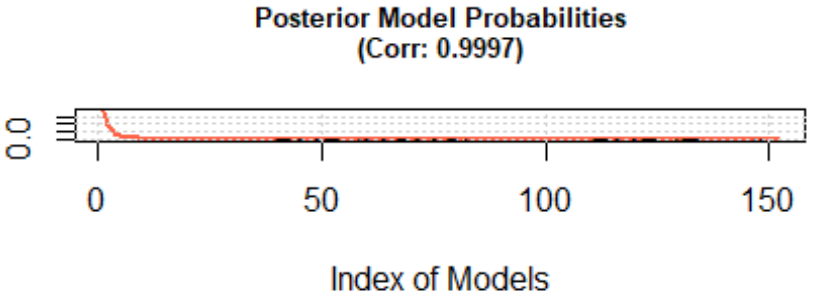
I chose to start with « mathematique » :

```r
idx = 21:26
d2_MA = numeric_MA1[-idx, ]
d3_MA = numeric_MA1[idx, ]

regMA.b = bms(X.data= d2_MA, burn=1e4, iter=5e4, nmodel=2e3)
```

```
##                                    PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.22584  2.348553e+00 4.6946008213
## taux_acces_attendu_seconde_bac  0.13478  1.075821e+00 2.9190313300
## taux_reussite_attendu_serie_es  0.07374  4.471324e-01 1.7857353638
## taux_reussite_attendu_serie_s   0.05770  3.029813e-01 1.3868650859
## taux_acces_brut_seconde_bac     0.03996  1.949601e-01 1.0877350334
## taux_acces_brut_premiere_bac    0.03968  2.555845e-01 1.4529102805
## taux_brut_de_reussite_serie_s   0.03488  1.713922e-01 1.0195587884
## taux_brut_de_reussite_serie_es  0.03356  1.469461e-01 0.8889392236
## taux_reussite_attendu_serie_l   0.01602 -4.891086e-02 0.9124137343
## effectif_presents_serie_l       0.00954  3.008842e-03 0.1000467852
## effectif_presents_serie_s       0.00886  2.865180e-03 0.0562047776
## commune                         0.00724  4.020320e-06 0.0002820149
## effectif_presents_serie_es      0.00628  1.431856e-03 0.0586240638
## taux_brut_de_reussite_serie_l   0.00512  8.843867e-03 0.1846294156
## effectif_de_premiere            0.00476 -1.147984e-05 0.0119352751
## effectif_de_seconde             0.00348  1.655919e-05 0.0111529044
##                                Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_es      0.9658259   9
## taux_reussite_attendu_serie_s       0.9660312  10
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_acces_brut_premiere_bac        0.9667339  15
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_reussite_attendu_serie_l       0.3470662   8
## effectif_presents_serie_l           0.6792453   2
## effectif_presents_serie_s           0.6726862   4
## commune                             0.8232044   1
## effectif_presents_serie_es          0.6528662   3
## taux_brut_de_reussite_serie_l       1.0000000   5
## effectif_de_premiere                0.3067227  12
## effectif_de_seconde                 0.4712644  11
##
## Mean no. regressors               Draws             Burnins
Time
##           "0.7014"             "50000"             "10000"     "3.037953
secs"
##  No. models visited      Modelspace 2^K          % visited         % Topm
odels
##              "5260"             "65536"                "8"
"100"
##           Corr PMP            No. Obs.         Model Prior             g-
```

```
Prior
##            "0.9997"               "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
##
## Time difference of 3.037953 secs
```

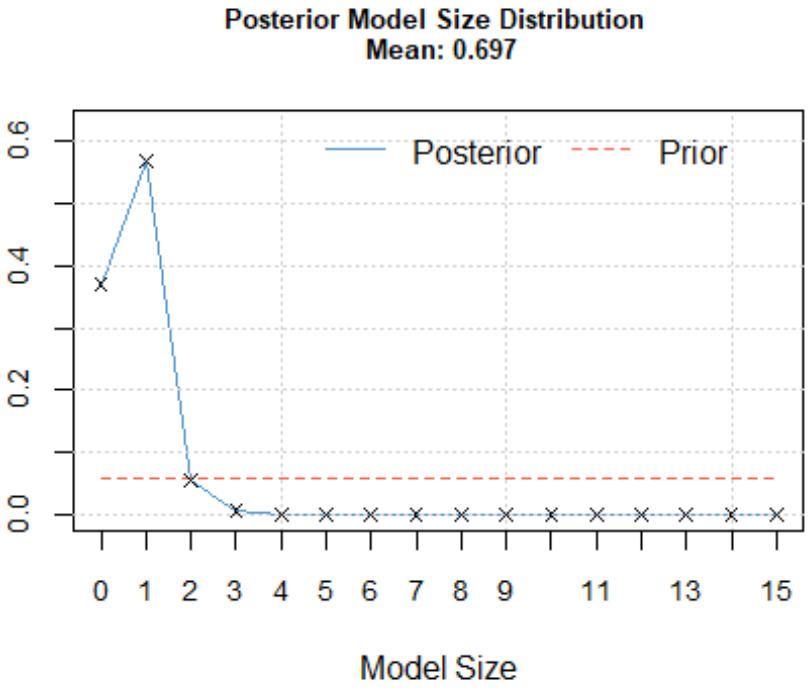





```
regMA.2 = bms(X.data= d2_MA, burn=1e4, iter=5e4, nmodel = 2)
```

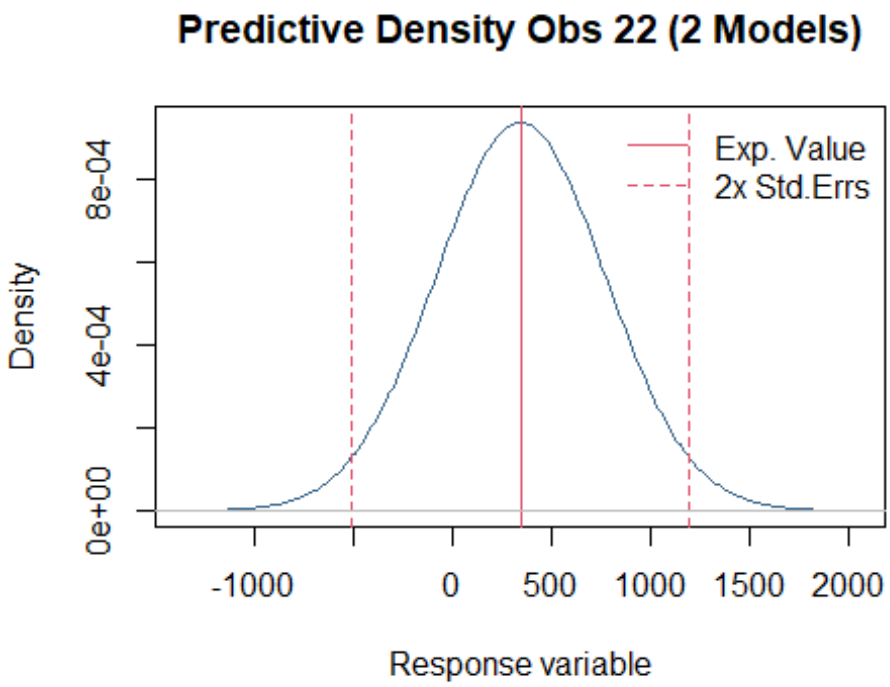
```
##                                      PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.21370  2.220840e+00 4.6027343156
## taux_acces_attendu_seconde_bac  0.13856  1.100598e+00 2.9628019392
## taux_reussite_attendu_serie_es  0.05814  3.488068e-01 1.5640596531
## taux_reussite_attendu_serie_s   0.05484  2.915527e-01 1.3567139388
## taux_acces_brut_seconde_bac     0.05372  2.701863e-01 1.2738524890
## taux_brut_de_reussite_serie_s   0.04212  2.131464e-01 1.1276428843
## taux_acces_brut_premiere_bac    0.03980  2.687495e-01 1.4751115332
## taux_brut_de_reussite_serie_es  0.03182  1.401460e-01 0.8674054077
## taux_reussite_attendu_serie_l   0.01570 -5.459440e-02 0.9029884302
## taux_brut_de_reussite_serie_l   0.00964  1.487128e-02 0.2432384773
## effectif_presents_serie_es      0.00784  1.634644e-03 0.0648609794
## effectif_de_seconde             0.00776  1.020713e-04 0.0196594847
## effectif_presents_serie_l       0.00726  2.329504e-03 0.0892689661
## effectif_presents_serie_s       0.00680  2.382697e-03 0.0481479358
## effectif_de_premiere            0.00560  4.599773e-05 0.0191669204
## commune                         0.00374  2.614171e-06 0.0002010158
```

```
##                                 Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_es      0.9735122   9
## taux_reussite_attendu_serie_s       0.9792123  10
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_acces_brut_premiere_bac        0.9773869  15
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_reussite_attendu_serie_l       0.2624204   8
## taux_brut_de_reussite_serie_l       1.0000000   5
## effectif_presents_serie_es          0.6938775   3
## effectif_de_seconde                 0.4484536  11
## effectif_presents_serie_l           0.6721763   2
## effectif_presents_serie_s           0.7323529   4
## effectif_de_premiere                0.4642857  12
## commune                             0.8877005   1
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.6970"             "50000"             "10000"     "2.845446
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "5200"             "65536"               "7.9"
"56"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "1.0000"               "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
##
## Time difference of 2.845446 secs
```



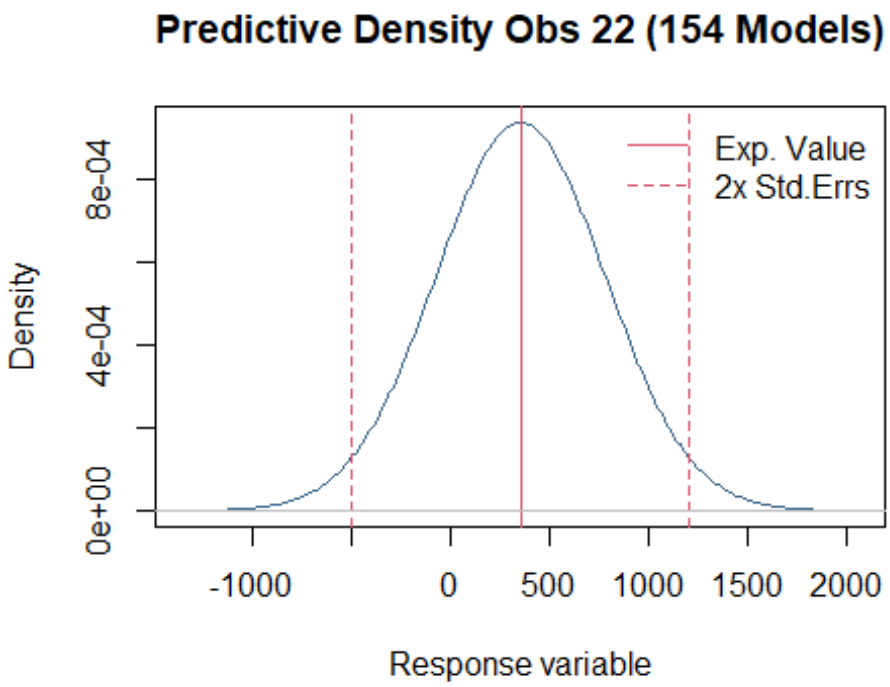
Prediction for "bms":

```
pdens1 = pred.density(regMA.2, d3_MA)
pdens1

## Call:
## pred.density(regMA.2, d3_MA)
##
## Densities for conditional forecast(s)
## 300 data points, based on 2 models;
##    Exp.Val. Std.Err.
## 21 347.2493 426.2786
## 22 347.2493 426.2786
## 23 343.7713 426.2284
## 24 343.7713 426.2284
## 25 343.7713 426.2284
## 26 343.7713 426.2284
```

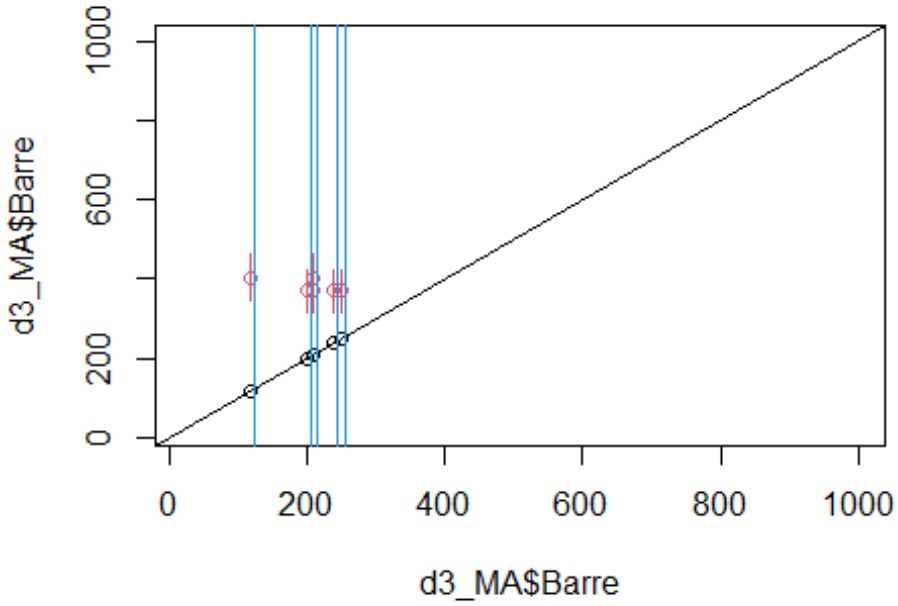



```
pdens.all = pred.density(regMA.b, d3_MA)
pdens.all
```

```
## Call:
## pred.density(regMA.b, d3_MA)
##
## Densities for conditional forecast(s)
## 300 data points, based on 154 models;
##    Exp.Val. Std.Err.
## 21 356.0016 425.5240
## 22 356.0016 425.5240
## 23 354.6916 425.5037
## 24 354.6916 425.5037
## 25 354.6916 425.5037
## 26 354.6916 425.5037
```



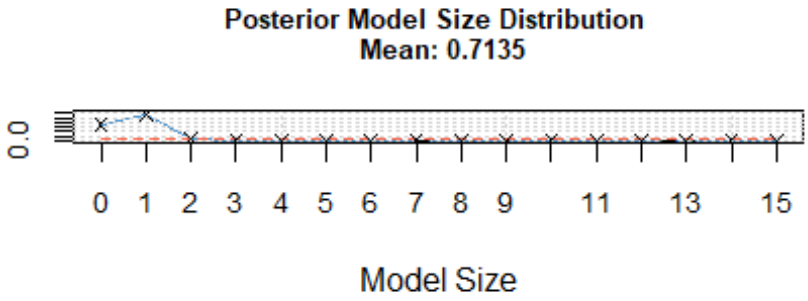

```
regMA.f = lm(Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_premi
ere_bac, data = numeric_MA1)
summary(regMA.f)

##
## Call:
## lm(formula = Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_pr
emiere_bac,
##     data = numeric_MA1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -387.32 -196.56 -130.83  -14.95 1696.20
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -494.324    260.593  -1.897  0.05840 .
## taux_reussite_attendu_serie_l     -7.882      4.360  -1.808  0.07124 .
## taux_acces_attendu_premiere_bac   17.833      5.407   3.298  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 419.5 on 513 degrees of freedom
## Multiple R-squared:  0.02539,    Adjusted R-squared:  0.02159
## F-statistic: 6.681 on 2 and 513 DF,  p-value: 0.001366

pred.f = predict(regMA.f, d3_MA, se.fit=T)
```

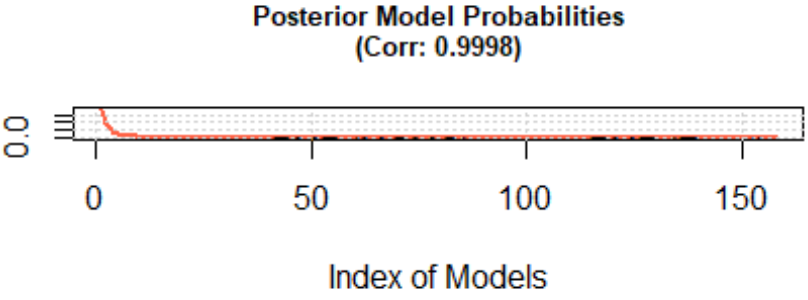



```
idx = 21:26
d2_AN = numeric_AN1[-idx, ]
d3_AN = numeric_AN1[idx, ]

regAN.b = bms(X.data= d2_AN, burn=1e4, iter=5e4, nmodel=2e3)
```

```
##                                     PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.21956  2.292928e+00 4.6934702625
## taux_acces_attendu_seconde_bac  0.14036  1.121683e+00 3.0011393390
## taux_reussite_attendu_serie_es  0.07152  4.309389e-01 1.7480493294
## taux_reussite_attendu_serie_s   0.05918  3.100876e-01 1.3982690359
## taux_acces_brut_premiere_bac    0.04252  2.647125e-01 1.4993951933
## taux_brut_de_reussite_serie_s   0.03940  1.932425e-01 1.0790120058
## taux_acces_brut_seconde_bac     0.03834  1.873814e-01 1.0680071168
## taux_brut_de_reussite_serie_es  0.03306  1.432612e-01 0.8776894793
## taux_reussite_attendu_serie_l   0.01952 -7.142247e-02 1.0043357575
## effectif_presents_serie_s       0.01144  4.639834e-03 0.0667038913
## effectif_presents_serie_l       0.00932  2.413373e-03 0.0987854418
## taux_brut_de_reussite_serie_l   0.00776  1.384853e-02 0.2321052681
## commune                         0.00722  3.913431e-06 0.0002809868
## effectif_de_premiere            0.00588  5.772384e-05 0.0145336358
## effectif_presents_serie_es      0.00534  1.565073e-03 0.0545328340
## effectif_de_seconde             0.00306  7.595992e-06 0.0090496793
##                                 Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_es      0.9675615   9
```

```
## taux_reussite_attendu_serie_s       0.9665428  10
## taux_acces_brut_premiere_bac        0.9285042  15
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_acces_brut_seconde_bac         1.0000000  13
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_reussite_attendu_serie_l       0.2500000   8
## effectif_presents_serie_s           0.7395105   4
## effectif_presents_serie_l           0.5579399   2
## taux_brut_de_reussite_serie_l       1.0000000   5
## commune                             0.8254848   1
## effectif_de_premiere                0.4829932  12
## effectif_presents_serie_es          0.7977528   3
## effectif_de_seconde                 0.3071895  11
##
## Mean no. regressors               Draws             Burnins
Time
##            "0.7135"             "50000"             "10000"     "3.938673
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "5318"             "65536"               "8.1"
"100"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "0.9998"               "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
##
## Time difference of 3.938673 secs
```

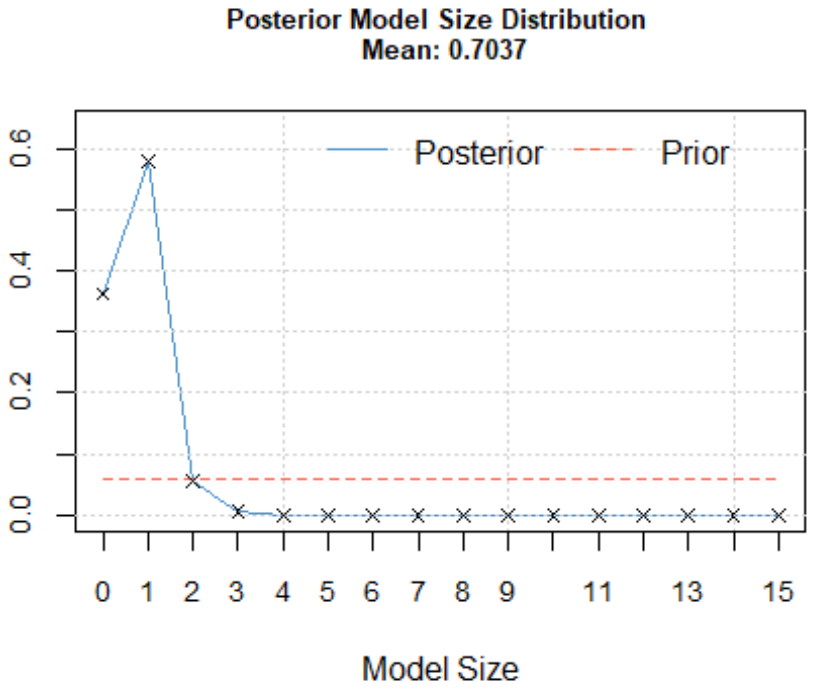





```
regAN.2 = bms(X.data= d2_AN, burn=1e4, iter=5e4, nmodel = 2)
```

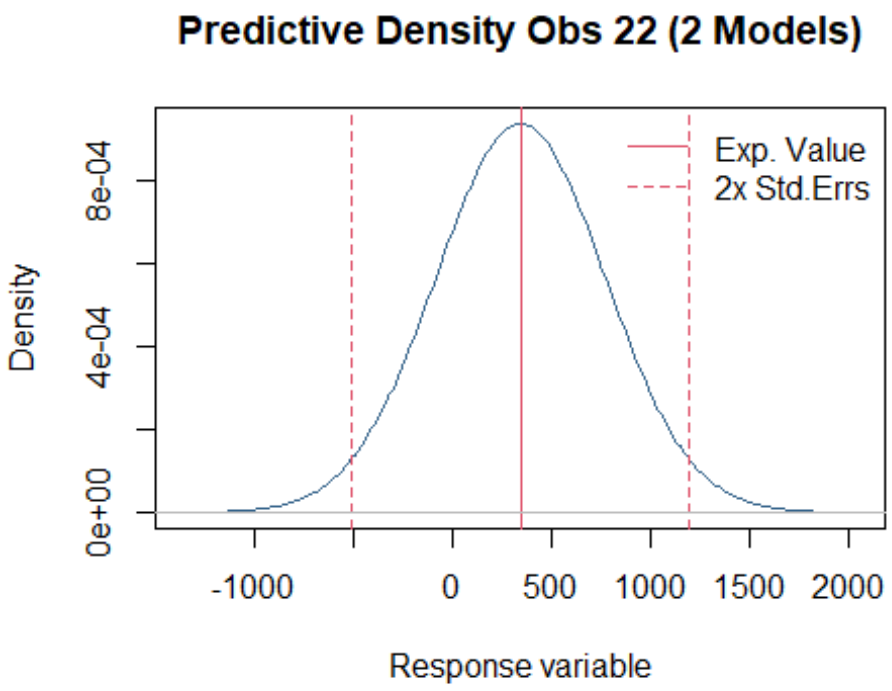
```
##                                    PIP     Post Mean      Post SD
## taux_acces_attendu_premiere_bac 0.21480  2.271273e+00 4.7093862556
## taux_acces_attendu_seconde_bac  0.14476  1.155713e+00 3.0160312539
## taux_reussite_attendu_serie_es  0.06942  4.321465e-01 1.7325700155
## taux_reussite_attendu_serie_s   0.06110  3.286820e-01 1.4223022722
## taux_brut_de_reussite_serie_es  0.04086  1.738666e-01 0.9644386585
## taux_brut_de_reussite_serie_s   0.03902  1.944642e-01 1.0830162806
## taux_acces_brut_premiere_bac    0.03446  2.198483e-01 1.3515394107
## taux_acces_brut_seconde_bac     0.03134  1.603210e-01 0.9904168013
## taux_reussite_attendu_serie_l   0.02152 -8.800077e-02 1.1179960920
## effectif_presents_serie_s       0.01230  4.622169e-03 0.0652734898
## taux_brut_de_reussite_serie_l   0.00784  1.371982e-02 0.2293403618
## effectif_de_seconde             0.00652  1.274275e-04 0.0135627784
## effectif_presents_serie_l       0.00630  2.238977e-03 0.0823134201
## effectif_presents_serie_es      0.00542  1.084794e-03 0.0521692163
## commune                         0.00500  1.753186e-06 0.0002340472
## effectif_de_premiere            0.00304  5.400233e-05 0.0096419042
##                                 Cond.Pos.Sign Idx
## taux_acces_attendu_premiere_bac     1.0000000  16
## taux_acces_attendu_seconde_bac      1.0000000  14
## taux_reussite_attendu_serie_es      0.9876116   9
## taux_reussite_attendu_serie_s       0.9846154  10
## taux_brut_de_reussite_serie_es      1.0000000   6
## taux_brut_de_reussite_serie_s       1.0000000   7
## taux_acces_brut_premiere_bac        0.9611143  15
```

```
## taux_acces_brut_seconde_bac          1.0000000  13
## taux_reussite_attendu_serie_l        0.2834572   8
## effectif_presents_serie_s            0.8195122   4
## taux_brut_de_reussite_serie_l        1.0000000   5
## effectif_de_seconde                  0.4877301  11
## effectif_presents_serie_l            0.7015873   2
## effectif_presents_serie_es           0.6346863   3
## commune                              0.7480000   1
## effectif_de_premiere                 0.4013158  12
##
## Mean no. regressors                Draws             Burnins
Time
##            "0.7037"             "50000"             "10000"     "2.183761
secs"
##  No. models visited      Modelspace 2^K           % visited         % Topm
odels
##              "5139"             "65536"               "7.8"
"55"
##            Corr PMP            No. Obs.         Model Prior             g-
Prior
##            "1.0000"               "510"        "random / 8"
"UIP"
##     Shrinkage-Stats
##          "Av=0.998"
##
## Time difference of 2.183761 secs
```



```
pdens1 = pred.density(regAN.2, d3_AN)
pdens1
```

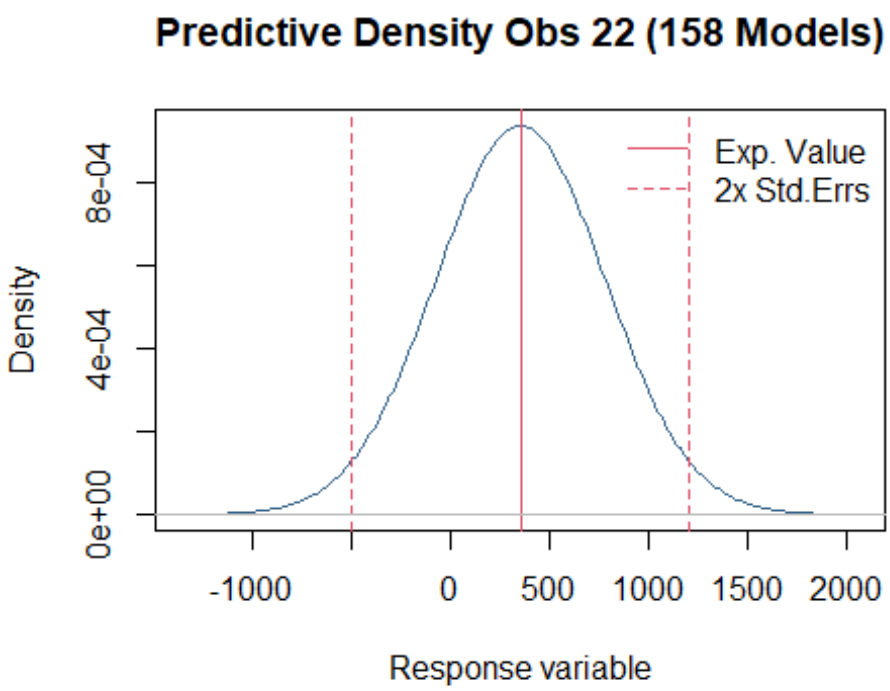
```
## Call:
## pred.density(regAN.2, d3_AN)
##
## Densities for conditional forecast(s)
## 300 data points, based on 2 models;
##    Exp.Val. Std.Err.
## 21 347.2493 426.2786
## 22 347.2493 426.2786
## 23 343.7713 426.2284
## 24 343.7713 426.2284
## 25 343.7713 426.2284
## 26 343.7713 426.2284
```




```
pdens.all = pred.density(regAN.b, d3_AN)
pdens.all
```

```
## Call:
## pred.density(regAN.b, d3_AN)
##
## Densities for conditional forecast(s)
## 300 data points, based on 158 models;
##    Exp.Val. Std.Err.
## 21 355.9847 425.5250
## 22 355.9847 425.5250
```

```
## 23 354.6925 425.5055
## 24 354.6925 425.5055
## 25 354.6925 425.5055
## 26 354.6925 425.5055
```



We have a small defference (0.0018 or 0.001) that I choose to ignore, so we have the same prediction for both dataset of "mathematique" and "anglais".

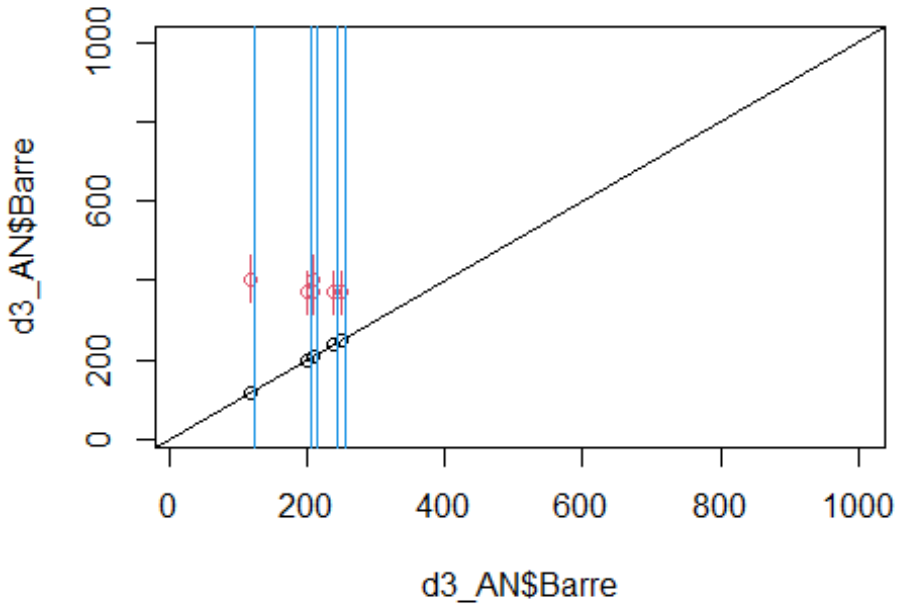
Let's do a linear regression then it's prediction to see if it got un the same result:

```
regAN.f = lm(Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_premi
ere_bac, data = numeric_AN1)
summary(regAN.f)
```

```
##
## Call:
## lm(formula = Barre ~ taux_reussite_attendu_serie_l + taux_acces_attendu_pr
emiere_bac,
##     data = numeric_AN1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -387.32 -196.56 -130.83  -14.95 1696.20
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -494.324    260.593  -1.897  0.05840 .
```

```
## taux_reussite_attendu_serie_l     -7.882      4.360  -1.808  0.07124 .
## taux_acces_attendu_premiere_bac   17.833      5.407   3.298  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 419.5 on 513 degrees of freedom
## Multiple R-squared:  0.02539,    Adjusted R-squared:  0.02159
## F-statistic: 6.681 on 2 and 513 DF,  p-value: 0.001366
```

```
pred.f = predict(regAN.f, d3_AN, se.fit=T)
```



CONCLUSION:

- I notice that we got the same result from all the different models for both dataset, there was a tiny difference that we can ignore.
- Even the plots of the prediction are the same.
- We can say that the covariates are acting in the same way on the different dataset, so on the different category “mathematique” and “anglais”.