

Rapport CYS

02 février, 2020

Contents

Etude des données W1_version1	1
Les données	1

```
##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##   pairs
##
## corrrplot 0.84 loaded
```

Etude des données W1_version1

Les données

```
CYS = read.csv("W1_version1.csv")
head(CYS)
```

```
##      prénom  Semestre Filière snapshot.1 snapshot.2 Snapshot.2...4m CYS.S3
## 1  YASSIR S3 2018-19    EEA      4.75      5.00                    non
## 2  CHEIMA S3 2018-19    EEA      3.00      3.00                    non
## 3  RAZA AKRAM S3 2018-19    EEA      3.50      3.75                    non
## 4  ANGELIQUE S3 2018-19    EEA      8.50     10.50                    non
## 5    DIANA S3 2018-19    EEA      3.50      9.50                    non
## 6    AXEL S3 2018-19    EEA      6.75     11.25                    non
##  CYS.S4 TP.S3 TP.S4 CMI Groupe.S3 Groupe.S4 Prof.TP
## 1      FR      non    Alba
## 2      FR      non    Siuban
## 3      FR      non    Akane
## 4      FR      non    Akane
## 5      FR      non    Alba
## 6      FR      non    Akane
```

```
str(CYS)
```

```
## 'data.frame':   181 obs. of  14 variables:
## $ prénom       : Factor w/ 160 levels "ABDELLAH","Abdou",...: 154 34 128 22 45 28 113 53 131 86 ..
## $ Semestre     : Factor w/ 2 levels "S3 2017-18","S3 2018-19": 2 2 2 2 2 2 2 2 2 ...
## $ Filière      : Factor w/ 1 level "EEA": 1 1 1 1 1 1 1 1 ...
## $ snapshot.1   : num  4.75 3 3.5 8.5 3.5 6.75 3.5 6 5 7.5 ...
## $ snapshot.2   : num  5 3 3.75 10.5 9.5 ...
## $ Snapshot.2...4m: Factor w/ 30 levels "","-", "10","10.75",...: 1 1 1 1 1 1 1 1 1 ...
## $ CYS.S3       : Factor w/ 2 levels "non","oui": 1 1 1 1 1 1 1 2 1 1 ...
```

```
## $ CYS.S4      : Factor w/ 3 levels "", "non", "oui": 1 1 1 1 1 1 1 1 1 ...
## $ TP.S3       : Factor w/ 2 levels "FR", "GB": 1 1 1 1 1 1 2 1 1 ...
## $ TP.S4       : Factor w/ 3 levels "", "FR", "GB": 1 1 1 1 1 1 1 1 1 ...
## $ CMI         : Factor w/ 2 levels "non", "oui": 1 1 1 1 1 1 1 1 1 ...
## $ Groupe.S3   : Factor w/ 8 levels "", "Akane", "Alba", ...: 3 6 2 2 3 2 3 2 3 6 ...
## $ Groupe.S4   : Factor w/ 3 levels "", "Nadia", "Virginia": 1 1 1 1 1 1 1 1 1 ...
## $ Prof.TP     : Factor w/ 3 levels "", "Didier", "Pierre": 1 1 1 1 1 1 1 1 1 ...
```

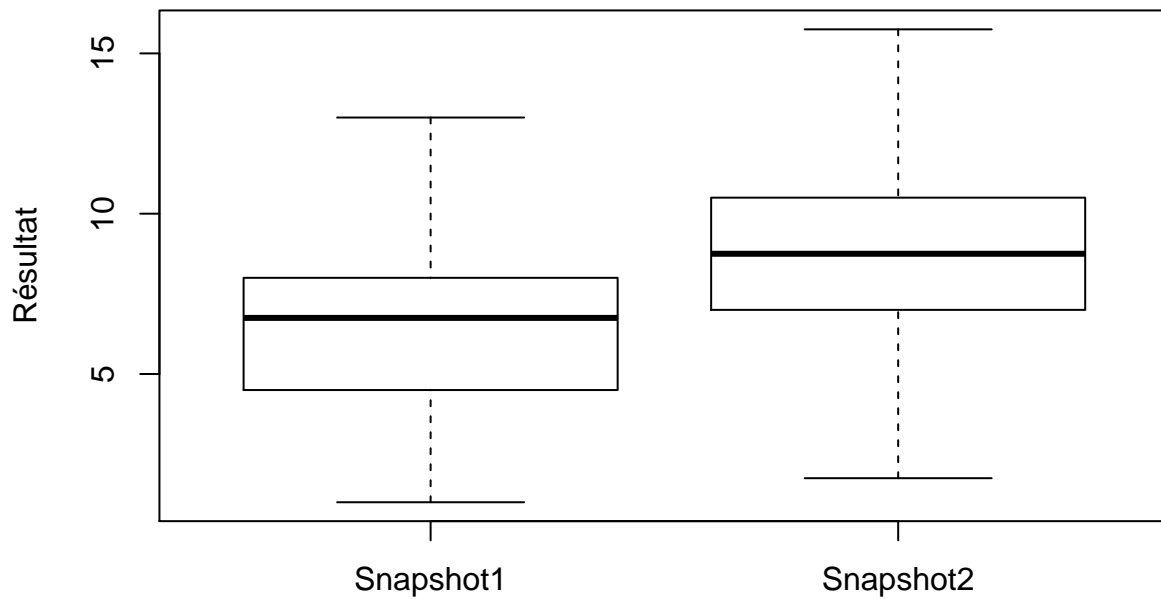
```
summary(CYS)
```

```
##      prénom      Semestre  Filière      snapshot.1      snapshot.2
## ALEXANDRE: 3    S3 2017-18: 38    EEA:181    Min. : 1.000    Min. : 1.75
## ALEXIS : 3    S3 2018-19:143
## HUGO : 3
## LUCAS : 3
## NICOLAS : 3
## VINCENT : 3
## (Other) :163
## Snapshot.2...4m CYS.S3    CYS.S4    TP.S3    TP.S4    CMI    Groupe.S3
##      :127    non:120      :137    FR:132    :143    non:150    Siuban :49
## - : 15    oui: 61    non: 21    GB: 49    FR: 15    oui: 31    Akane :34
## 11 : 3      oui: 23      GB: 23
## 11.75 : 3
## 14.5 : 3
## 6.75 : 3
## (Other): 27
##      Groupe.S4    Prof.TP
##      :154      :168
## Nadia : 15    Didier: 6
## Virginia: 12    Pierre: 7
##
##
##
##
```

```
# Stat. descriptives à compléter
```

```
boxplot(CYS$snapshot.1,CYS$snapshot.2,names=c("Snapshot1","Snapshot2"), ylab="Résultat",
        main="Résultats au Snapshot1 et au Snapshot2")
```

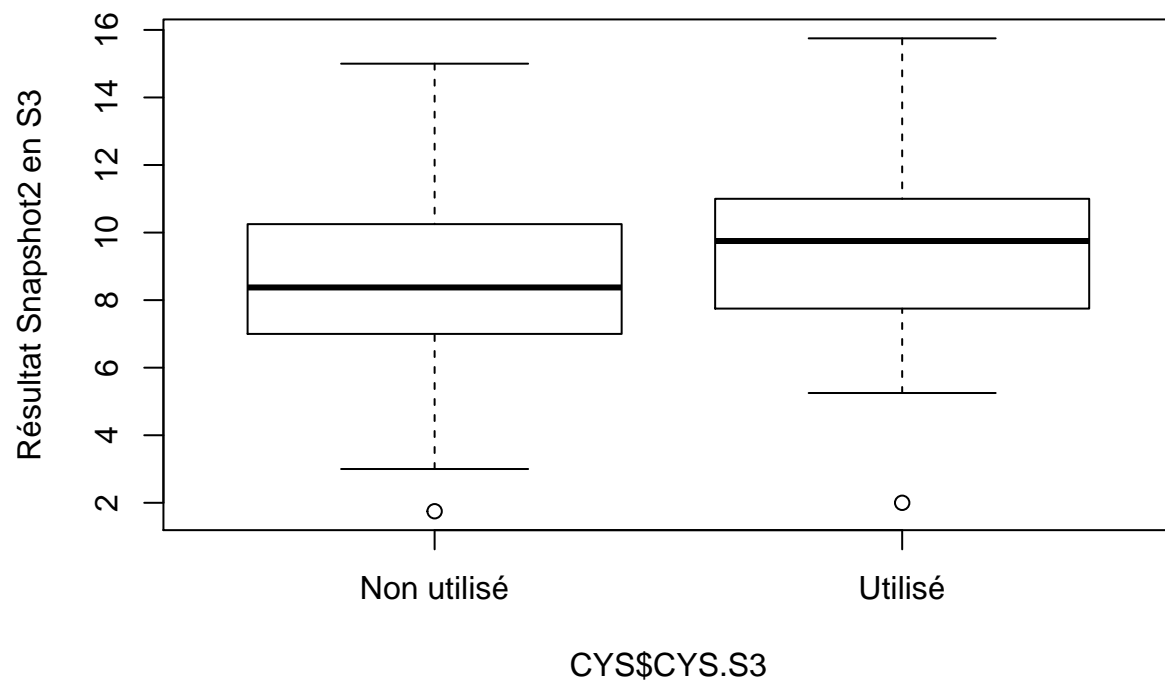
Résultats au Snapshot1 et au Snapshot2



Au regard de boxplot, on constate que le Snapshot2 prend souvent les valeurs plus grandes que le Snapshot1 d'où la progression obtenue en résultat.

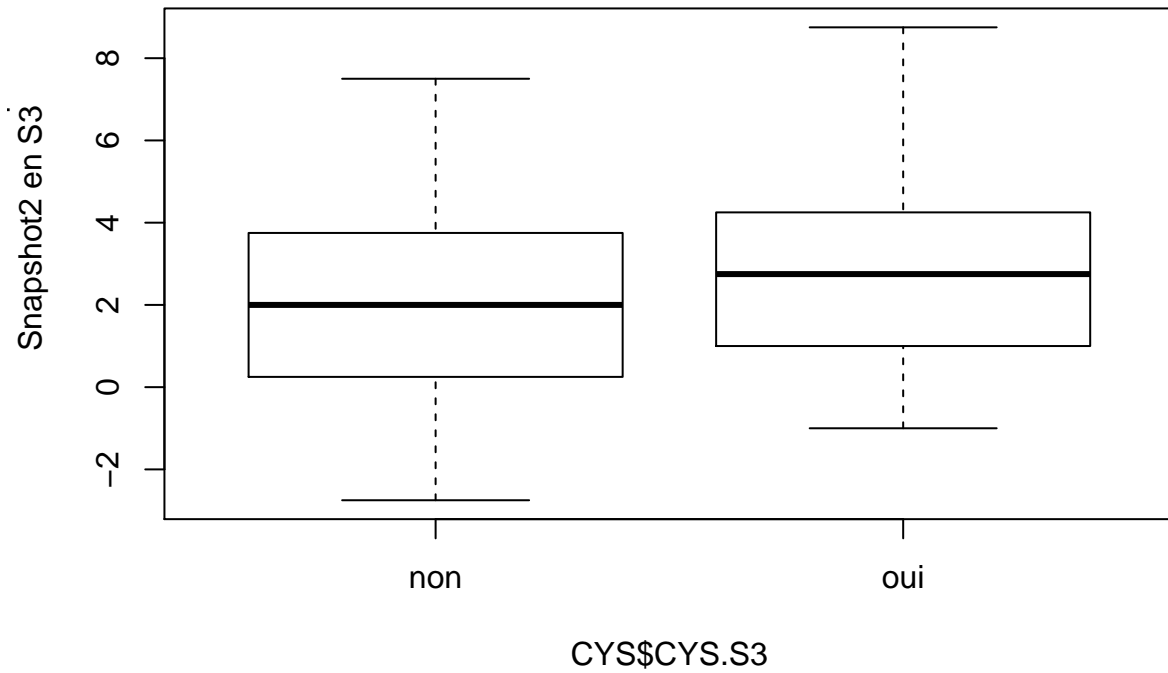
```
# Stat. descriptives à compléter
boxplot(CYS$snapshot.2~CYS$CYS.S3,names=c("Non utilisé","Utilisé"),
        ylab="Résultat Snapshot2 en S3",
        main=" Résultat Snapshot2 en S3 selon l'utilisation de CYS en S3")
```

Résultat Snapshot2 en S3 selon l'utilisation de CYS en S3



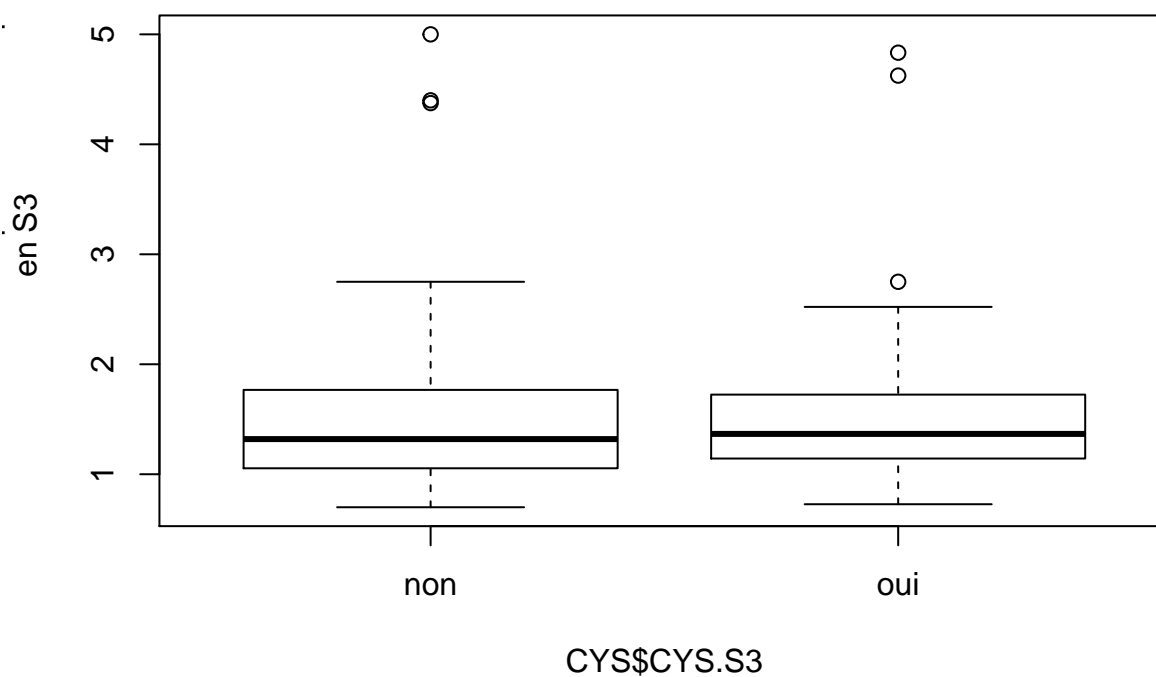
```
# Stat. descriptives à compléter
dif_snap=CYS$snapshot.2-CYS$snapshot.1
ratio_snap=CYS$snapshot.2/CYS$snapshot.1
boxplot(dif_snap~CYS$CYS.S3, ylab="Différence de résultat entre Snapshot1 et
  Snapshot2 en S3",
  main=" Différence de résultat entre Snapshot1 et
  Snapshot2 en S3 selon l'utilisation de CYS en S3")
```

Différence de résultat entre Snapshot1 et Snapshot2 en S3 selon l'utilisation de CYS en S3



```
boxplot(ratio_snap~CYS$CYS.S3, ylab="Ratio de résultat entre Snapshot1 et Snapshot2  
en S3",  
main=" Ratio de résultat entre Snapshot1 et Snapshot2  
en S3 selon l'utilisation de CYS en S3")
```

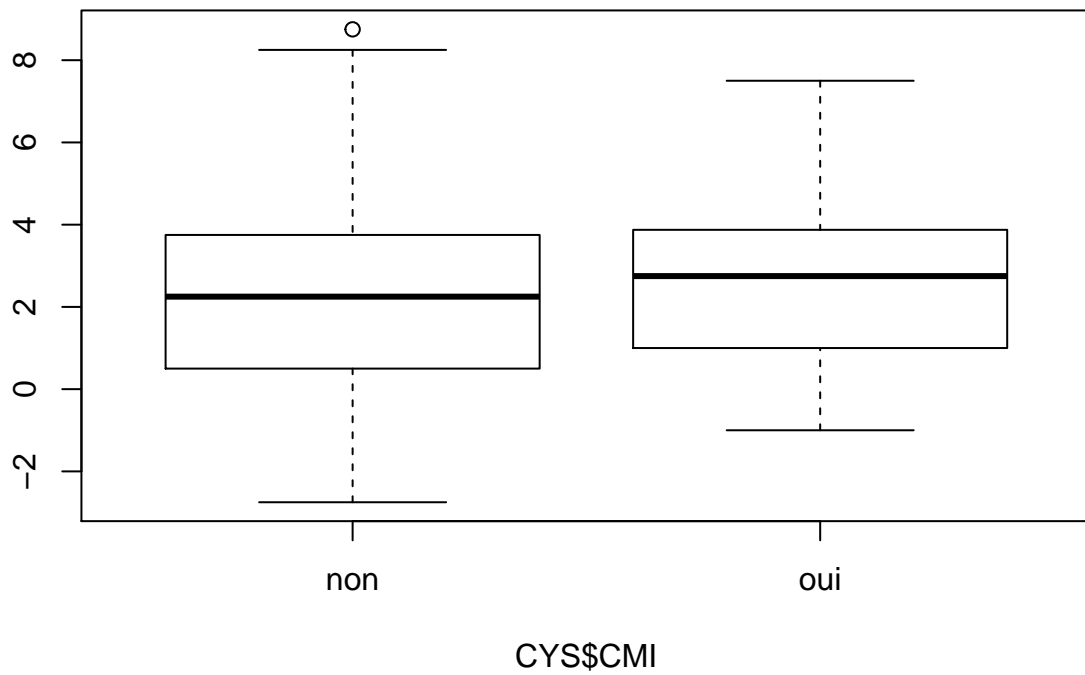
Ratio de résultat entre Snapshot1 et Snapshot2 en S3 selon l'utilisation de CYS en S3



```
# Stat. descriptives à compléter
boxplot(dif_snap~CYS$CMI, ylab="Différence de résultat entre Snapshot1 et Snapshot2 en S3",
        main=" Différence de résultat entre Snapshot1 et Snapshot2 en S3 selon CMI en S3")
```

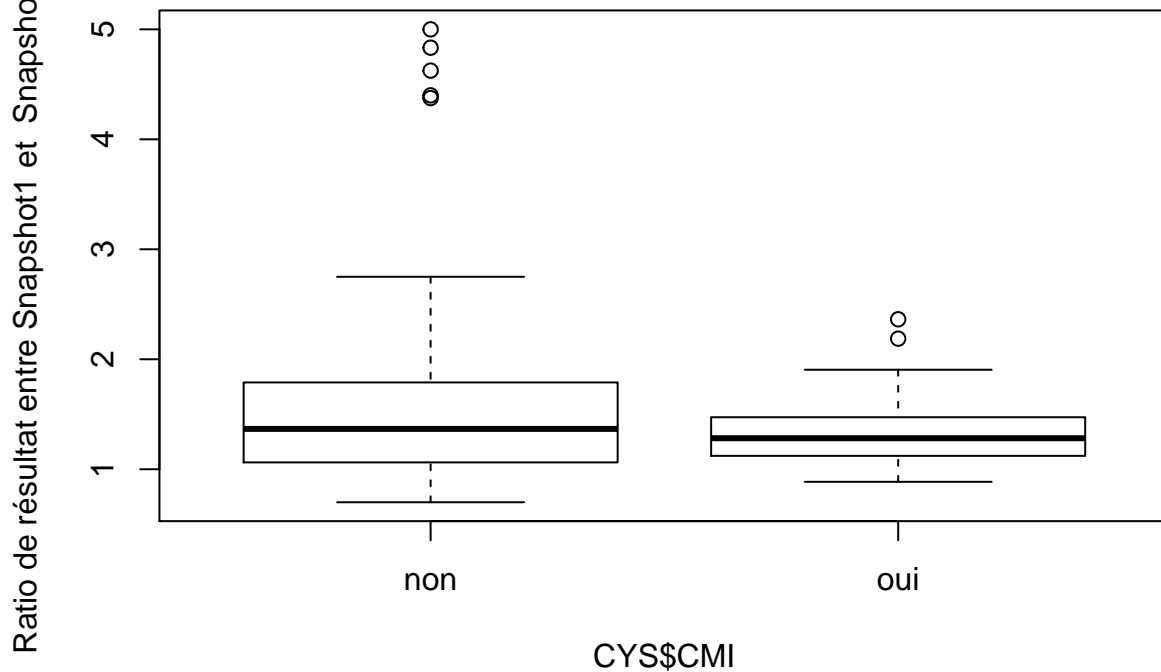
Différence de résultat entre Snapshot1 et Snapshot2 en S3

Différence de résultat entre Snapshot1 et Snapshot2 en S3 selon CMI



```
# Stat. descriptives à compléter  
boxplot(ratio_snap~CYS$CMI, ylab="Ratio de résultat entre Snapshot1 et Snapshot2 en S3",  
        main=" Ratio de résultat entre Snapshot1 et Snapshot2 en S3 selon CMI en S3")
```

Ratio de résultat entre Snapshot1 et Snapshot2 en S3 selon CMI en



```
## Test d'un modèle ANOVA de 3 facteurs(CYS S3, CMI, TP S3)
```

```
# A completer
```

```
mod1=lm(dif_snap~(CYS$CYS.S3+CYS$TP.S3+ CYS$CMI)^2,data=CYS)
```

```
summary(mod1)
```

```
##
```

```
## Call:
```

```
## lm(formula = dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2,
```

```
## data = CYS)
```

```
##
```

```
## Residuals:
```

```
## Min 1Q Median 3Q Max
```

```
## -4.875 -1.739 0.000 1.255 5.005
```

```
##
```

```
## Coefficients: (1 not defined because of singularities)
```

```
## Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 1.9954 0.2063 9.675 <2e-16 ***
```

```
## CYS$CYS.S3oui 0.7437 0.4941 1.505 0.134
```

```
## CYS$TP.S3GB -0.9954 2.1632 -0.460 0.646
```

```
## CYS$CMIoui 3.1250 2.2584 1.384 0.168
```

```
## CYS$CYS.S3oui:CYS$TP.S3GB 2.4328 2.2702 1.072 0.285
```

```
## CYS$CYS.S3oui:CYS$CMIoui -5.5515 2.3652 -2.347 0.020 *
```

```
## CYS$TP.S3GB:CYS$CMIoui NA NA NA NA
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```



```
## Residual standard error: 2.153 on 175 degrees of freedom
## Multiple R-squared:  0.1238, Adjusted R-squared:  0.09876
## F-statistic: 4.945 on 5 and 175 DF,  p-value: 0.0002929

# A completer - fonction lm
step.backward = step(mod1)

## Start:  AIC=283.56
## dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
##
##
## Step:  AIC=283.56
## dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$TP.S3 +
##           CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CYS.S3:CYS$TP.S3  1     5.3247 816.76 282.74
## <none>                                811.43 283.56
## - CYS$CYS.S3:CYS$CMI    1    25.5452 836.98 287.17
##
## Step:  AIC=282.74
## dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## <none>                                816.76 282.74
## - CYS$TP.S3              1    15.852 832.61 284.22
## - CYS$CYS.S3:CYS$CMI    1    52.880 869.64 292.09
```

Selon le test d'AIC, on trouve le meilleur modèle modAIC1:

$$dif_snap \sim CYS\$CYS.S3 + CYS\$TP.S3 + CYS\$CMI + CYS\$CYS.S3 : CYS\$CMI$$

```
# A completer
modAIC1=lm(dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI,data=CYS)
summary(modAIC1)

##
## Call:
## lm(formula = dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI,
##     data = CYS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.875 -1.725  0.000  1.275  5.025
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.9753     0.2055   9.613 < 2e-16 ***
## CYS$CYS.S3oui      0.8590     0.4825   1.780 0.076740 .
## CYS$TP.S3GB        1.2134     0.6565   1.848 0.066252 .
## CYS$CMIoui         0.9362     0.9641   0.971 0.332834
## CYS$CYS.S3oui:CYS$CMIoui -3.2340     0.9580 -3.376 0.000906 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.154 on 176 degrees of freedom
## Multiple R-squared: 0.118, Adjusted R-squared: 0.098
## F-statistic: 5.889 on 4 and 176 DF, p-value: 0.0001802
```

```
anova(modAIC1,mod1)
```

```
## Analysis of Variance Table
##
## Model 1: dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
## Model 2: dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      176 816.76
## 2      175 811.43  1    5.3247 1.1484 0.2854
```

La p_valeur de Test Fisher est 0,2854 supérieure que 0,05 donc on accepte le modèle modAIC1.

```
step.backward = step(mod1,direction="backward",k=log(nrow(CYS)))
```

```
## Start: AIC=302.75
## dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
##
##
## Step: AIC=302.75
## dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$TP.S3 +
##   CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CYS.S3:CYS$TP.S3  1    5.3247 816.76 298.73
## <none>                                811.43 302.75
## - CYS$CYS.S3:CYS$CMI    1   25.5452 836.98 303.16
##
## Step: AIC=298.73
## dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$TP.S3          1   15.852 832.61 297.01
## <none>                  816.76 298.73
## - CYS$CYS.S3:CYS$CMI  1   52.880 869.64 304.89
##
## Step: AIC=297.01
## dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## <none>                  832.61 297.01
## - CYS$CYS.S3:CYS$CMI  1   76.921 909.53 307.81
```

Selon le test d'BIC, on trouve le meilleur modèle:

```
dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3 : CYS$CMI
```

```
# A completer
modBIC1=lm(dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI,data=CYS)
anova(modBIC1,mod1)
```

```
## Analysis of Variance Table
##
## Model 1: dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
## Model 2: dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    177 832.61
## 2    175 811.43  2    21.177 2.2835  0.105
```

La p_valeur de Test Fisher est 0,105 supérieure que 0,05 donc on accepte le modèle modBIC1.

```
# A completer
```

```
anova(modBIC1,modAIC1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: dif_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
```

```
## Model 2: dif_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
```

```
## 1    177 832.61
```

```
## 2    176 816.76  1    15.852 3.4158 0.06625 .
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La p_valeur de Test Fisher est 0,105 supérieure que 0,05 donc on accepte le modèle modBIC1.

```
# A completer
```

```
mod2=lm(ratio_snap~(CYS$CYS.S3+CYS$TP.S3+ CYS$CMI)^2,data=CYS)
```

```
summary(mod1)
```

```
##
```

```
## Call:
```

```
## lm(formula = dif_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2,
```

```
##     data = CYS)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -4.875 -1.739  0.000  1.255  5.005
```

```
##
```

```
## Coefficients: (1 not defined because of singularities)
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      1.9954     0.2063   9.675  <2e-16 ***
```

```
## CYS$CYS.S3oui      0.7437     0.4941   1.505    0.134
```

```
## CYS$TP.S3GB      -0.9954     2.1632  -0.460    0.646
```

```
## CYS$CMIoui       3.1250     2.2584   1.384    0.168
```

```
## CYS$CYS.S3oui:CYS$TP.S3GB  2.4328     2.2702   1.072    0.285
```

```
## CYS$CYS.S3oui:CYS$CMIoui -5.5515     2.3652  -2.347    0.020 *
```

```
## CYS$TP.S3GB:CYS$CMIoui      NA         NA      NA      NA
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 2.153 on 175 degrees of freedom
```

```
## Multiple R-squared:  0.1238, Adjusted R-squared:  0.09876
```

```
## F-statistic: 4.945 on 5 and 175 DF,  p-value: 0.0002929
```

```
step.backward = step(mod2,direction="backward",k=log(nrow(CYS)))
```

```
## Start:  AIC=-101.45
```

```
## ratio_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
```

```
##
```

```
##
```

```
## Step:  AIC=-101.45
```

```
## ratio_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$TP.S3 +
## CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CYS.S3:CYS$TP.S3  1  0.05397 87.032 -106.54
## - CYS$CYS.S3:CYS$CMI    1  0.52871 87.507 -105.56
## <none>                    86.978 -101.45
##
## Step: AIC=-106.54
## ratio_snap ~ CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$TP.S3            1  0.18258 87.214 -111.36
## - CYS$CYS.S3:CYS$CMI  1  1.61594 88.648 -108.41
## <none>                    87.032 -106.54
##
## Step: AIC=-111.36
## ratio_snap ~ CYS$CYS.S3 + CYS$CMI + CYS$CYS.S3:CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CYS.S3:CYS$CMI  1  1.438 88.652 -113.60
## <none>                    87.214 -111.36
##
## Step: AIC=-113.6
## ratio_snap ~ CYS$CYS.S3 + CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CYS.S3  1  0.72326 89.376 -117.33
## - CYS$CMI      1  1.57982 90.232 -115.60
## <none>          88.652 -113.60
##
## Step: AIC=-117.33
## ratio_snap ~ CYS$CMI
##
##           Df Sum of Sq    RSS    AIC
## - CYS$CMI  1  1.0724 90.448 -120.36
## <none>          89.376 -117.33
##
## Step: AIC=-120.37
## ratio_snap ~ 1
```

Selon le test d’AIC, on trouve le meilleur modèle modBIC2:

ratio_snap ~ CYS\$CMI

De même façon, on trouve le meilleur modèle pour modéliser le ratio_snap:

```
# A compléter
modBIC2=lm(ratio_snap ~ CYS$CMI,data=CYS)
anova(modBIC2,mod2)
```

```
## Analysis of Variance Table
##
## Model 1: ratio_snap ~ CYS$CMI
## Model 2: ratio_snap ~ (CYS$CYS.S3 + CYS$TP.S3 + CYS$CMI)^2
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
```

```
## 1      179 89.376
## 2      175 86.978  4      2.3978 1.2061  0.31
```

La p_valeur de Test Fisher est 0,31 supérieure que 0,05 donc on accepte le modèle modBIC1.

Conclusion: Grâce au test ANOVA on trouve que:

- Si on considère la différence entre les deux snapshots on obtient le modèle

$$dif_snap \sim CYS\$CYS.S3 + CYS\$CMI + CYS\$CYS.S3 : CYS\$CMI$$

Cela montre l'impact de CMI et CYS S3 sur l'évolution de résultat.

- SI on considère la ratio entre les deux snapshots on obtient le modèle

$$ratio_snap \sim CYS\$CMI$$

Cela montre l'impact de CMI S3 sur l'évolution de résultat.

Dans ces deux cas on ne trouve pas l'effet de la variable TP S3.