Analyse de la résistance à la compression du béton

LI Yi, Muwala Abel

03/04/2018

Contents

1 Introduction

1.1 Problématiques

Résumé: Le béton est le matériau le plus important en génie civil. la résistance à la compression du béton est une fonction hautement non linéaire de l'âge et de ces composents.Notamment le ciment, le laitier de haut fourneau, les cendres volantes,l'eau, le superplastifiant, l'agrégat grossier et l'agrégat fin.

Question : Y'a-t-il une relation entre "Concrete compressive strength" et les autres attributs ? ou simplement y'a t-il une relation entre un attribut (Input variable) et Concrete compressive strength.

Y'a t-il un attribut qui détermine la résistance du béton à la compression?

Sujet : Concrete Compressive Strength (la résistance à la compression du béton)

1.2 Données

1.2.1 Contexte

La résistance réelle à la compression du béton (MPa) pour un mélange donné a été déterminé à partir du laboratoire en fonction de son âge spécifique (jours). Les données sont sous forme brute (non échelonnées).

Statistiques récapitulatives: Nombre de cas (observations): 1030 Nombre d'attributs: 9 Répartition des attributs: 8 variables d'entrée quantitatives et 1 variable de sortie quantitative Valeurs d'attribut manquantes: Aucune

1.2.2 Description des données

Nombre de cas (observations): 1030 Nombre d'attributs: 9 Répartition des attributs: 8 variables d'entrée quantitatives et 1 variable de sortie quantitative Valeurs d'attribut manquantes: Aucune

On donne pour chaque variable le nom, le type, l'unité de mesure et une brève description. La résistance à la compression du béton est le problème de la régression. L'ordre des variables correspond à l'ordre des chiffres le long des lignes de la base de données.

Pour des questions de simplicicté dans notre traitement , nous allons utiliser d'autres variables à la place des attributs : A1 pour "Cement"; A2 pour "Blast Furnace Slag"; A3 pour "Fly Ash"; A4 pour "Water"; A5 pour "Superplasticizer"; A6 pour "Coarse Aggregate"; A7 pour "Fine Aggregate"; A8 pour "Age"; A9 pour "Concrete compressive strength".

1.2.3 Sources:

Original Owner and Donor Prof. I-Cheng Yeh Department of Information Management Chung-Hua University, Hsin Chu, Taiwan 30067, R.O.C. e-mail:icyeh@chu.edu.tw TEL:886-3-5186511

Date Donated: August 3, 2007

1.3 Méthodologie

Méthode d'analyse et les visualisations - statistic simple et description des données. - visualisation des données en utilisant les graphiques et les diagrammes qu'on a vue dans les cours. - regression linéaire : qualité = a x Cement + b x Blast Furnace Slag + c x Fly Ash …

'library' utiles pour notre analyse: - ggplot2 - xlsx - car

2 Summary et visualisation des données

2.1 Les packages que nous utilisons

Nous utilisons le package ggplot2 pour la visulisation des données, le package xlsx pour le lecture des données.

```
# package
library(ggplot2)
library(xlsx)
```

```
## Loading required package: rJava

## Loading required package: xlsxjars
```

2.2 Lecture et Summary des donnée

Nous utilisons la fonction read.xlsx() pour lire les données dans le fichier "Concrete_Data.xls". On affiche les six première lignes.

```
df <- read.xlsx("Concrete_Data.xls",1)
str(df);
```

```
## 'data.frame':    1030 obs. of  9 variables:
##  $ A1: num  540 540 332 332 199 ...
##  $ A2: num  0 0 142 142 132 ...
##  $ A3: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A4: num  162 162 228 228 192 228 228 228 228 228 ...
```

```
##  $ A5: num  2.5 2.5 0 0 0 0 0 0 0 0 ...
##  $ A6: num  1040 1055 932 932 978 ...
##  $ A7: num  676 676 594 594 826 ...
##  $ A8: num  28 28 270 365 360 90 365 28 28 28 ...
##  $ A9: num  80 61.9 40.3 41.1 44.3 ...
```

```
head(df);
```

```
##      A1    A2 A3  A4  A5     A6    A7  A8       A9
## 1 540.0   0.0  0 162 2.5 1040.0 676.0  28 79.98611
## 2 540.0   0.0  0 162 2.5 1055.0 676.0  28 61.88737
## 3 332.5 142.5  0 228 0.0  932.0 594.0 270 40.26954
## 4 332.5 142.5  0 228 0.0  932.0 594.0 365 41.05278
## 5 198.6 132.4  0 192 0.0  978.4 825.5 360 44.29608
## 6 266.0 114.0  0 228 0.0  932.0 670.0  90 47.02985
```

Nous utilisons la fonctions summary pour afficher certains caractéristiques des données, par exemple: Le nombre minimal, le nombre maximaume, Médiane etc.

```
myvars <- c("A1","A2","A3","A4","A5","A6","A7","A8","A9")
summary(df[myvars])
```

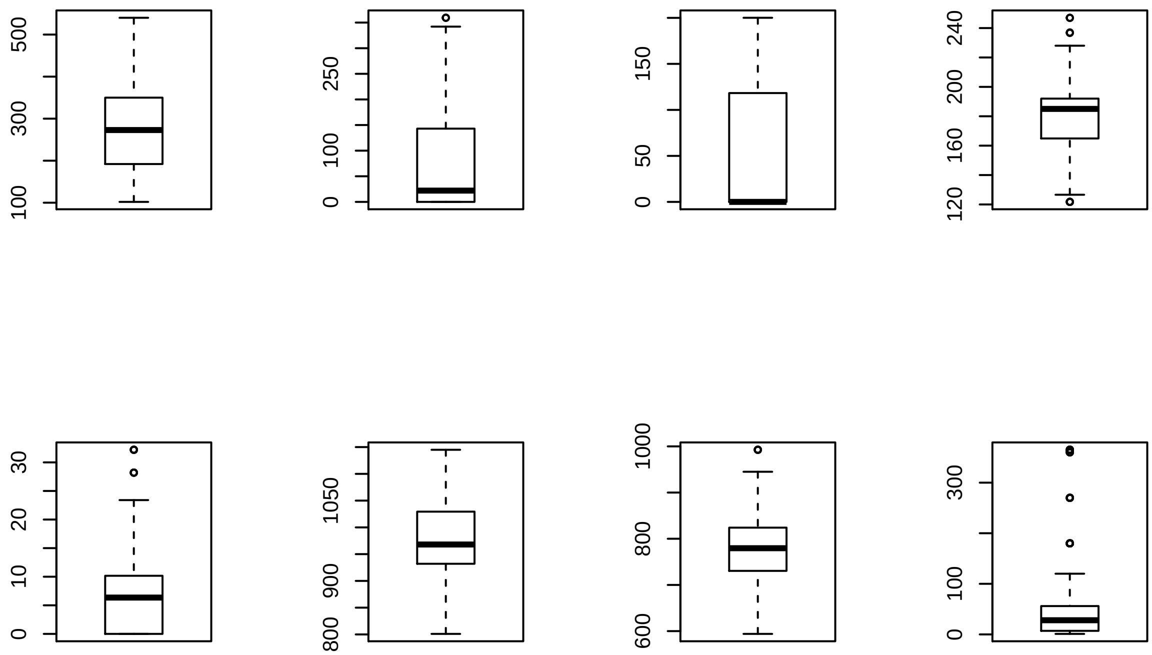
```
##        A1              A2              A3               A4
##  Min.   :102.0   Min.   :  0.0   Min.   :  0.00   Min.   :121.8
##  1st Qu.:192.4   1st Qu.:  0.0   1st Qu.:  0.00   1st Qu.:164.9
##  Median :272.9   Median : 22.0   Median :  0.00   Median :185.0
##  Mean   :281.2   Mean   : 73.9   Mean   : 54.19   Mean   :181.6
##  3rd Qu.:350.0   3rd Qu.:142.9   3rd Qu.:118.27   3rd Qu.:192.0
##  Max.   :540.0   Max.   :359.4   Max.   :200.10   Max.   :247.0
##        A5               A6               A7              A8
##  Min.   : 0.000   Min.   : 801.0   Min.   :594.0   Min.   :  1.00
##  1st Qu.: 0.000   1st Qu.: 932.0   1st Qu.:731.0   1st Qu.:  7.00
##  Median : 6.350   Median : 968.0   Median :779.5   Median : 28.00
##  Mean   : 6.203   Mean   : 972.9   Mean   :773.6   Mean   : 45.66
##  3rd Qu.:10.160   3rd Qu.:1029.4   3rd Qu.:824.0   3rd Qu.: 56.00
##  Max.   :32.200   Max.   :1145.0   Max.   :992.6   Max.   :365.00
##        A9
##  Min.   : 2.332
##  1st Qu.:23.707
##  Median :34.443
##  Mean   :35.818
##  3rd Qu.:46.136
##  Max.   :82.599
```

2.3 La visualisation des données

Nous utilisons la boîte à moustaches pour résumer quelques caractéristiques de position du caractère étudié (médiane, quartiles, minimum, maximum ou déciles).

```
attach(df)
opar <- par(no.readonly=TRUE)
par(mfrow=c(2,4))
boxplot(df["A1"])
boxplot(df["A2"])
boxplot(df["A3"])
boxplot(df["A4"])
```

```r
boxplot(df["A5"])
boxplot(df["A6"])
boxplot(df["A7"])
boxplot(df["A8"])
```

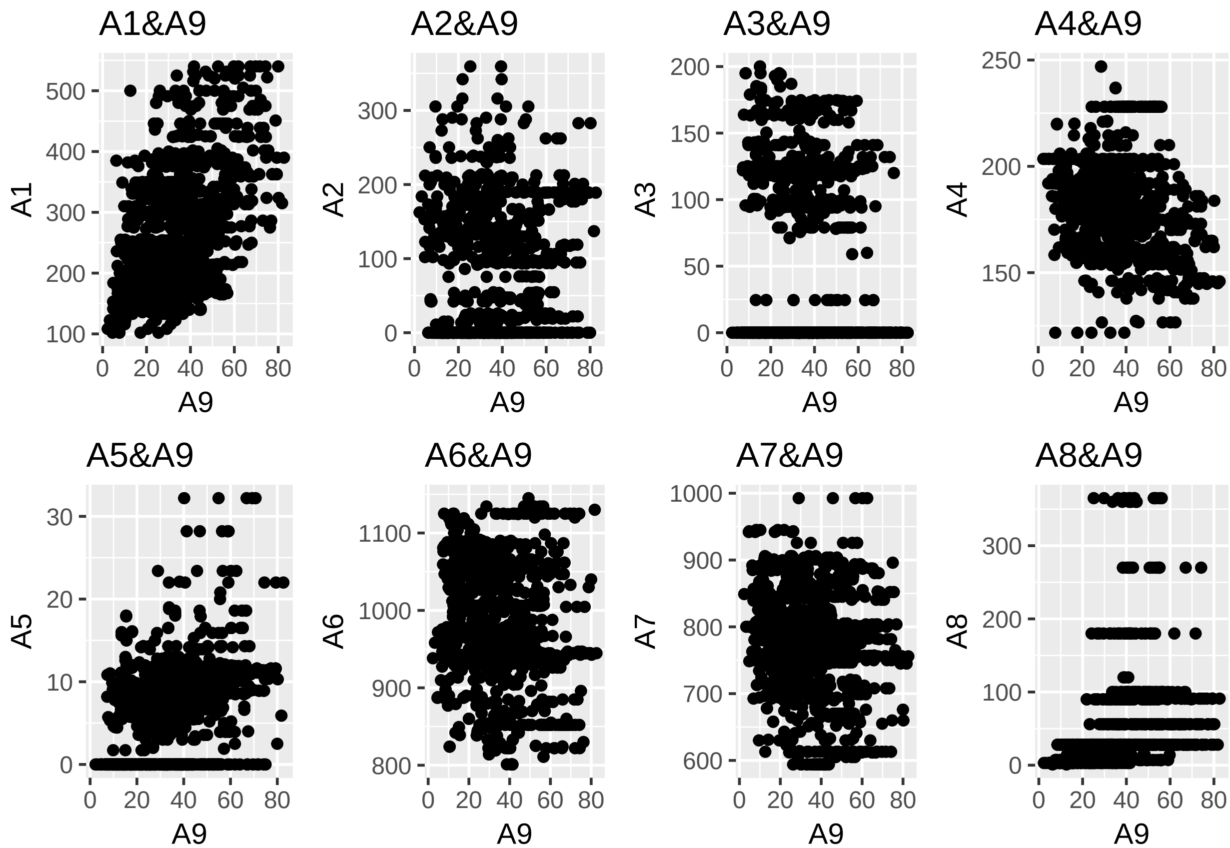


```r
par(opar)
detach(df)
```

Nous utilisons le nuage de points pour faire la représentation de données dépendant de plusieurs variables. Il nous permet de mettre en évidence le degré de corrélation entre A9 et les autres variables(A1,A2,A3. . . A8).

```r
library(ggplot2)
attach(df)
opar <- par(no.readonly=TRUE)
par(mfrow=c(2,4))

p1 <- ggplot(data=df, aes(x=A9, y=A1)) + geom_point() + labs(title="A1&A9", x="A9", y="A1")
p2 <- ggplot(data=df, aes(x=A9, y=A2)) + geom_point() + labs(title="A2&A9", x="A9", y="A2")
p3 <- ggplot(data=df, aes(x=A9, y=A3)) + geom_point() + labs(title="A3&A9", x="A9", y="A3")
p4 <- ggplot(data=df, aes(x=A9, y=A4)) + geom_point() + labs(title="A4&A9", x="A9", y="A4")
p5 <- ggplot(data=df, aes(x=A9, y=A5)) + geom_point() + labs(title="A5&A9", x="A9", y="A5")
p6 <- ggplot(data=df, aes(x=A9, y=A6)) + geom_point() + labs(title="A6&A9", x="A9", y="A6")
p7 <- ggplot(data=df, aes(x=A9, y=A7)) + geom_point() + labs(title="A7&A9", x="A9", y="A7")
p8 <- ggplot(data=df, aes(x=A9, y=A8)) + geom_point() + labs(title="A8&A9", x="A9", y="A8")
library(gridExtra)
grid.arrange(p1, p2, p3,p4,p5,p6,p7,p8, ncol=4)
```




```r
detach(df)
```

3 Les relations entre les variable

3.1 les relations

Avant de faire la régression linéaire, nous examinons les relations entre les variable. Nous utilisons la fonction cor() pour calculer les corrélation entre les variable. C'est étudier l'intensité de la liaison qui peut exister entre ces variables.

```r
vars <- df[,1:9]
vars <- as.matrix(sapply(vars, as.numeric))
cor(vars)
```

```
##             A1          A2           A3          A4          A5
## A1  1.00000000 -0.27519344 -0.397475440 -0.08154361  0.09277137
## A2 -0.27519344  1.00000000 -0.323569468  0.10728594  0.04337574
## A3 -0.39747544 -0.32356947  1.000000000 -0.25704400  0.37733956
## A4 -0.08154361  0.10728594 -0.257043997  1.00000000 -0.65746444
## A5  0.09277137  0.04337574  0.377339559 -0.65746444  1.00000000
## A6 -0.10935604 -0.28399823 -0.009976788 -0.18231167 -0.26630276
## A7 -0.22272017 -0.28159326  0.079076351 -0.45063498  0.22250149
## A8  0.08194726 -0.04424580 -0.154370165  0.27760443 -0.19271652
## A9  0.49783272  0.13482445 -0.105753348 -0.28961348  0.36610230
##              A6          A7          A8        A9
## A1 -0.109356039 -0.22272017  0.081947264  0.4978327
```

```
## A2 -0.283998230 -0.28159326 -0.044245801  0.1348244
## A3 -0.009976788  0.07907635 -0.154370165 -0.1057533
## A4 -0.182311668 -0.45063498  0.277604429 -0.2896135
## A5 -0.266302755  0.22250149 -0.192716518  0.3661023
## A6  1.000000000 -0.17850575 -0.003015507 -0.1649278
## A7 -0.178505755  1.00000000 -0.156094049 -0.1672490
## A8 -0.003015507 -0.15609405  1.000000000  0.3288770
## A9 -0.164927821 -0.16724896  0.328876976  1.0000000
```

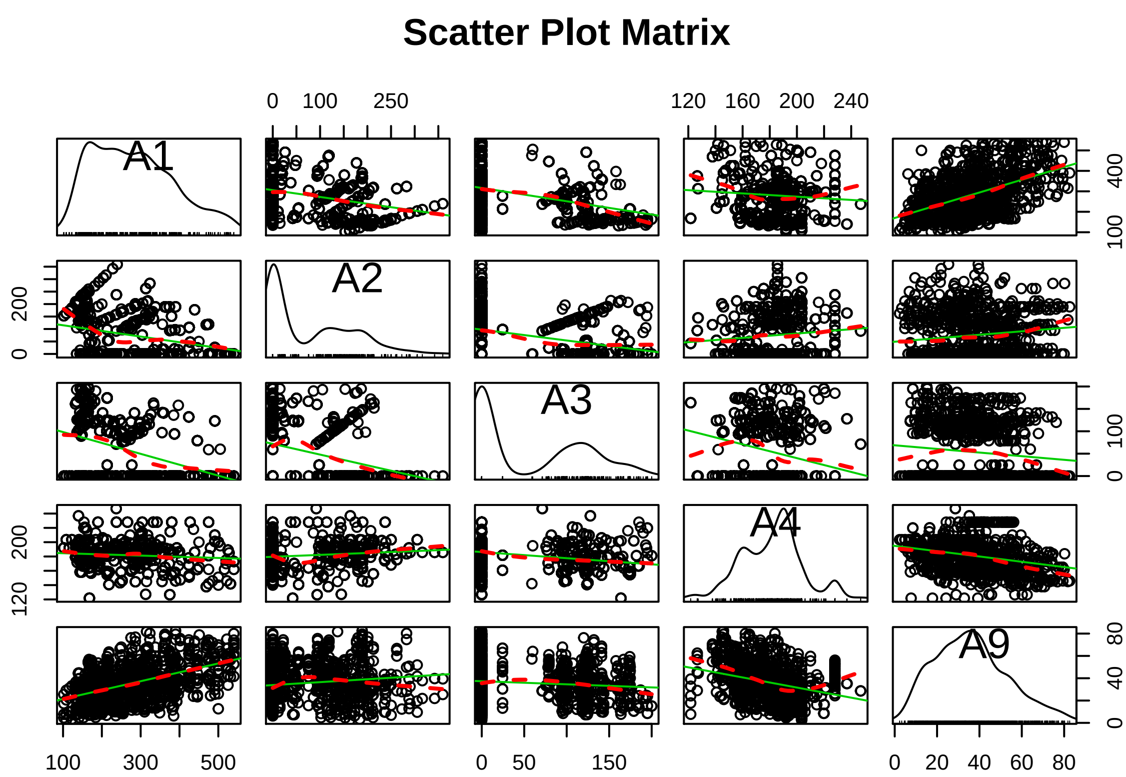
3.2 La visualiation des relations

Nous utilisons la fonctions scatterplotMatrix () pour afficher les schemas qui representent les relations entre les variables.

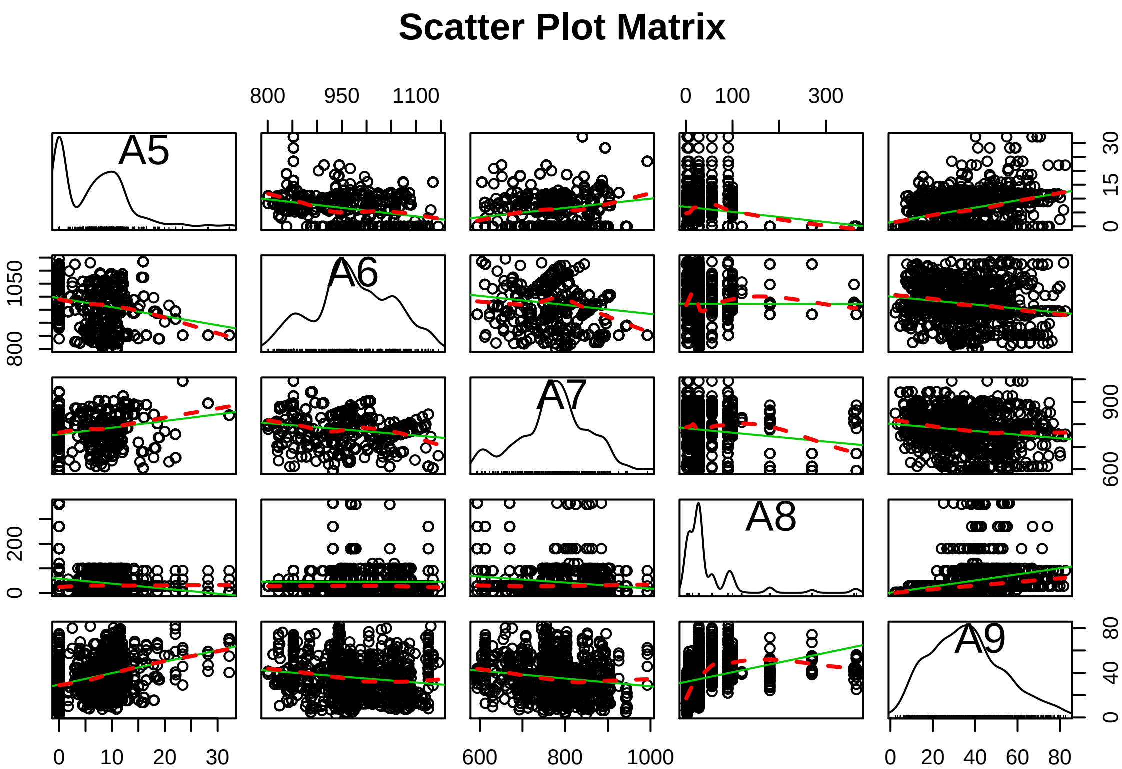
```r
library(car)
```

```
## Warning: package 'car' was built under R version 3.4.4
```

```r
myvars1 <- c("A1","A2","A3","A4","A9")
myvars2 <- c("A5","A6","A7","A8","A9")
#summary(df[myvars])
vars1 <- df[myvars1]
vars2 <- df[myvars2]
scatterplotMatrix(vars1, spread=FALSE, smoother.args=list(lty=2),
main="Scatter Plot Matrix")
```



```
scatterplotMatrix(vars2, spread=FALSE, smoother.args=list(lty=2),
main="Scatter Plot Matrix")
```



4 Régression linéaire multiple

4.1 Toutes les variables

Nous utilisons toutes les variables pour faire la Régression linéaire Variable dépendante(variable expliquée): A9 Variables indépendantes(variable explicative):A1,A2,A3,A4,A5,A6,A7,A8

```
vars <- df[,1:9]
fit1 <- lm(A9 ~ A1+A2+A3+A4+A5+A6+A7+A8,data=vars)
summary(fit1)
```

```
##
## Call:
## lm(formula = A9 ~ A1 + A2 + A3 + A4 + A5 + A6 + A7 + A8, data = vars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.653  -6.303   0.704   6.562  34.446
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) -23.163756  26.588421  -0.871 0.383851
```

```
## A1            0.119785   0.008489  14.110  < 2e-16 ***
## A2            0.103847   0.010136  10.245  < 2e-16 ***
## A3            0.087943   0.012585   6.988 5.03e-12 ***
## A4           -0.150298   0.040179  -3.741 0.000194 ***
## A5            0.290687   0.093460   3.110 0.001921 **
## A6            0.018030   0.009394   1.919 0.055227 .
## A7            0.020154   0.010703   1.883 0.059968 .
## A8            0.114226   0.005427  21.046  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 1021 degrees of freedom
## Multiple R-squared:  0.6155, Adjusted R-squared:  0.6125
## F-statistic: 204.3 on 8 and 1021 DF,  p-value: < 2.2e-16
```

4.2 Les variables avec des relations fortes

Nous ne utilisons que les variables qui ont une relation relativement fortes avec la variable dépendante.(|cor|>0.2) Variable dépendante(variable expliquée): A9 Variables indépendantes(variable explicative):A1,A4,A5,A8

```
fit2 <- lm(A9 ~ A1+A4+A5+A8,data=vars)
summary(fit2)
```

```
##
## Call:
## lm(formula = A9 ~ A1 + A4 + A5 + A8, data = vars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.766  -8.182  -0.276   7.200  40.689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.061564   4.716991   6.797 1.81e-11 ***
## A1           0.067793   0.003574  18.969  < 2e-16 ***
## A4          -0.138258   0.023484  -5.887 5.32e-09 ***
## A5           0.803663   0.082218   9.775  < 2e-16 ***
## A8           0.105404   0.006126  17.205  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.85 on 1025 degrees of freedom
## Multiple R-squared:  0.4986, Adjusted R-squared:  0.4966
## F-statistic: 254.8 on 4 and 1025 DF,  p-value: < 2.2e-16
```

4.3 Les termes du second degré

Nous ajoutons les termes du second degré

```
fit3 <- lm(A9 ~ A1+A4+A5+A8+I(A1^2)+I(A2^2)+I(A3^2)+I(A4^2),data=vars)
 summary(fit3)
```

```
##
```

```
## Call:
## lm(formula = A9 ~ A1 + A4 + A5 + A8 + I(A1^2) + I(A2^2) + I(A3^2) +
##     I(A4^2), data = vars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.714  -7.563   0.448   7.170  37.701
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.794e+01  2.008e+01   1.391   0.1645
## A1           1.214e-01  1.900e-02   6.390 2.52e-10 ***
## A4          -2.065e-01  2.104e-01  -0.982   0.3265
## A5           5.875e-01  8.565e-02   6.860 1.19e-11 ***
## A8           1.109e-01  5.932e-03  18.704  < 2e-16 ***
## I(A1^2)     -5.187e-05  2.916e-05  -1.779   0.0755 .
## I(A2^2)      2.666e-04  2.080e-05  12.817  < 2e-16 ***
## I(A3^2)      2.210e-04  5.057e-05   4.369 1.38e-05 ***
## I(A4^2)      6.629e-05  5.656e-04   0.117   0.9067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11 on 1021 degrees of freedom
## Multiple R-squared:  0.5699, Adjusted R-squared:  0.5666
## F-statistic: 169.1 on 8 and 1021 DF,  p-value: < 2.2e-16
```

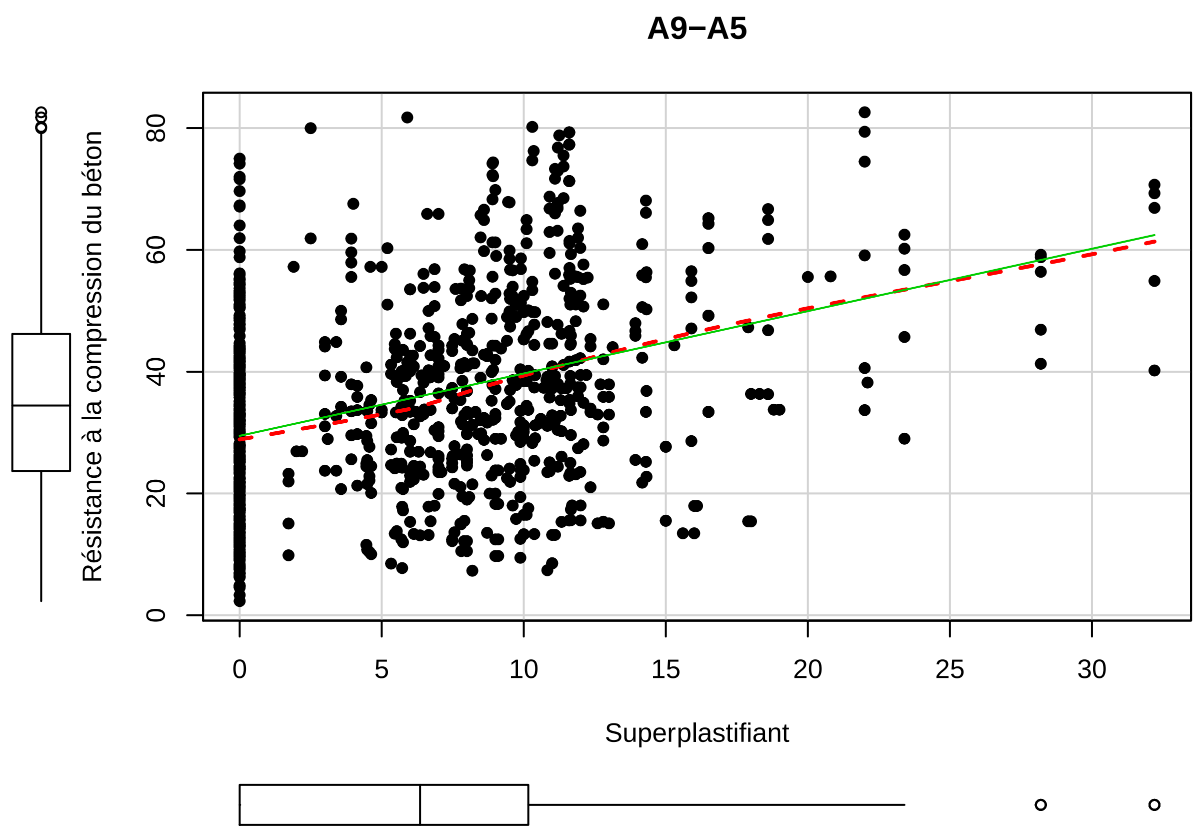
4.4 Explication de résultat de régression linéaire

On a réaliser 3 types de régression linéaire , chacun avec différents ensemble de variables. On compare les résultats de ces analyses , notamment le 'Multiple R-squared' et la première régression linéaire a une plus grande valeur du 'Multiple R-squared'. La première régression linéaire (avec tous les attibuts) est donc la meilleure car plus proche de '1'. Chaque fois que l'attribut A1 augmente de une unité , l'augmentation estimée de A9 est de '0.119785'. C'est aussi le cas avec les autres attributs à l'exception de A4 (l'eau). L'augmentation de A4 de une unité diminue la résistance à la compression de '0.150298' Ce résultat n'est obtenue quand ne considérant que l'augmentation estimée. Le fait de prendre en compte 'standard error' et 't value' et la distribution des échantillons ca donnerait un résultat différent et le processus d'analyse serait compliqué à notre niveau !

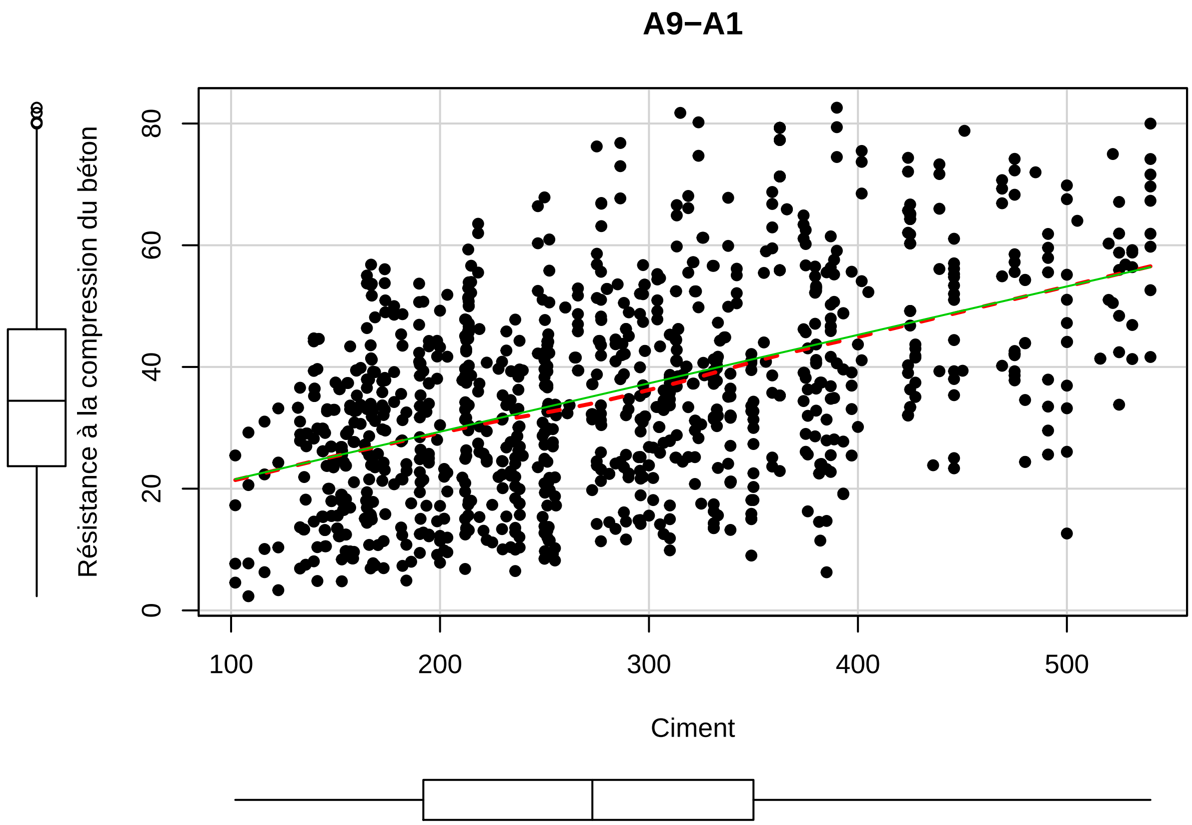
4.5 Affichage dans les graphes

Nous choisissons deux variables indépendantes et les afficher dans les graphes Avec les deux graphes ci-dessous on voit les tendances entre : -la résistance à la compression du béton(A9) et le superplastifiant(A5). -la résistance à la compression du béton(A9) et l'eau(A4).

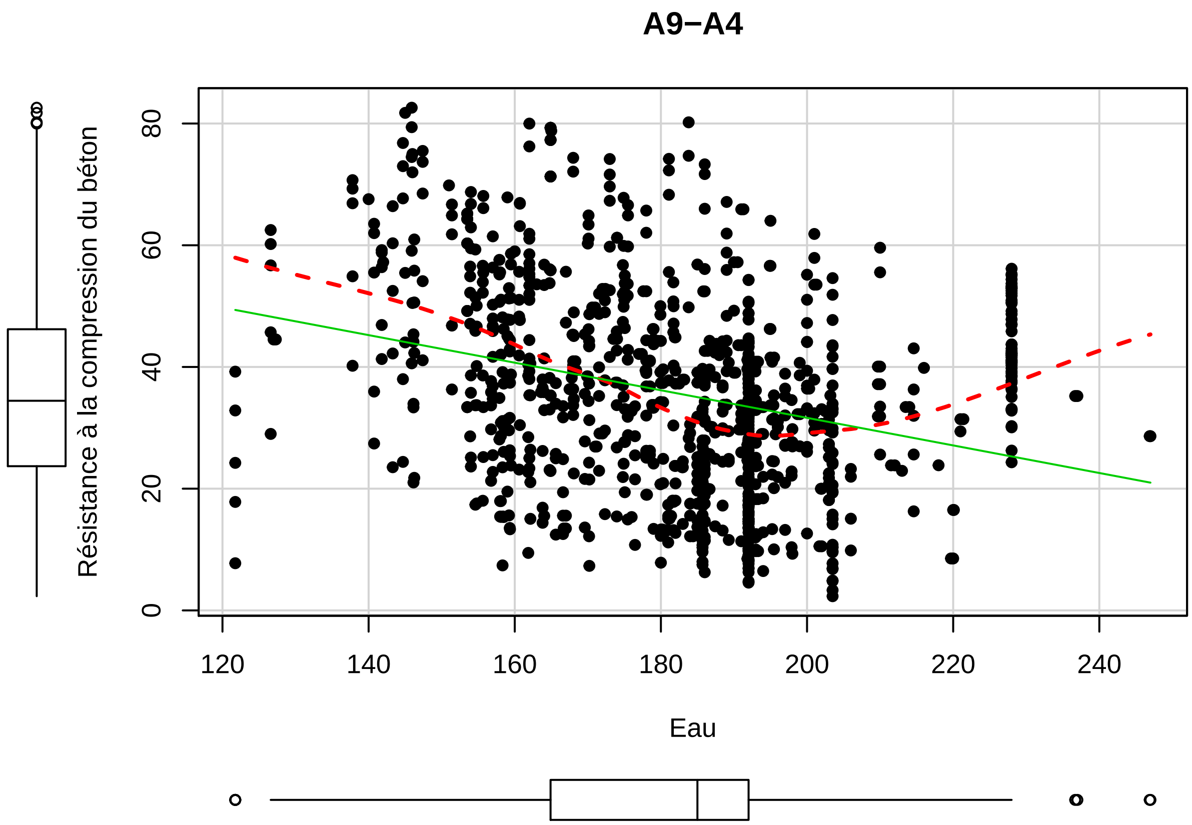
```r
library(car)
scatterplot(A9 ~ A5, data=vars,
spread=FALSE, smoother.args=list(lty=2), pch=19,
main="A9-A5",
xlab="Superplastifiant",
ylab="Résistance à la compression du béton")
```




```r
scatterplot(A9 ~ A1, data=vars,
spread=FALSE, smoother.args=list(lty=2), pch=19,
main="A9-A1",
xlab="Ciment",
ylab="Résistance à la compression du béton")
```




```
scatterplot(A9 ~ A4, data=vars,
spread=FALSE, smoother.args=list(lty=2), pch=19,
main="A9-A4",
xlab="Eau",
ylab="Résistance à la compression du béton")
```



5 Conclusion

Après l'analyse effetuée on peut déterminer les composants qui influence la résistance à la compression du béton (Augmentation ou diminution). On observe que le 'Superplastifiant' est le composant qui fait augmenter le plus la résistance à la compression du béton. On observe aussi que l'augmentation de l'eau diminue sensible la résistance à la compression du béton. Avec les graphiques 'scatter plot' on peut observer que les courbes des tendances entre la résistance à la compression du béton et d'un coté le superplastifiant et de l'autre l'eau sont contraire . Ce qui est cohérent étant donné que le superplastifiant pour diminuer la quantité d'eau que contient le béton et ainsi augmenter leur résistance.

Notons que les modèles de régression linéaire que nous avons utilisés ne sont pastrès rigoureux et qu'on devra les améliorés pour des analyses plus pertinentes.

6 References

[1] Statistical Tools For High-Throughput Data Analysis.Logicil R. disponible sur :http://www.sthda.com/french/wiki/logiciel-r

[2] I-Cheng Yeh.Concrete Compressive Strength Data set. 3/08/2007. disponible : https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

[3] Robert Prue . Boxplots Using the Amasing R and R command.1703/2014 disponible :https://www.youtube.com/watch?v=WmHcl1xLFEU