

# Regression Linéaire

Dimitri DELPECH,      Timothé FADENIPO,  
Matthis ARVOIS,      Ismael MADOU GAGI GREMA,  
Cheikh LO

2025-04-10

## Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Introduction</b>   | <b>3</b> |
| <b>2</b> | <b>Developpement</b>  | <b>3</b> |
| 2.1      | Revue empirique . . . . .   | 3        |
| 2.2      | Resulats des travaux . . . . .  | 6        |
| <b>3</b> | <b>Conclusion</b>   | <b>7</b> |
| <b>4</b> | <b>Annexe</b>   | <b>8</b> |
| 4.1      | Analyse univariée . . . . .   | 8        |
| 4.2      | Analyse bivariee . . . . .  | 17       |
| 4.3      | Ingénierie des données . . . . .                                      | 26       |
| 4.4      | Analyse multivariée . . . . .   | 28       |
| 4.5      | Histogramme des valeurs propres . . . . .                             | 29       |
| 4.6      | Analyse sur les 2 Premiers Axes . . . . .                             | 30       |
| 4.7      | Implications pour la Modélisation de la Résistance du Béton . . . . . | 35       |

---

|          |  |           |
|----------|--|-----------|
| <b>5</b> | <b>Modélisation linéaire</b>   | <b>36</b> |
| 5.1      | Résultats de la Régression . . . . .   | 36        |
| 5.2      | Interprétations de la régression . . . . .   | 36        |
| 5.3      | Diagnostic initial du modèle sans interactions (avec toutes les variables) . . . . . | 37        |
| 5.4      | Selection du Meilleur modèle . . . . .   | 44        |
| 5.5      | Diagnostic du modèle final retenu . . . . .  | 53        |
| <b>6</b> | <b>Statistique en Grande dimension</b>   | <b>56</b> |
| 6.1      | methode ridge . . . . .  | 56        |
| 6.2      | methode lasso . . . . .  | 58        |

# 1 Introduction

Le béton est un matériau très utilisé dans la construction, car il est solide, résistant dans le temps et peut être façonné facilement. Parmi ses propriétés, la résistance à la compression est la plus importante. Elle correspond à la capacité du béton à supporter une charge sans se casser. C'est cette résistance qui garantit la stabilité et la durabilité des ouvrages.

Cette résistance dépend de plusieurs facteurs : la quantité de ciment, d'eau, de granulats, d'adjuvants (ou additifs), mais aussi de l'âge du béton. Chaque composant joue un rôle, et c'est souvent leur combinaison qui influence le résultat final. Comprendre ces effets est essentiel pour concevoir un béton performant.

Dans ce travail, on cherche à identifier les facteurs qui ont un impact sur la résistance du béton à la compression, à partir de données mesurées en laboratoire.

On commence d'abord par une revue empirique qui présente ce que disent les études sur les effets des différents composants du béton. Ensuite, on passe à l'analyse des résultats obtenus avec notre modèle linéaire.

## 2 Developpemnt

### 2.1 Revue empirique

La résistance du béton est une propriété reconnue depuis longtemps. Si le béton est un élément important du développement de nos sociétés, c'est qu'il possède une résistance mécanique, en particulier à la compression, qui a permis aux architectes et concepteurs de concevoir des structures de plus en plus importantes et durables dans le temps. La propriété de la résistance du béton reste la propriété la plus importante du matériau pour du point de vue de l'ingénieur. Depuis longtemps, la relation entre la composition du béton et la résistance au béton fait écho dans le monde du génie civil et intéresse de nombreux chercheurs de ce domaine. Cependant aucune théorie fondamentale et universellement adoptée n'existe, en la matière, au-delà de la notion commune de rapport eau/ciment. Cette première partie a pour but de mettre en lumière les effets des variables explicatives sur la variable cible en l'occurrence sur la résistance à la compression du béton en se basant sur les travaux effectués dans ce sens.

Bien que dans toute approche fondamentale de la résistance à la compression, la nature du granulat représente un rôle secondaire néanmoins ceci reste important pour notre étude. Dans les années 1960, Walker et Bloem ont publié dans leur article (Effect of Aggregate size on properties of concrete, journal of AIC, septembre)

un résultat assez percutant. Il s'agit de la démonstration de l'effet négatif du volume de la dimension maximale sur la résistance. Trois effets du granulat sur la résistance du béton ont été énumérés à savoir un effet d'adhérence ; l'effet de confinement et l'effet plafond (le Bulletin des Laboratoires des Ponts et Chaussées, numéro 219, en janvier-février 1999, pages 41 à 52.).

A part son rôle important dans le phénomène de l'hydratation, l'eau est l'un des facteurs les plus importants au niveau de l'ouvrabilité du béton. L'augmentation du dosage en eau augmente la fluidité du béton et entraîne la diminution de la concentration en solides. Cependant, l'introduction excessive de l'eau provoque la chute de la résistance mécanique et sa durabilité (effect on composition parameters on fresh state properties of self-compacting concrete N.Bouhamou et al.). Le bulletin publié par FEBELCEM (Fédération de l'Industrie Cimentière Belge) avec l'auteur Ir C.Ployaert que la durabilité d'un béton dépend d'une faible porosité. De plus, pour les surfaces de bétonnage non coffrées sont les plus critiques du fait de leur exposition au soleil et au vent. Le contrôle de leur protection efficace contre toute évaporation intempestive de l'eau nécessaire à l'hydratation du ciment revêt d'une grande importance. D'où la liaison importante entre la résistance à la compression et l'eau.

Puis dans l'article publié par le département de génie civil à l'université de Mostaganem en Algérie, le rôle de superplastifiant a été mis en exergue. Le volume d'eau augmente avec l'augmentation du dosage en superplastifiant. L'une des explications avancées est l'augmentation de la viscosité de l'eau due au superplastifiant. De plus, il est noté que l'augmentation du dosage en superplastifiant a engendré une augmentation du taux de ségrégation statique dans le cas où le dosage est élevé entraînant ainsi une réduction de l'homogénéité du béton et par conséquent une réduction sa résistance à la compression. Par ailleurs, un dosage modéré en superplastifiants apporte un bénéfice supplémentaire, particulièrement aux premiers âges (48h, 72h), en raison d'une meilleure compacité et d'une dispersion plus efficace des grains de ciment.

Ensuite, Mehta et Monteiro (2014) soulignent que l'augmentation de la teneur en ciment améliore significativement la résistance mécanique, en particulier dans les premiers jours de durcissement. De même, Siddique et al. (2011) ont observé une corrélation positive entre la teneur en ciment et la résistance à la compression, confirmant que cette relation est particulièrement marquée dans les premières phases de durcissement. Toutefois, au-delà d'un certain seuil, une concentration excessive de ciment peut entraîner des effets négatifs, notamment une élévation de la chaleur d'hydratation et une augmentation du retrait, favorisant ainsi l'apparition de fissures (Neville, 2011). Il a également spécifié qu'il existe une diminution dans l'efficacité du ciment en appliquant de très haut dosage même en présence de su-

perplastifiant dans ADDIS B.J , Alexandre M.G (1994) . Le dosage du ciment dans le béton est très souvent relié à ses propriétés mécaniques et sa durabilité (effect on composition parameters on fresh state properties of self-compacting concrete N.Bouhamou et al.).

Par ailleurs, le laitier de haut fourneau est souvent utilisé pour remplacer une partie du ciment afin d'améliorer certaines propriétés du béton. Plusieurs études ont montré que son ajout réduit la résistance du béton dans les premiers jours, car sa réaction est plus lente que celle du ciment classique. Cependant, à plus long terme, il aide à renforcer le béton grâce à une réaction chimique qui produit des éléments solides supplémentaires, améliorant ainsi sa résistance (Jin et al., 2003).

De plus, l'utilisation des cendres volantes pour remplacer une partie du ciment a aussi un impact sur la résistance du béton. D'après Tan et al. (2016), leur ajout réduit la résistance dans les premiers jours, car elles réagissent plus lentement que le ciment. Toutefois, avec le temps, elles renforcent la structure du béton et améliorent sa durabilité grâce à une réaction chimique progressive. Zhao et al. (2019) précisent que leur effet dépend de la quantité utilisée et de la finesse des particules, ce qui influence directement la résistance finale du béton.

En outre, Les granulats fins, communément appelés "Fine Aggregate", constituent la fraction sableuse dans la formulation du béton. On considère généralement qu'il s'agit de particules de dimensions inférieures à 5 mm. La quantité de granulats fins, principalement constitués de sable, influence la résistance du béton. Selon Chatterji (2013), une bonne proportion de granulats fins améliore la compacité du mélange, réduit la porosité et augmente la résistance à la compression. Cependant, un excès de sable peut affaiblir la structure en diminuant la cohésion entre les particules de ciment et les granulats plus gros (Neville, 2011). Concrètement, une teneur adéquate en sable améliore le contact entre la pâte cimentaire et les grains, réduisant la porosité et favorisant ainsi un meilleur transfert des contraintes. Il est donc essentiel de trouver un équilibre pour garantir une répartition homogène des matériaux et optimiser la résistance du béton. Quant à l'étude du bulletin des laboratoires des ponts et chaussées (1999) intitulée "Influence du granulat sur la résistance à la compression des bétons", elle révèle qu'entre 60 % et 75 % de volume total d'agréats, une variation de la proportion de granulats fins peut entraîner un écart de l'ordre de 2 à 3 MPa en résistance à la compression. La raison tient à la formation d'une couche de pâte cimentaire plus ou moins épaisse selon la quantité de sables incorporée, conditionnant ainsi la compacité et la performance mécanique du béton. Enfin, L'âge du béton (Age) se définit comme la durée écoulée depuis le coulage et le compactage jusqu'à la réalisation des essais mécaniques, habituellement exprimée en nombre de jours (1, 2, 7, 28, etc.). Il revêt une grande importance, car plus le béton mûrit, plus la réaction

d'hydratation du ciment se poursuit, consolidant la microstructure et augmentant de manière significative la résistance en compression. Plusieurs études, dont celles de Neville (2011), Mindess et Young (2019), montrent que la résistance du béton augmente avec le temps en raison du processus d'hydratation du ciment. La majeure partie du gain de résistance se produit au cours des 28 premiers jours, période pendant laquelle le ciment continue de réagir avec l'eau pour former une structure solide. Cependant, le taux de durcissement ralentit après cette période, bien que certaines formulations, notamment celles contenant du laitier ou des cendres volantes, puissent encore voir leur résistance s'améliorer au-delà de 90 jours. Selon les observations présentées dans un mémoire de recherche intitulé "Impact du superplastifiant sur les propriétés physico-mécaniques du ciment" à l'université de Blida 1(2022-2023), la résistance à la compression connaît une croissance rapide entre 1 et 7 jours, passant de 20 % de la résistance ultime à près de 70 % autour de la première semaine. Typiquement, un saut d'environ 10 MPa (environ 50 %) est constaté entre 2 et 7 jours. Après ce cap, le rythme de durcissement ralentit, mais on atteint souvent plus de 90 % de la résistance finale à 28 jours.

## 2.2 Résultats des travaux

Nous présentons ici les résultats obtenus avec le modèle final, construit à partir des critères AIC/BIC, des résultats de l'analyse descriptives (notamment l'ACP) et des apports de la littérature.

La variable ciment (ciment) est fortement significative avec un effet positif (coefficient estimé à 0.097). Cela confirme que plus on augmente la quantité de ciment, plus la résistance à la compression augmente. Ce résultat est en accord avec la littérature, notamment Mehta et Monteiro (2014), qui ont montré que le ciment est un facteur clé de la résistance, surtout dans les premiers jours.

L'interaction age \* ciment est aussi significative (coefficient de 0.00041). Cela veut dire que l'effet du ciment sur la résistance dépend du temps : plus le béton vieillit, plus cet effet devient important.

La variable age seule est également significative (0.093). Cela confirme ce qu'indiquent Neville (2011) et Mindess et Young (2019) : la résistance du béton augmente avec le temps, surtout durant les 28 premiers jours, grâce à la poursuite du processus d'hydratation du ciment.

Le blast\_furnace (laitier de haut fourneau) a un effet positif (0.054) et très significatif. Cela reflète bien son rôle connu : il a peu d'impact au début, mais renforce le béton progressivement. L'interaction age \* blast\_furnace (0.00089) est aussi significative, ce qui montre que cet effet s'accroît avec le temps, comme le soulignent les travaux de Jin et al. (2003).

Concernant `fly_ash` (cendres volantes), son effet seul est légèrement positif (0.014) mais faiblement significatif ( $p = 0.10$ ). Par contre, l'interaction `age * fly_ash` est bien significative (0.00093), ce qui veut dire que les cendres volantes contribuent davantage à la résistance avec le temps. Cela correspond aux observations dans la littérature qui indiquent un effet lent mais croissant.

L'eau a été centrée avant d'ajouter le terme au carré, pour éviter les problèmes de colinéarité entre `water` et `water2`. Le coefficient de `water_c` est négatif (-0.295), ce qui montre que lorsque la quantité d'eau augmente, la résistance diminue. Le terme `water_c2` est positif (0.00168), ce qui indique que l'effet de l'eau n'est pas linéaire. Cela signifie qu'à très faibles ou très fortes doses, l'impact sur la résistance varie. Ce résultat est cohérent avec la notion de rapport eau/ciment, souvent citée dans la littérature, où trop ou pas assez d'eau nuit à la performance du béton.

Le `super_plast` (superplastifiant) a un effet positif et significatif (0.443), ce qui montre qu'il améliore la résistance du béton, surtout à court terme. L'interaction `super_plast * water_c` (0.01795) est aussi significative, ce qui veut dire que l'effet du superplastifiant dépend de la quantité d'eau. Cela rejoint les travaux montrant que le superplastifiant agit sur la fluidité, mais que son effet combiné avec l'eau est déterminant.

En résumé, le modèle final contient six variables principales (`age`, `cement`, `water`, `super_plast`, `blast_furnace`, `fly_ash`) et cinq interactions significatives. Tous les effets inclus sont cohérents avec la littérature et les résultats de l'analyse exploratoire.

Sur le plan de la performance, la validation croisée à 10 groupes donne une erreur moyenne de 109.48 pour le modèle basé uniquement sur les critères AIC, BIC et Fisher. Avec le modèle final, qui inclut des interactions et un effet quadratique, cette erreur descend à 64.56. Par ailleurs, le  $R^2$  ajusté du modèle final est de 0.7768, ce qui signifie que le modèle explique environ 78 % de la variabilité observée dans la résistance à la compression.

### 3 Conclusion

Ce travail avait pour objectif d'identifier les facteurs qui influencent la résistance du béton à la compression, à partir d'une base de données expérimentales. Pour cela, nous avons combiné des critères statistiques (AIC/BIC), les résultats de l'analyse descriptive, ainsi que des éléments issus de la littérature.

Le modèle final comprend six variables principales, plusieurs interactions et un effet quadratique de l'eau. Il permet d'expliquer environ 78 % de la variabilité de

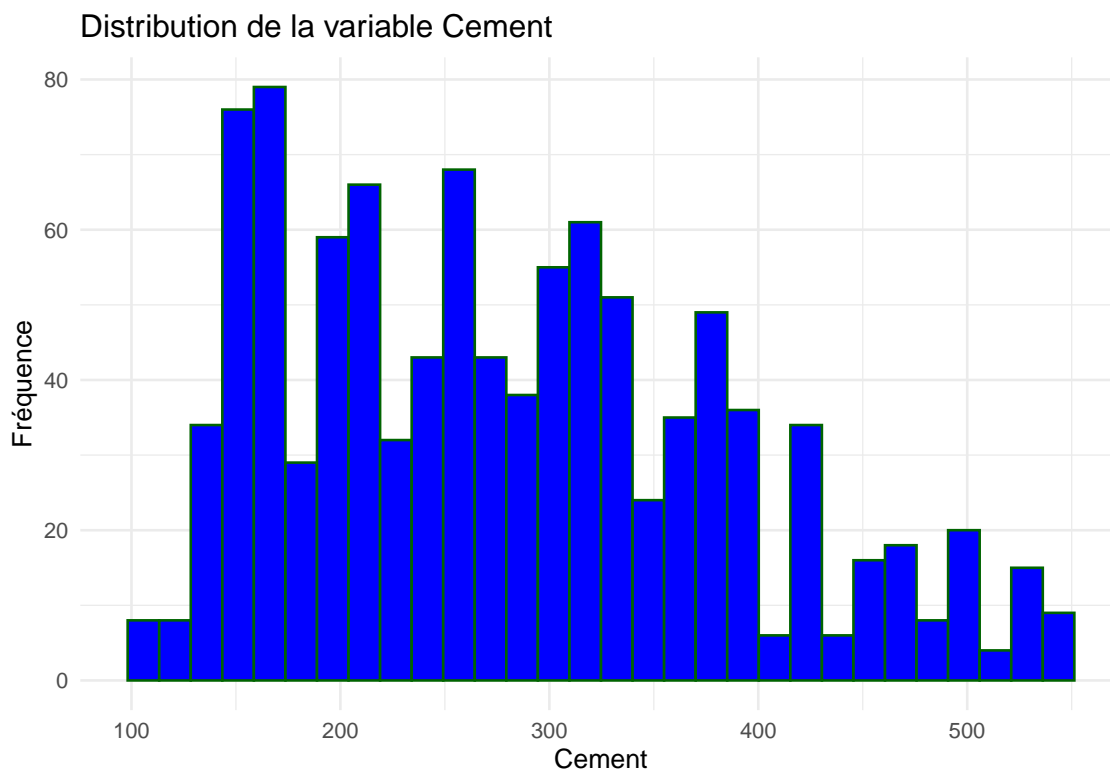
la résistance ( $R^2$  ajusté = 0.7768), avec une erreur moyenne de 64.56 en validation croisée à 10 groupes.

Les effets estimés sont globalement en accord avec les connaissances du domaine. Ils mettent en évidence le rôle important du ciment, de l'eau, du superplastifiant, du laitier et des cendres volantes, ainsi que des interactions liées à l'évolution dans le temps et à l'effet non linéaire de l'eau. Malgré certaines limites sur les résidus (autocorrélation, hétéroscédasticité, et normalité imparfaite), le modèle offre une base solide pour analyser et interpréter la résistance du béton à la compression.

## 4 Annexe

### 4.1 Analyse univariée

#### 4.1.1 Cement

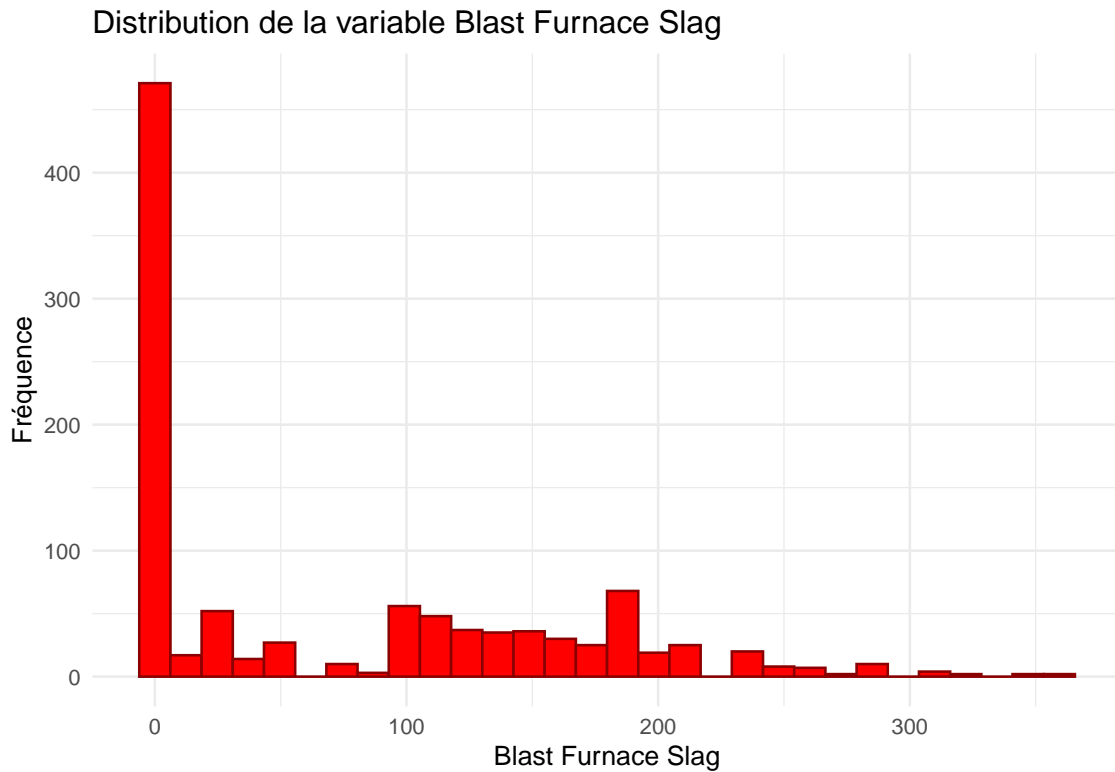


L'histogramme montre une distribution étalée avec plusieurs pics, indiquant une variabilité dans les proportions de ciment utilisées. Les valeurs les plus courantes se situent entre 100 et 400 kg/m<sup>3</sup>. Cette répartition suggère que différentes for-



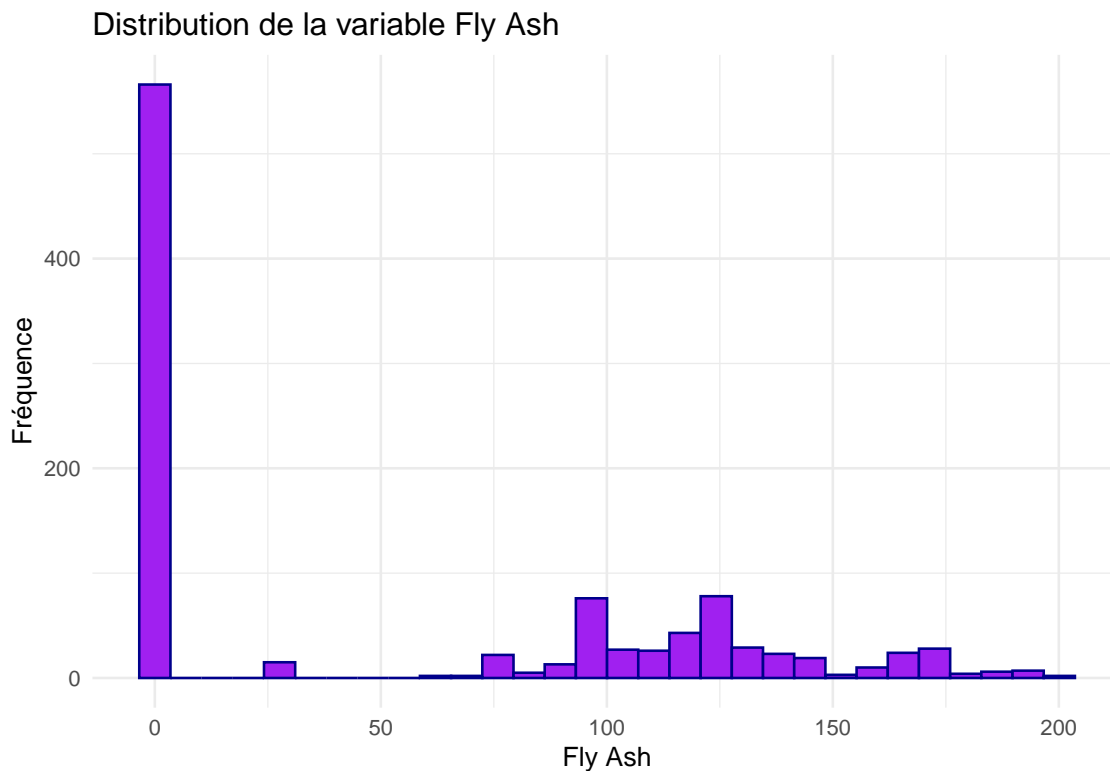
mulations de béton nécessitent des proportions de ciment variées, influençant potentiellement la résistance finale du béton.

#### 4.1.2 Blast Furnace Slag



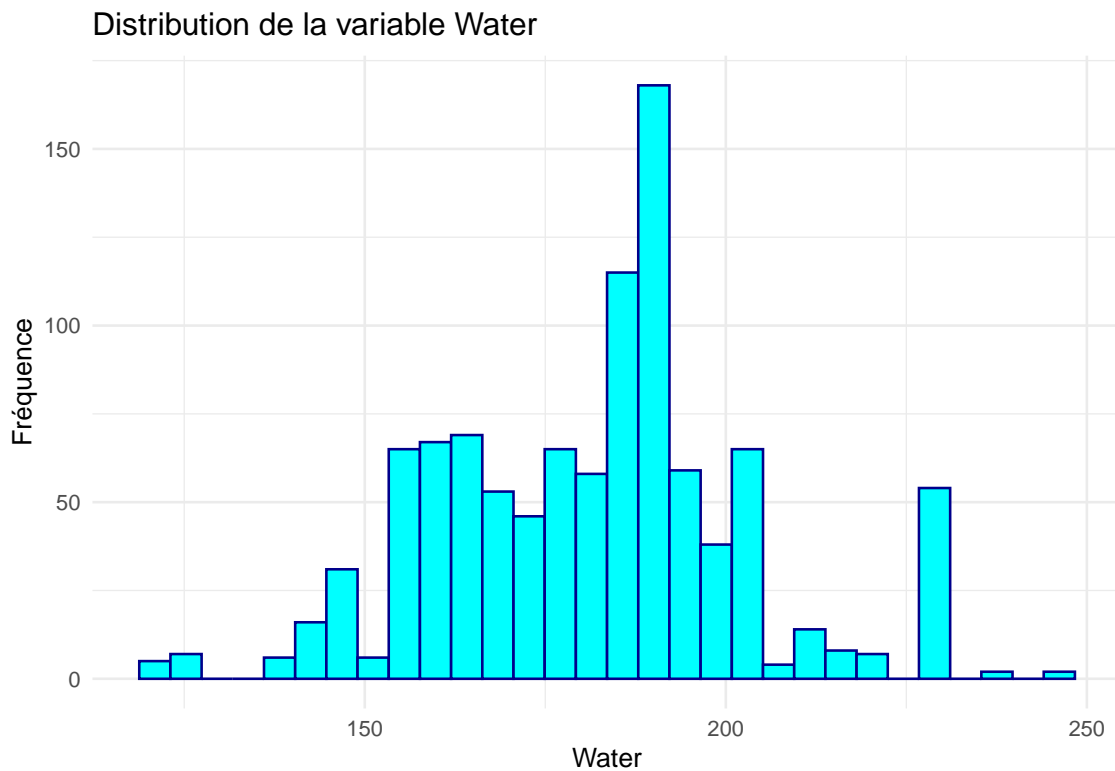
La majorité des valeurs sont proches de zéro, suggérant que le blast furnace slag est rarement utilisé en grande quantité. Cependant, quelques observations montrent une présence significative de ce composant, ce qui pourrait indiquer des formulations spécifiques cherchant à améliorer certaines propriétés du béton, comme la durabilité ou la résistance aux sulfates.

### 4.1.3 Fly Ash



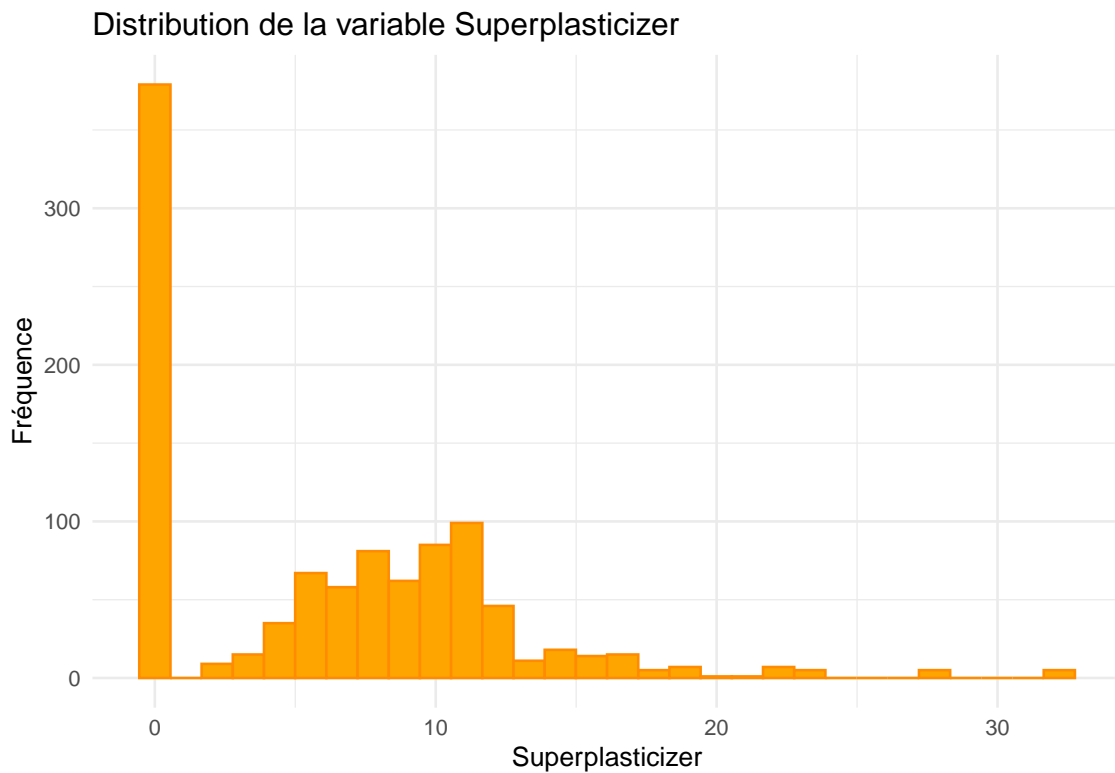
La distribution est fortement biaisée vers zéro, montrant que le fly ash est peu utilisé dans la majorité des échantillons. Seules quelques observations présentent des valeurs plus élevées, ce qui indique que cet additif est employé dans des mélanges spécifiques, probablement pour optimiser la maniabilité ou réduire les coûts en substituant une partie du ciment.

#### 4.1.4 Water



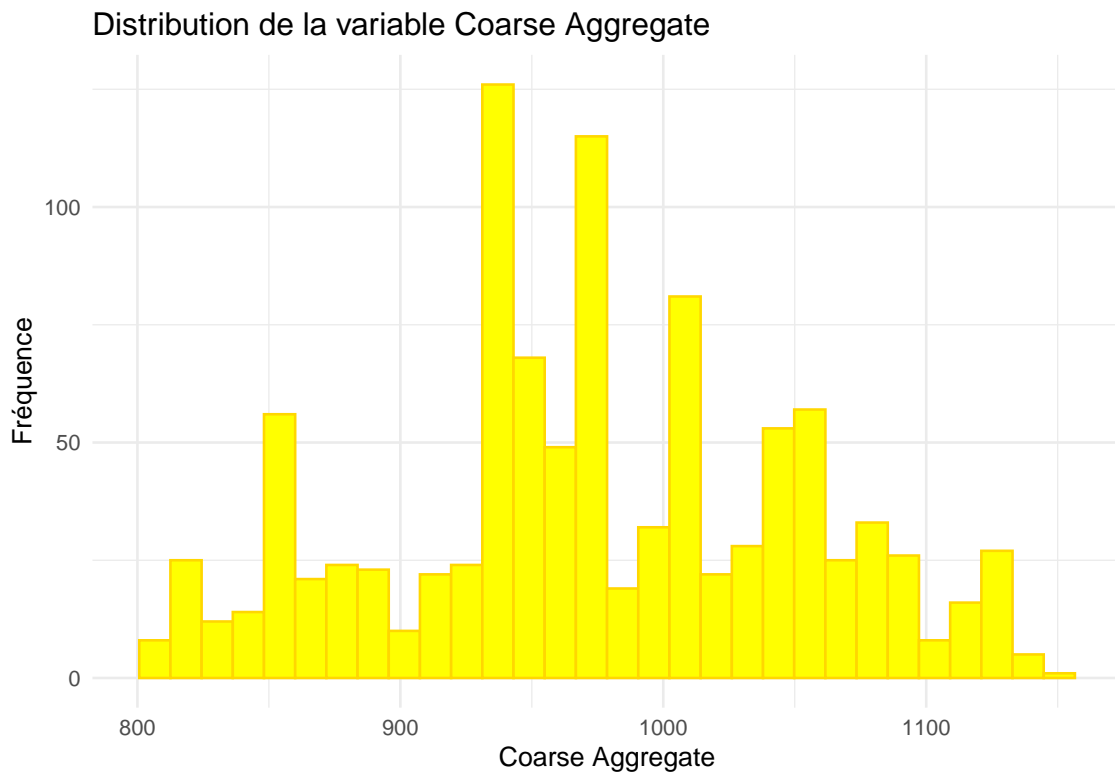
La répartition est relativement normale avec un pic autour de 200 kg/m<sup>3</sup>, ce qui indique une quantité d'eau standardisée dans la plupart des mélanges. Une consommation d'eau bien contrôlée est essentielle pour assurer une bonne hydratation du ciment et éviter une porosité excessive pouvant affaiblir le matériau.

#### 4.1.5 Superplasticizer



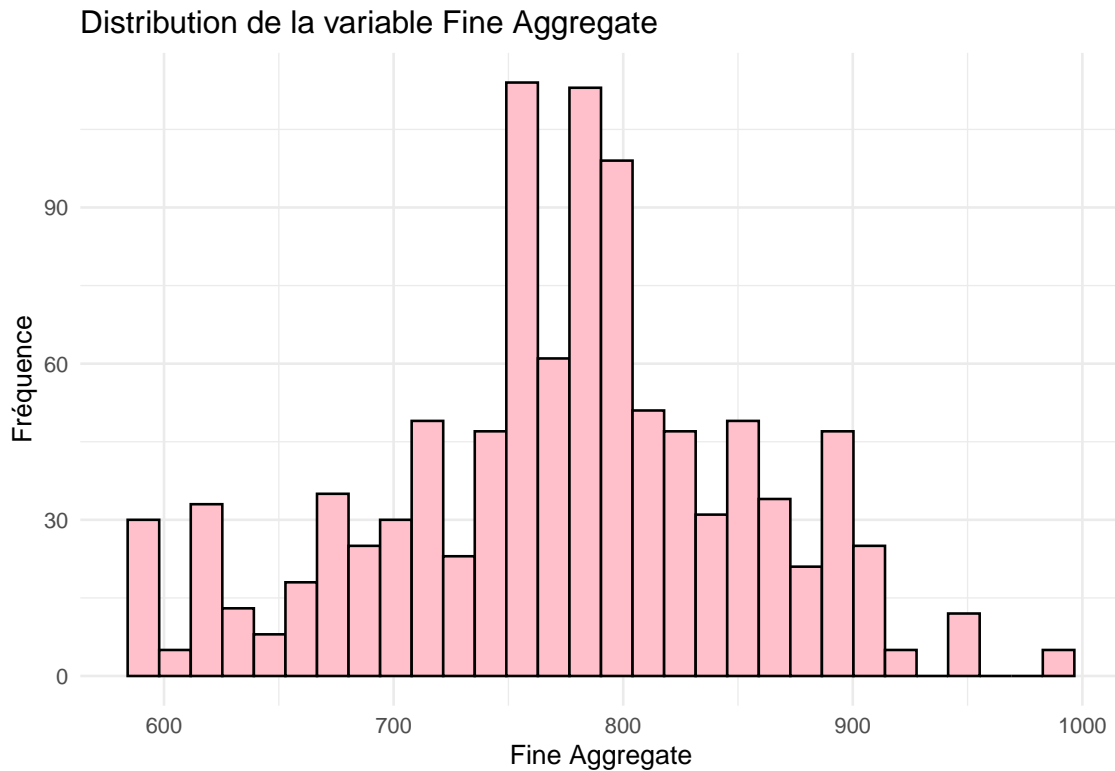
La plupart des valeurs sont proches de zéro, ce qui signifie que le superplasticizer est peu utilisé dans de nombreux échantillons. Cependant, certaines observations indiquent une utilisation plus importante, ce qui peut être lié à des mélanges nécessitant une meilleure fluidité tout en réduisant le rapport water/cement pour maximiser la résistance finale.

#### 4.1.6 Coarse Aggregate



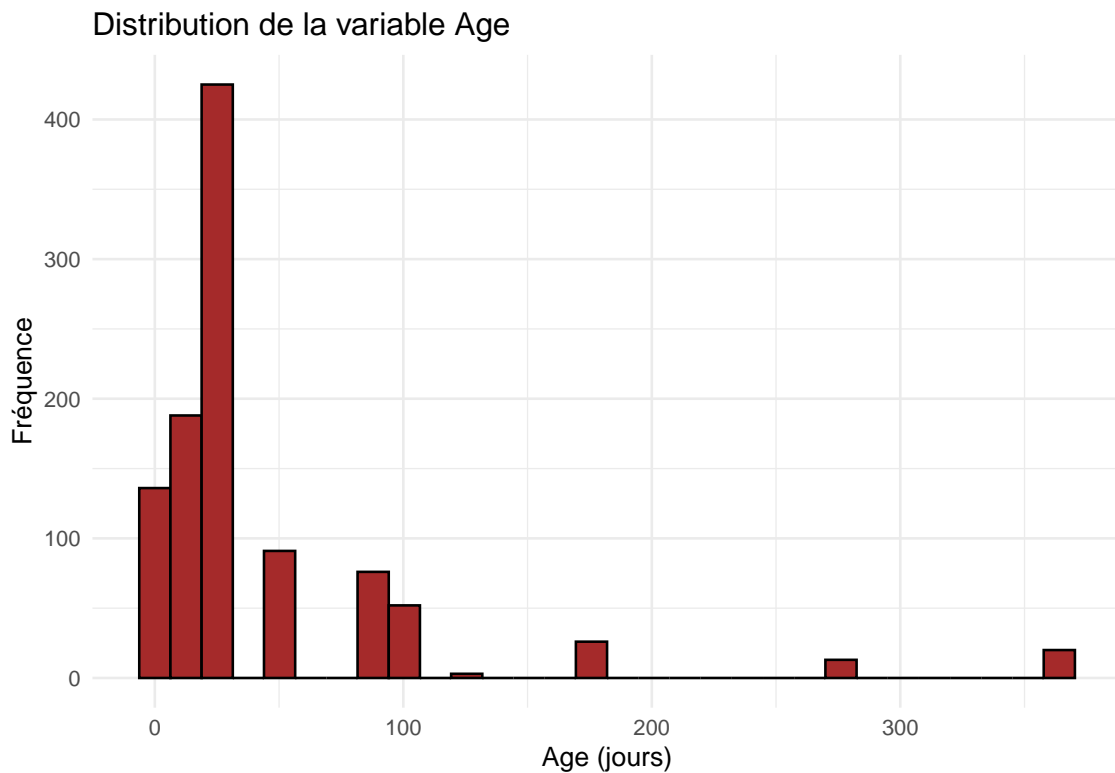
Le coarse aggregate suit une distribution centrée autour de 950-1000 kg/m<sup>3</sup>, ce qui montre une standardisation dans son utilisation. La présence d'une quantité relativement constante de coarse aggregate est essentielle pour garantir la stabilité et la résistance mécanique du béton tout en minimisant la fissuration.

#### 4.1.7 Fine Aggregate



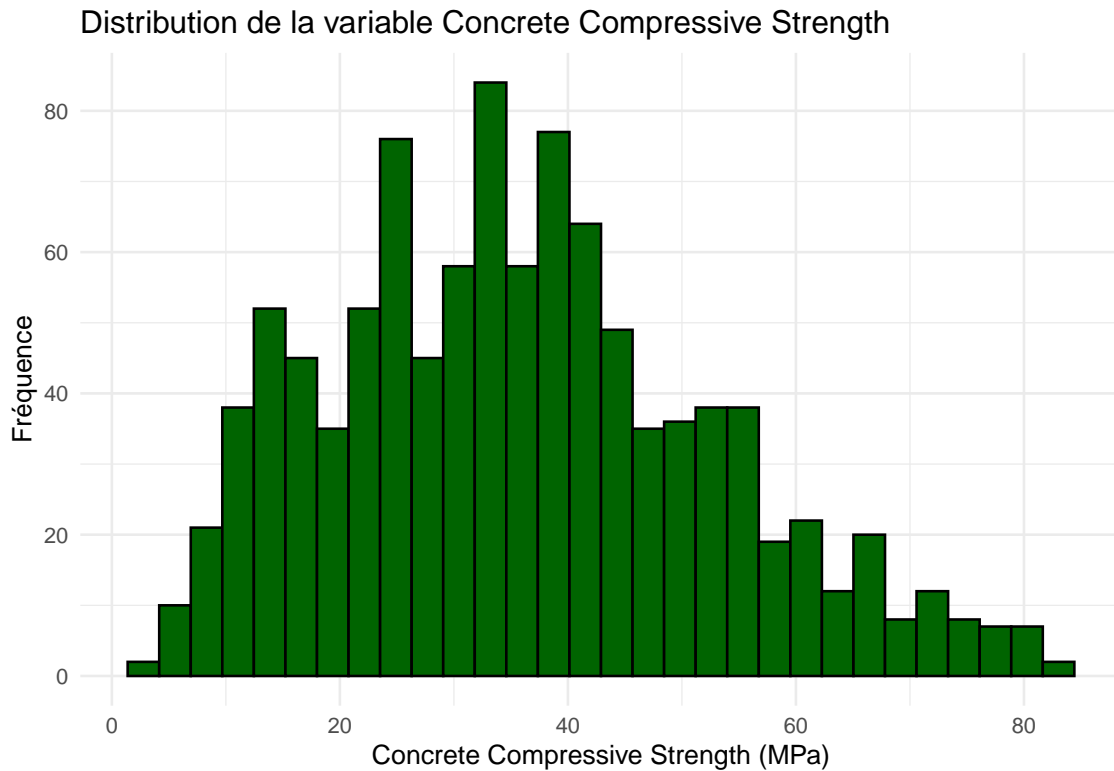
La distribution est également centrée autour de 750-800 kg/m<sup>3</sup>, suggérant une utilisation homogène du fine aggregate dans les différents mélanges de béton. Une quantité bien maîtrisée de fine aggregate améliore la compacité du béton et influe sur sa capacité de mise en œuvre et son imperméabilité.

#### 4.1.8 Age



L'âge du béton est majoritairement faible, avec une concentration autour de 28 jours. Quelques échantillons présentent des âges plus avancés, jusqu'à un an. Cela reflète l'importance de la période de cure du béton, puisque la résistance continue d'augmenter avec le temps grâce à l'hydratation prolongée du ciment.

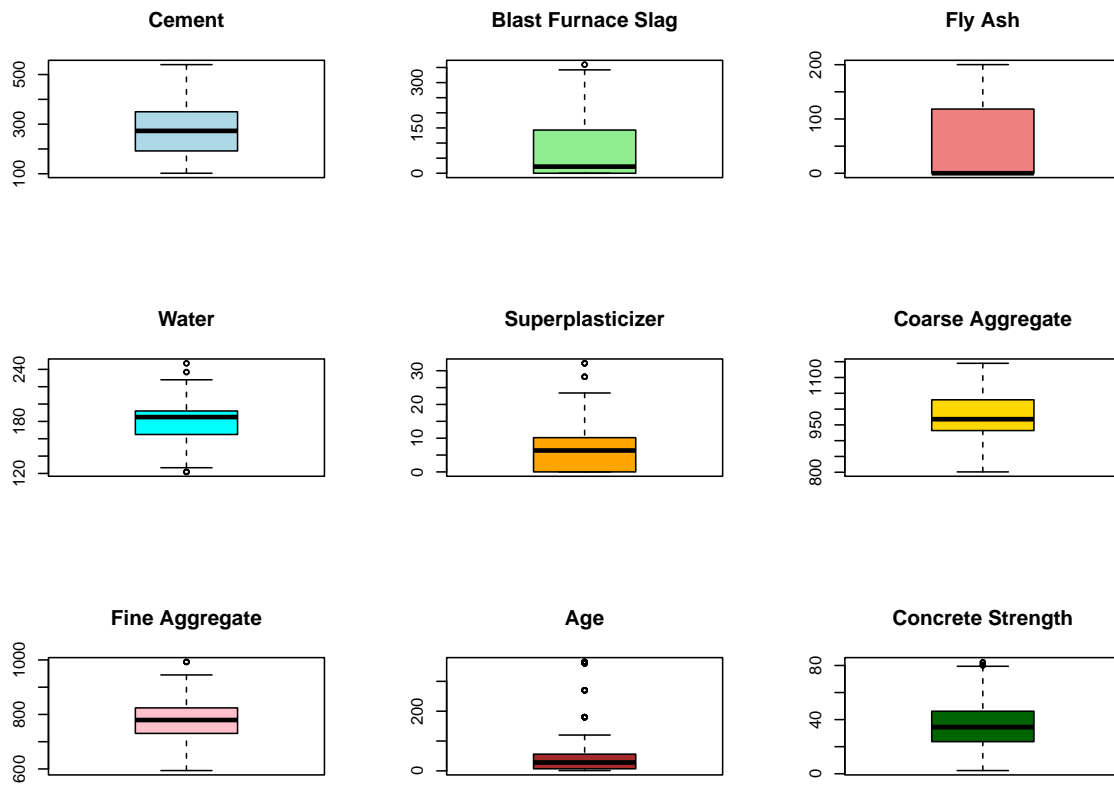
#### 4.1.9 Concrete Compressive Strength



La concrete compressive strength suit une distribution proche d'une loi normale, avec un pic autour de 40 MPa. Cela indique une variabilité contrôlée de la résistance du béton. La concrete compressive strength étant un paramètre clé, cette distribution montre que la plupart des mélanges atteignent une performance attendue, bien qu'il existe des échantillons avec des valeurs plus élevées ou plus faibles en fonction des formulations utilisées.



### 4.1.10 Fig 10 : Détection des valeurs aberrantes (Boxplots)



## 4.2 Analyse bivariee

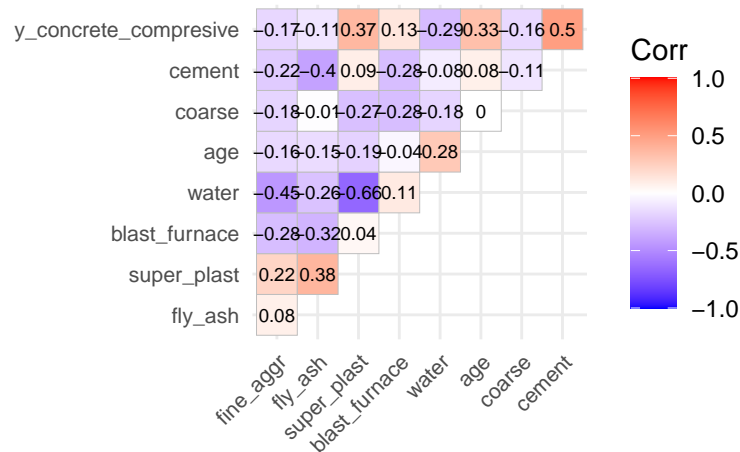
### 4.2.1 Matrice de corrélation

```
##          cement blast_furnace    fly_ash    water
## cement      1.00000000 -0.27519344 -0.397475440 -0.08154361
## blast_furnace -0.27519344  1.00000000 -0.323569468  0.10728594
## fly_ash      -0.39747544 -0.32356947  1.000000000 -0.25704400
## water        -0.08154361  0.10728594 -0.257043997  1.00000000
## super_plast   0.09277137  0.04337574  0.377339559 -0.65746444
## coarse       -0.10935604 -0.28399823 -0.009976788 -0.18231167
## fine_aggr    -0.22272017 -0.28159326  0.079076351 -0.45063498
## age          0.08194726 -0.04424580 -0.154370165  0.27760443
## y_concrete_compresive 0.49783272  0.13482445 -0.105753348 -0.28961348
##          super_plast    coarse    fine_aggr    age
## cement      0.09277137 -0.109356039 -0.22272017  0.081947264
## blast_furnace 0.04337574 -0.283998230 -0.28159326 -0.044245801
```

```

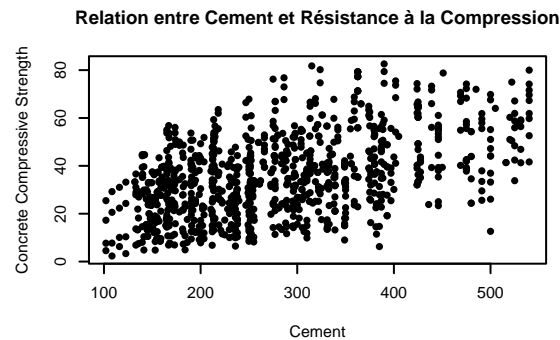
## fly_ash          0.37733956 -0.009976788  0.07907635 -0.154370165
## water           -0.65746444 -0.182311668 -0.45063498  0.277604429
## super_plast     1.00000000 -0.266302755  0.22250149 -0.192716518
## coarse          -0.26630276  1.000000000 -0.17850575 -0.003015507
## fine_aggr       0.22250149 -0.178505755  1.00000000 -0.156094049
## age             -0.19271652 -0.003015507 -0.15609405  1.000000000
## y_concrete_compresive 0.36610230 -0.164927821 -0.16724896  0.328876976
##
##                  y_concrete_compresive
## cement                0.4978327
## blast_furnace         0.1348244
## fly_ash               -0.1057533
## water                 -0.2896135
## super_plast           0.3661023
## coarse               -0.1649278
## fine_aggr            -0.1672490
## age                  0.3288770
## y_concrete_compresive 1.0000000

```



La matrice de corrélation en annexe va nous aider à voir les différents liens de corrélation potentiels entre nos variables. Dans cette partie, nous allons étudier et essayer de trouver les couples de variables significativement dépendants. Il est important de trouver tous les couples liés afin d'expliquer de manière optimale comment les variables de notre base peuvent influencer sur la résistance à la compression du béton.

### 4.2.2 Quantité de ciment et résistance du béton.

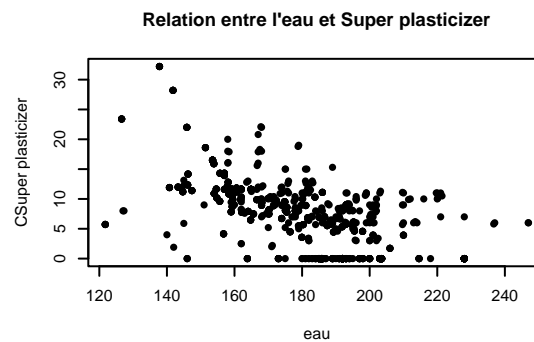


Nous allons ici examiner si la quantité de ciment utilisée pour créer le béton est significativement reliée à la résistance à la compression du béton. En effet, nous pouvons d'abord observer la distribution de ces deux variables à l'aide du graphique ci-dessous :

Cette visualisation des données nous donne une première impression du lien potentiel entre ces deux variables. En étudiant plus précisément ces deux variables avec un test de corrélation de Pearson, on se rend compte qu'il y a un lien significatif : ces deux variables sont positivement liées.

On peut alors dire que plus la quantité de ciment utilisée pour la fabrication du béton est grande, plus la résistance à la compression de ce béton sera importante. Le test final a donné un coefficient de corrélation  $\rho$  de Pearson égal à **0,5** (annexe), ce qui indique un lien relativement fort et non négligeable. Ainsi, une fabrication de béton comprenant une grande quantité de ciment favoriserait grandement sa résistance à la compression.

### 4.2.3 Additif chimique et eau.



Dans le processus de fabrication du béton, il est parfois nécessaire d'ajouter un additif chimique pour améliorer sa fluidité et sa maniabilité. La réflexion porte ici sur la corrélation entre l'ajout de cet additif chimique et la quantité d'eau utilisée pour la fabrication : y aurait-il un quelconque lien entre ces deux variables ?

Pour ce faire, il semble juste d'examiner les différents ajouts d'additif en fonction de la quantité d'eau utilisée pour chaque béton. Le graphique ci-dessous illustre cette relation, et nous pouvons aisément supposer qu'il existe une corrélation entre ces deux variables. En effet, dans l'ensemble, on remarque une diminution de l'ajout d'additif lorsque la quantité d'eau augmente.

Pour avoir une certitude, nous effectuons alors un test de corrélation de Kendall, qui nous indique, premièrement, qu'il existe un lien significatif entre ces deux variables et, deuxièmement, que le coefficient de corrélation  $\tau$  étant de **-0.53** (annexe), la négativité de la liaison est prouvée.

En clair, plus la quantité d'eau utilisée pour la fabrication du béton est importante, moins d'additif chimique a été ajouté lors de cette fabrication. Ce lien fort nous aidera dans la suite de notre étude.

#### 4.2.4 L'âge, une variable importante.



Il serait tout à fait naturel de penser que l'âge ait une quelconque importance sur la résistance à la compression du béton. En effet, l'imaginaire collectif nous amène d'abord à penser que ce béton, en conséquence de l'âge, deviendrait de plus en plus fragile et, de ce fait, moins résistant à la compression. Notre hypothèse serait alors de dire que l'âge est négativement lié à la résistance à la compression du béton, *i.e.*, plus l'âge augmente, moins le béton est résistant.

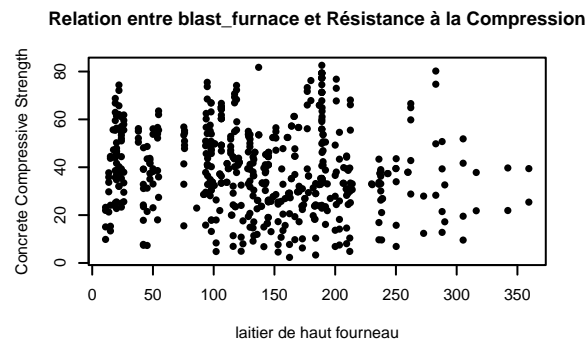
Dans un premier temps, l'observation de la distribution du temps par rapport à la compression nous aiderait à confirmer ou infirmer l'hypothèse de cette partie. Voyons le graphique ci-dessous.

Cette première visualisation ne nous permet pas de déterminer s'il existe réellement un lien entre ces deux variables. Une exagération pourrait peut-être nous amener à supposer l'existence d'une relation quadratique entre elles, mais cette hypothèse ne peut être vérifiée dans cette partie.

En conséquence, il est nécessaire de réaliser un test de corrélation.

Le test effectué nous permet de constater que notre hypothèse est infirmée, car il existe bien un lien significatif entre ces deux variables, et la corrélation indique que ce lien est positif. Cette corrélation est positive mais modérée, avec un coefficient de Kendall  $\tau$  de 0,449 (voir annexe). Les valeurs croissent significativement ensemble, bien que certaines exceptions subsistent.

#### 4.2.5 La résistance à la compression du béton peut-elle s'expliquer grâce au laitier de haut fournaux ?

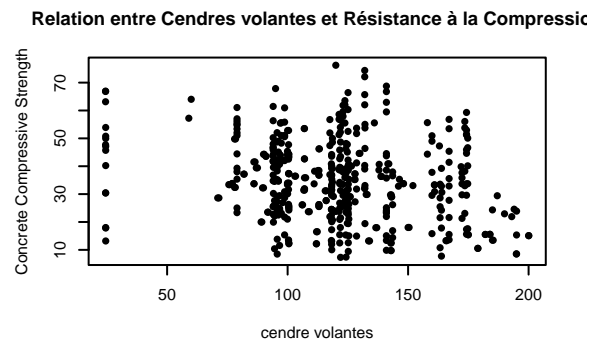


On examine ici la relation entre le laitier de haut fourneau et la résistance à la compression du béton :

Ce graphique ne suggère aucun lien de corrélation entre la quantité de laitier de haut fourneau et la résistance à la compression du béton, mis à part une légère baisse de la résistance du béton lorsque la quantité de laitier de haut fourneau augmente. Afin de détecter une potentielle relation de dépendance, il pourrait être judicieux d'effectuer un test de corrélation de Spearman.

En effet, ce test nous fournit une p-value inférieure à 0.05 et un coefficient  $\rho$  de 0.16. Ces valeurs indiquent qu'il existe bien un lien significatif positif entre ces deux variables, mais cette corrélation reste faible compte tenu de la proximité de  $\rho$  avec zéro.

#### 4.2.6 Les cendres volantes peuvent-elles expliquer la résistance du béton ?

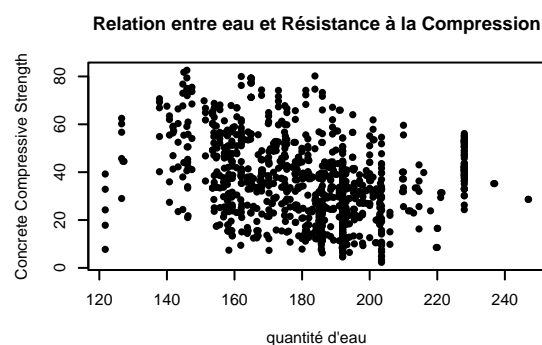


Pour tenter de répondre à cette question, on peut d'abord examiner la distribution de la quantité de cendres volantes en fonction de la résistance à la compression du béton.

La représentation de la distribution de notre variable à expliquer par rapport à la variable explicative suggère une légère corrélation linéaire négative. On pourrait alors en déduire que plus la quantité de cendres volantes augmente, moins le béton résiste à la compression.

Pour s'assurer de la justesse de cette analyse, nous réalisons un test de corrélation de Spearman, qui met en évidence un lien significatif entre ces deux variables. Comme indiqué dans notre interprétation, ce lien est négatif, avec un coefficient  $\rho$  de -0.19, indiquant ainsi une corrélation faible mais non négligeable.

#### 4.2.7 L'eau, un facteur de résistance ?



Dans l'imaginaire collectif, il est tout à fait logique de penser que si l'on ajoute de l'eau lors de la création du béton, celui-ci sera moins résistant que si l'on ne le

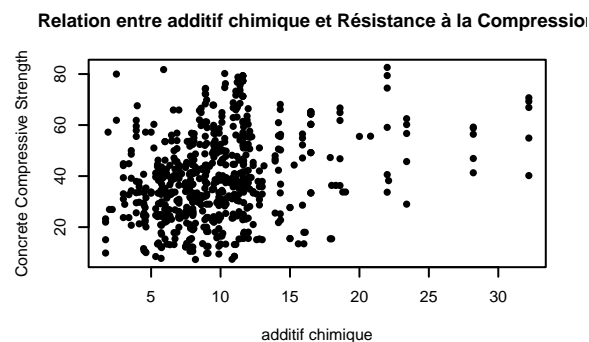
fait pas. Cependant, il est indispensable d'incorporer une certaine quantité d'eau. L'objectif est donc de trouver un juste milieu dans cette quantité. Examinons d'abord la distribution de la résistance des bétons par rapport à leur résistance à la compression :

Comme dans la partie précédente, le graphique est relativement désordonné, mais on peut tout de même y relever une légère dépendance négative, comme mentionné dans l'introduction de l'étude de cette dépendance. Pour être sûr de nos analyses, il serait nécessaire d'effectuer un test de corrélation de Spearman.

Le résultat de ce test indique qu'il existe bien une corrélation avec un coefficient de corrélation de  $-0.31$ , ce qui montre un lien négatif non négligeable entre ces deux variables.

En conclusion, moins il y a d'eau dans la création du béton, plus celui-ci sera résistant. Cependant, les points situés à l'extrémité gauche du graphique suggèrent qu'un minimum d'eau est tout de même nécessaire pour que le béton atteigne une résistance suffisante.

#### 4.2.8 Superplastificité et Résistance

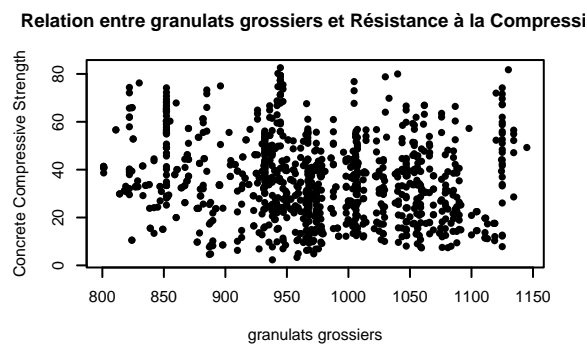


Le superplastifiant, utilisé dans les mélanges de béton, joue un rôle essentiel en améliorant la fluidité du béton sans ajout d'eau supplémentaire. Ce composant chimique permet d'augmenter la densité du béton, contribuant ainsi à une meilleure résistance à la compression.

Visuellement, le graphique ne permet pas de tirer de conclusions claires sur la relation entre la quantité de superplastifiant et la résistance à la compression du béton. Bien que la tendance semble indiquer une augmentation initiale de la résistance avec l'ajout de superplastifiant, la distribution des points reste dispersée. De plus, la présence de nombreux points proches du début de la courbe rend difficile l'identification d'une tendance nette.

Le test de Kendall révèle une corrélation positive modérée ( $\tau = 0.249$ ) entre la quantité de superplastifiant et la résistance à la compression du béton. Cette corrélation est significative, comme en témoigne la p-value extrêmement faible, bien inférieure à 0.05. Cela suggère que, de manière générale, une augmentation de la quantité de superplastifiant dans le mélange de béton pourrait être associée à une amélioration de sa résistance à la compression.

#### 4.2.9 Granulats grossiers et résistance à la compression



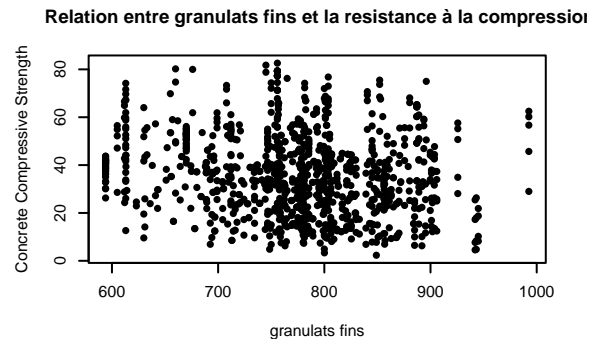
Les granulats grossiers, qui comprennent des particules de grande taille comme des graviers ou des pierres concassées, contribuent à la résistance à la compression du béton en réduisant les vides dans le mélange, augmentant ainsi sa compacité.

Le graphique représente la relation entre les granulats grossiers et la résistance à la compression du béton. La répartition des données semble dispersée. On peut néanmoins observer un lien quadratique entre ces deux variables, mais il est difficile d'en tirer une conclusion plus approfondie.

Le test de corrélation de Kendall a donné un coefficient  $\tau$  de -0.124. Bien que la p-value soit très faible ( $3.318e-09$ ), ce faible  $\tau$  indique une relation négative mais modérée. Cela signifie qu'il existe bien une certaine corrélation entre ces deux variables.



#### 4.2.10 Fig 20 : Relation entre granulats fins et la resistance à la compression



test correlation entre variable à expliquer et cement

```
cor.test(bdd$cement, bdd$y_concrete_compressive, method = "pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: bdd$cement and bdd$y_concrete_compressive
## t = 18.405, df = 1028, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.4504473 0.5424213
## sample estimates:
## cor
## 0.4978327
```

*#test de pearson car ont une distribution pseudo-normale*

test de correlation entre variable water et SuperPlasticizer

```
cor.test(bdd$water, bdd$super_plast, method = "kendall")
```

```
##
## Kendall's rank correlation tau
##
## data: bdd$water and bdd$super_plast
```

```
## z = -23.914, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.528651
```

*#test de Kendall car n'ont pas de distribution pseudo- normale*

test de correlation entre variable water et y

```
cor.test(bdd$age, bdd$y_concrete_compressive, method = "kendall")
```

```
##
## Kendall's rank correlation tau
##
## data: bdd$age and bdd$y_concrete_compressive
## z = 19.826, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.4490164
```

*#test de Kendall car n'ont pas de distribution pseudo- normale*

## 4.3 Ingénierie des données

La base de données contient **1030 observations et 9 variables numériques** initialement. Dans cette partie, nous préparons les données pour garantir leur qualité et leur pertinence. Cette étape comprend deux étapes principales :

- **Détection et traitement des valeurs aberrantes** à l'aide de boxplots.
- **Analyse des corrélations** pour éviter la redondance des variables.

### 4.3.1 Détection des valeurs aberrantes

Les valeurs aberrantes peuvent fausser l'analyse et influencer la modélisation. Les **boxplots** (voir Analyse univariée) permettent d'identifier des observations extrêmes.

Certaines variables comme **Cement**, **Fly Ash** et **Coarse Aggregate** ont une distribution homogène et sont conservées sans modification.

La variable **Water** présente quelques valeurs extrêmes. Plutôt que de les exclure, nous analysons leur impact.

La variable **Superplasticizer** contient des valeurs très éloignées du reste des observations. Pour éviter toute influence excessive, ces valeurs extrêmes sont exclues en utilisant la règle des  $1.5 \times \text{IQR}$  :

```
# Calcul des seuils avec la règle des 1.5*IQR pour Superplasticizer
Q1_sp <- quantile(bdd$super_plast, 0.25)
Q3_sp <- quantile(bdd$super_plast, 0.75)
IQR_sp <- Q3_sp - Q1_sp

seuil_sup_sp <- Q3_sp + 1.5 * IQR_sp

# Filtrer les valeurs aberrantes pour Superplasticizer
bdd_ing <- bdd %>% filter(super_plast <= seuil_sup_sp)
```

Concernant **Blast Furnace Slag**, quelques valeurs sont élevées mais restent plausibles. Elles sont conservées sans modification.

La variable **Age** présente des valeurs extrêmes qu'il faut filtrer. En utilisant la règle des  $1.5 \times \text{IQR}$ , nous identifions les observations anormalement élevées et les supprimons pour assurer une meilleure cohérence des données :

```
# Calcul des seuils avec la règle des 1.5*IQR pour Age
Q1_age <- quantile(bdd_ing$age, 0.25)
Q3_age <- quantile(bdd_ing$age, 0.75)
IQR_age <- Q3_age - Q1_age

seuil_sup_age <- Q3_age + 1.5 * IQR_age

# Filtrer les valeurs aberrantes pour Age
bdd_ing <- bdd_ing %>% filter(age <= seuil_sup_age)
bdd_final <- bdd_ing
```

#### 4.3.2 Analyse des corrélations

L'objectif est d'identifier les variables fortement corrélées entre elles pour éviter la **multicolinéarité** dans le modèle.

La **matrice de corrélation** (voir Analyse bivariée) permet d'observer les liens entre les différentes variables.

La variable cible **y\_concrete\_compressive** présente des corrélations notables avec certaines variables explicatives : - **cement (0.50)** : Forte corrélation positive, suggérant une relation directe entre la quantité de ciment et la résistance à la compression.

- **cater (-0.29)** : Corrélation négative, ce qui est cohérent avec l'hypothèse qu'un excès d'eau diminue la résistance du béton.

- **super\_plast (0.37)** : Corrélation indiquant que cet additif pourrait améliorer la résistance du béton.

- **age (0.33)** : Relation positive, confirmant que la résistance augmente avec le temps.

Certaines variables explicatives montrent également des **corrélations importantes entre elles**, ce qui peut poser problème : - **water et super\_plast (-0.65)** : Corrélation négative forte, ce qui est logique car le superplastifiant permet de réduire la quantité d'eau nécessaire.

- **blast\_furnace et fly\_ash (-0.32)** : Corrélation négative modérée, indiquant que ces composants sont parfois utilisés en substitution l'un de l'autre.

Nous devons **évaluer l'impact de ces relations dans le modèle**. Si deux variables sont trop fortement corrélées, l'une d'elles pourrait être supprimée pour éviter des problèmes de multicolinéarité.

Ces observations pourront nous aider lors de la sélection des variables dans la prochaine étape.

### 4.3.3 Synthèse de l'Ingénierie des Données

Après le traitement des valeurs aberrantes et l'analyse des corrélations, la base de données passe de **1030 à 961 observations**.

Les valeurs extrêmes des variables **Age** et **Superplasticizer** ont été supprimées en appliquant la règle des  $1.5 \times \text{IQR}$ , garantissant une meilleure cohérence des données.

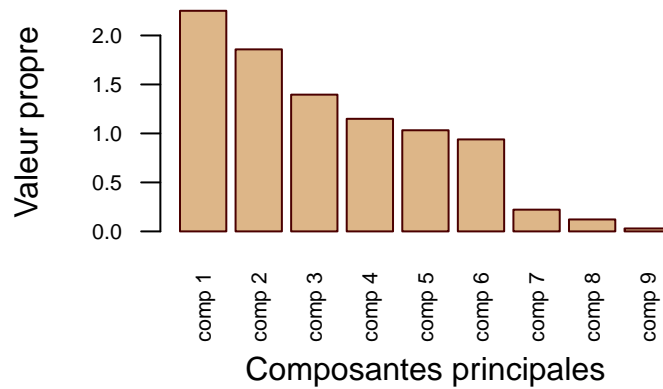
L'analyse des corrélations a révélé des relations notables entre plusieurs variables, notamment une forte corrélation entre **Water et Superplasticizer (-0.65)** et un lien important entre **Cement et la résistance à la compression (0.50)**.

## 4.4 Analyse multivariée

Dans cette partie, nous allons étudier l'ensemble des variables en utilisant une analyse en composantes principales (ACP). La variable à expliquer

(“y\_concrete\_compressive”) est également considérée comme active, aux côtés des huit autres variables explicatives.

## 4.5 Histogramme des valeurs propres

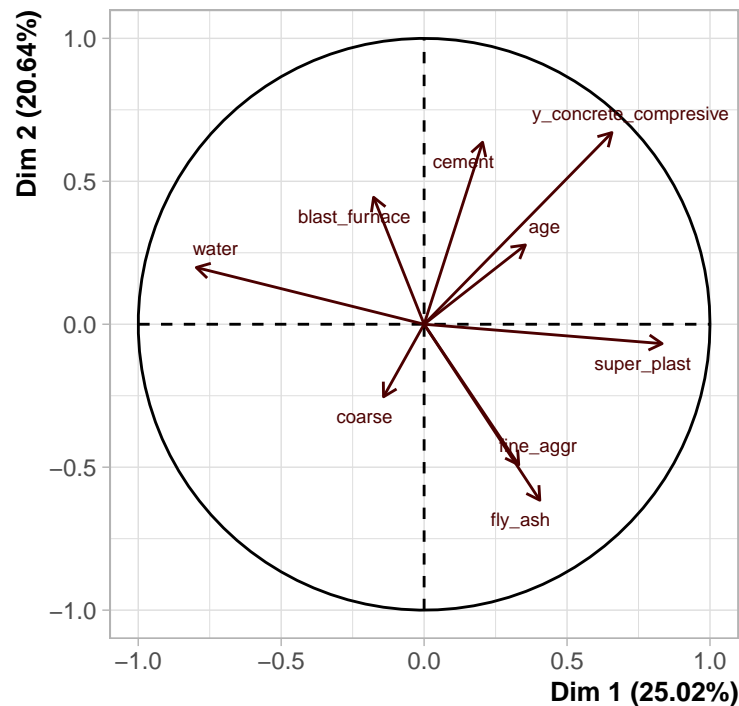


Le graphique illustre l’histogramme des valeurs propres issu de l’ACP, permettant d’identifier le nombre d’axes principaux à retenir.

- Le premier axe capte une part importante de la variance totale.
- Le deuxième axe conserve également une proportion significative d’information.
- Le troisième axe ajoute un complément notable de variance.
- À partir du quatrième axe, la contribution devient plus marginale.

En appliquant la **règle du coude**, il est pertinent de retenir les **trois premiers axes** pour l’analyse.

## 4.6 Analyse sur les 2 Premiers Axes



Le cercle de corrélation met en évidence les relations entre les variables étudiées, projetées sur les deux premiers axes principaux de l'ACP. Ces axes capturent ensemble **45,67 %** de la variance totale des données (**25,02 % pour Dim 1** et **20,64 % pour Dim 2**).

### 4.6.1 Interprétation du Premier Axe (Dim 1 : 25,02 %)

Le premier axe distingue **les formulations riches en eau (water)** de celles utilisant **des superplastifiants (super\_plast)**.

- **water** (28,20 % de contribution,  $\cos^2 = 0.62$ ) est la deuxième variable la plus influente sur Dim 1. Son orientation négative indique qu'une teneur élevée en eau tend à être associée à une baisse de la résistance à la compression.
- **super\_plast** (30,77 % de contribution,  $\cos^2 = 0.69$ ) est projetée dans la direction opposée à **water**, traduisant son rôle de réduction du besoin en eau, favorisant ainsi la résistance. c'est la variable la plus influente sur cet axe.
- **fly\_ash** (7,3 % de contribution,  $\cos^2 = 0.16$ ) et **fine\_aggr** (4,86% de contribution,  $\cos^2 = 0.11$ ) ne sont pas bien projetées sur cet axe.

- **age** (5,57 % de contribution,  $\cos^2 = 0.13$ ) est aussi bien représenté sur Dim 1, confirmant que la résistance du béton augmente avec le temps de durcissement.
- **y\_concrete\_compressive** (19,17 % de contribution,  $\cos^2 = 0.43$ ) est bien projetée sur Dim 1 ainsi que son  $\cos^2$  indique qu'elle est très bien représentée sur cet axe.

#### 4.6.2 Interprétation du Deuxième Axe (Dim 2 : 21,52 %)

Le deuxième axe distingue les **formulations riches en ciment (cement) et associées à une plus grande résistance (y\_concrete\_compressive)** des formulations utilisant **une proportion plus élevée des cendres volantes ('fly\_ash')** et les **granulats fins (fine\_aggr)**.

- **y\_concrete\_compressive** (24.70 % de contribution,  $\cos^2 = 0.45$ ) est la variable la plus influente sur Dim 2. Cela signifie que la résistance à la compression est fortement expliquée par cette dimension.
- **cement** (21.78 % de contribution,  $\cos^2 = 0.40$ ) est bien représentée sur Dim 2, indiquant que la quantité de ciment influence fortement la résistance mécanique. **fly\_ash** (20.39% de contribution et  $\cos^2 = 0.38$ ) et **fine\_aggr** (13.01% de contribution et  $\cos^2 = 0.24$ ) contribuent plus sur cet axe que sur le premier et elles sont très bien représentées sur cet axe contrairement au premier axe, suggérant que ces composants influencent la résistance du béton de la même façon puisque les flèches de ces deux variables sont confondues.
- **coarse** (5.05 % de contribution,  $\cos^2 = 0.10$ ) et **blast\_furnace** (10.61% de contribution et un  $\cos^2 = 0.10$ ) ne sont pas bien projetées sur cet axe et contribuent faiblement à la construction de cette dimension. Elles pourraient être expliquées par le troisième axe.

#### 4.6.3 Contribution des variables

```
# Tableau des contributions des variables
par(cex = 0.65)
contrib_var <- as.data.frame(res_pca$var$contrib)
print(contrib_var)
```

| ##        | Dim.1     | Dim.2      | Dim.3      | Dim.4      | Dim.5     |
|-----------|-----------|------------|------------|------------|-----------|
| ## cement | 1.8570924 | 21.8162521 | 19.4630539 | 17.0024498 | 0.3034758 |

```
## blast_furnace      1.3854568 10.6144114 38.1264598  0.8844007  4.7645646
## fly_ash            7.2754889 20.3973975  1.3084478  9.8815857  4.0570219
## water             28.2000179  2.1202816  1.4847945  2.0723167  4.4435413
## super_plast       30.7756668  0.2472336  6.2022982  0.9015953  8.4701606
## coarse            0.8939835  3.4632950 32.5054492 12.5162699 10.6005496
## fine_aggr         4.8648268 13.0123250  0.2499000 13.1394718 39.0454823
## age               5.5749786  4.1403732  0.4309223 40.7207759 28.1887145
## y_concrete_compresive 19.1724883 24.1884306  0.2286743  2.8811343  0.1264894
```

#### 4.6.4 Tableau des $\cos^2$ des variables sur

```
# Tableau des cos² des variables
par(cex = 0.65)
cos2_var <- as.data.frame(res_pca$var$cos2)
print(cos2_var)
```

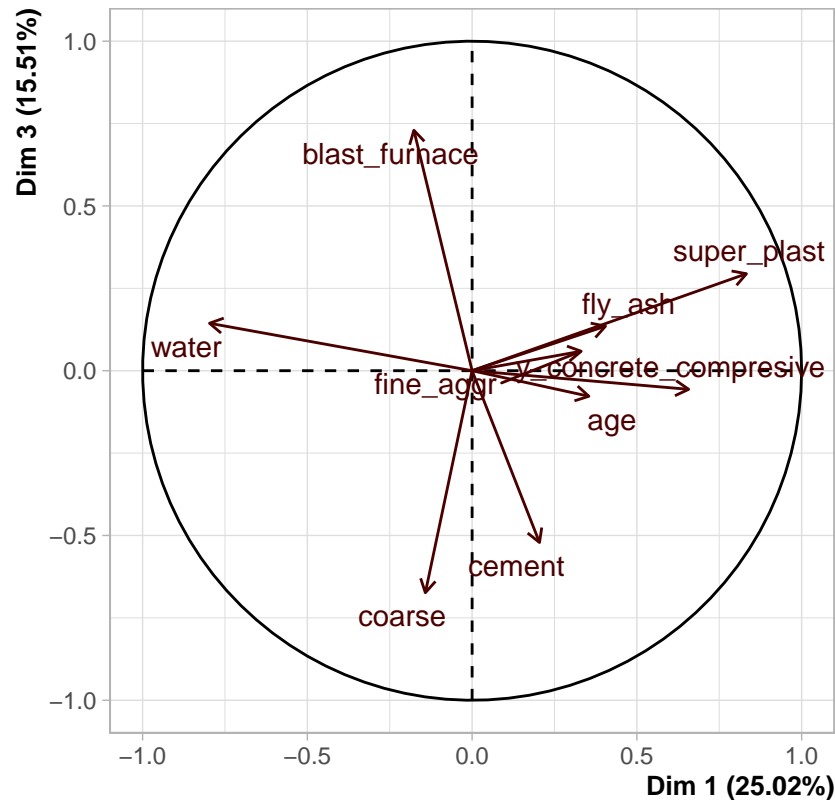
```
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## cement      0.04181872 0.405352873 0.271676732 0.19539670 0.003132659
## blast_furnace 0.03119825 0.197219125 0.532191509 0.01016377 0.049182696
## fly_ash      0.16383224 0.378990105 0.018264082 0.11356182 0.041879015
## water        0.63501880 0.039395503 0.020725633 0.02381562 0.045868900
## super_plast  0.69301824 0.004593679 0.086575320 0.01036137 0.087434082
## coarse       0.02013106 0.064349119 0.453730143 0.14384032 0.109425235
## fine_aggr    0.10954803 0.241773118 0.003488251 0.15100232 0.403050903
## age          0.12553950 0.076929445 0.006015067 0.46797404 0.290980829
## y_concrete_compresive 0.43173343 0.449428700 0.003191970 0.03311077 0.001305699
```

#### 4.6.5 Interprétation du Troisième Axe

```
plot.PCA(
  res_pca,
  axes = c(1, 3),           # On se concentre sur les axes 2 et 3
  choix = "var",            # Afficher les variables dans le plan factoriel
  col.var = "#4B0000",      # Couleur des variables alignée au style
  col.quant.sup = "#0000FF", # Couleur pour la variable quantitative supplém
  label = "all",
  title = "",
```



```
addgrid.col = "#DDB688"      # Couleur de la grille
)
```



Dim(1-3) : 40.5% % de la Variance Totale

Le troisième axe distingue les formulations riches en laitier du haut fourneau (**blast\_furnace**) et associées à une plus grande résistance (**y\_concrete\_compressive**) des formulations utilisant une proportion plus élevée des granulats grossiers (**Coarse**) et du ciment (**Cement**). ces deux axes représentent 40.5% de la variance totale.

-**blast\_furnace** (38.12% de contribution et  $\cos^2 = 0.53$ ) constitue la variable la plus influente et impacte positivement la résistance à la compression du béton. une hausse du laitier haut fourneau renforce la résistance.

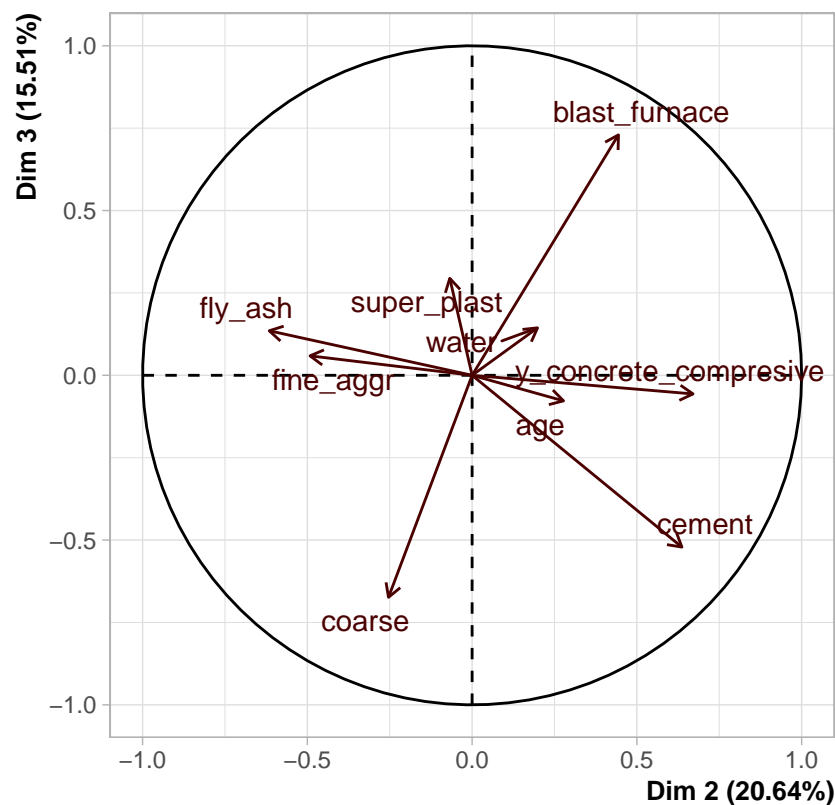
-**Coarse** (avec 32.5% de contribution et  $\cos^2 = 0.45$ ) représente la deuxième variable la plus importante et contribue en grande partie à la construction de cet axe. elle est projetée dans la direction opposée à **blast\_furnace** et **y\_concrete\_compressive**, suggérant que les formulations contenant une proportion élevée de granulats grossiers sont associées à une résistance plus faible.

- Enfin, nous avons la variable **Cement** qui est également bien représentée ( $\cos^2 = 0.27$ ) et contribue à la hauteur de 19.46% à la construction de cet axe.

#### 4.6.6 Interpretation PCA sur la dimension 2-3

```
# Représentation des variables sur le plan factoriel (axes 1 et 3)
plot.PCA(
  res_pca,
  axes = c(2, 3),
  choix = "var",
  col.var = "#4B0000",
  col.quant.sup = "#0000FF",
  label = "all",
  title = "",
  addgrid.col = "#DDB688"
)
```

*# On se concentre sur les 2 premiers axes*  
*# Afficher les variables dans le plan factoriel*  
*# Couleur des variables alignée au style*  
*# Couleur pour la variable quantitative supplémentaire*  
*# Couleur de la grille*



Dim(2-3) : 36.14 % de la Variance Totale (Voir annexe : Partie 22)

Le troisième axe distingue **les formulations riches en laitier du haut fourneau (blast\_furnace) et associées à une plus grande résistance (y\_concrete\_compressive)** des formulations utilisant **une proportion plus élevée des granulats grossiers (Coarse) et du ciment (Cement)**. ces deux axes représentent 40.5% de la variance totale.

nous remarquons la même conclusion que l'analyse des axes 1-3. Retenons que la variable **age (age)** n'est pas bien représentée sur les trois premiers axes mais on remarque qu'elle est très bien représentée sur le quatrième axe avec 40.72% de contribution et 0.46 comme qualité de représentation. compte tenu de la mauvaise représentation de la variable cible sur cet axe, nous pouvons conclure que l'âge n'influence pas de manière significative la résistance à la compression du béton.

Cette ACP met en évidence trois dimensions principales dans la composition du béton :

1. **Axe 1 (Dim 1)** : Opposition entre l'eau et les superplastifiants.
2. **Axe 2 (Dim 2)** : Opposition entre ciment et cendres volantes, granulats fins.
3. **Axe 3 (Dim 3)** : Opposition entre laitier du haut fourneau et granulats grossiers, ciment

## 4.7 Implications pour la Modélisation de la Résistance du Béton

- L'opposition entre **water** et **super\_plast** sur Dim 1 indique qu'un **terme d'interaction** entre ces deux variables pourrait être testé dans le modèle linéaire.
- **fine\_aggr** et **fly\_ash** étant bien projetées sur Dim 2, leurs interactions avec **cement** pourraient être pertinentes pour examiner leurs effets combinés sur la compacité et la résistance.
- **age** étant bien représentée sur Dim 4, une interaction avec **Cement** pourrait être étudiée pour voir comment l'évolution du béton avec le temps est influencée par le ciment.
- **y\_concrete\_compressive** étant principalement projeté sur Dim 2, un modèle linéaire devrait inclure **cement, blast\_furnace, fly\_ash, fine\_aggr** comme facteurs explicatifs car d'après nos analyses, ce sont les variables susceptibles d'influer sur la résistance à la compression.
- L'opposition entre **coarse** et **super\_blast** sur Dim 2 suggère qu'un effet combiné de ces deux variables pourrait influencer la résistance mécanique.

## 5 Modélisation linéaire

Dans cette section, nous nous concentrons sur l'application de la régression linéaire pour modéliser la relation entre la résistance à la compression du béton et ses différentes variables explicatives. Cette approche permet de comprendre l'impact de chaque variable exogène sur la variable cible, d'identifier les relations significatives ainsi que d'évaluer leur pertinence statistique et les hypothèses qui en découlent.

### 5.1 Résultats de la Régression

**Table 1:** Tableau des résultats de la régression linéaire

| Coefficient   | Estimate   | Std_Error  | t_value | Pr       |
|---------------|------------|------------|---------|----------|
| (Intercept)   | 13.5338116 | 21.7514794 | 0.622   | 0.534    |
| cement        | 0.1104577  | 0.0069542  | 15.884  | < 2e-16  |
| blast_furnace | 0.0840965  | 0.0084035  | 10.007  | < 2e-16  |
| fly_ash       | 0.0571293  | 0.0104619  | 5.461   | 6.05e-08 |
| water         | -0.1748622 | 0.0323918  | -5.398  | 8.49e-08 |
| super_plast   | 0.3208792  | 0.0814200  | 3.941   | 8.71e-05 |
| coarse        | 0.0012942  | 0.0077510  | 0.167   | 0.867    |
| fine_aggr     | -0.0005301 | 0.0088325  | -0.060  | 0.952    |
| age           | 0.3112546  | 0.0093762  | 33.196  | < 2e-16  |

### 5.2 Interprétations de la régression

#### Interprétation des p-values

Les résultats de la régression montrent que :

- **Intercept** : La p-value est de **0.534**, ce qui est **supérieur à 0.05**. Nous retenons donc l'**hypothèse H0**, selon laquelle cette valeur n'est **pas significative**.
- **Les variables qui sont significatives ( $p < 0.05$ )** : cement, blast\_furnace, fly\_ash, water, super\_plast et age.

Nous acceptons l'**hypothèse H1**, qui indique que ces variables ont un **effet significatif** sur la résistance du béton.

- **Variables non significatives au seuil de 5 %** : `coarse` et `fine_aggr`. Elles ne sont pas significatives, même au seuil de 10 %, et n'ont donc **aucun effet statistique** sur la résistance du béton.

Ces résultats suggèrent que certaines variables ont un impact plus marqué sur la **résistance à la compression du béton**, tandis que d'autres jouent un rôle moins déterminant.

Au regard de ces variables significatives, nous avons constaté des effets divers sur la variable réponse.

Pour la variable **Cement** (ciment), on remarque un effet positif sur la variable réponse qui n'est rien d'autre que la résistance à la compression. La valeur de cette relation est de l'ordre de **0.110**, cela signifie que si le ciment augmente d'une unité alors la résistance à la compression augmente de **0.110 unité**.

Ensuite, pour la variable **blast\_furnace** (laitier de haut fourneau), la relation est positive. Cela signifie qu'une hausse de la variable en question d'une unité entraîne une augmentation de **0.084 unité** de la variable cible.

Puis, on s'intéresse à la variable **fly\_ash** (cendres volantes) qui est positivement corrélée avec la variable endogène (**0.057**). Cela pourrait être traduit comme étant une augmentation moins proportionnelle de la résistance comparée aux cendres volantes.

Par la suite, nous remarquons une relation négative entre la variable **water** et la résistance à la compression. Cette relation est de l'ordre de **-0.175**, c'est-à-dire que l'augmentation de l'eau d'une unité entraîne une baisse de la résistance de **0.175 unité**. Cela rejoint notre conclusion de l'analyse multivariée.

Par ailleurs, la relation de notre variable cible avec la variable **super\_plast** est de **0.321**, qui est la relation la plus importante dans notre modèle ! Cette relation est positive, ce qui signifie que **l'additif chimique est très important pour augmenter la résistance de notre béton**. Cela confirme nos propos suite à l'analyse multivariée.

Enfin, on s'aperçoit que la variable **age** est positivement corrélée avec la variable réponse avec une valeur de **0.311**, ce qui confirme les résultats des revues qui stipulent que la résistance augmente avec le temps.

### 5.3 Diagnostic initial du modèle sans interactions (avec toutes les variables)

#### Test de Durbin-Watson

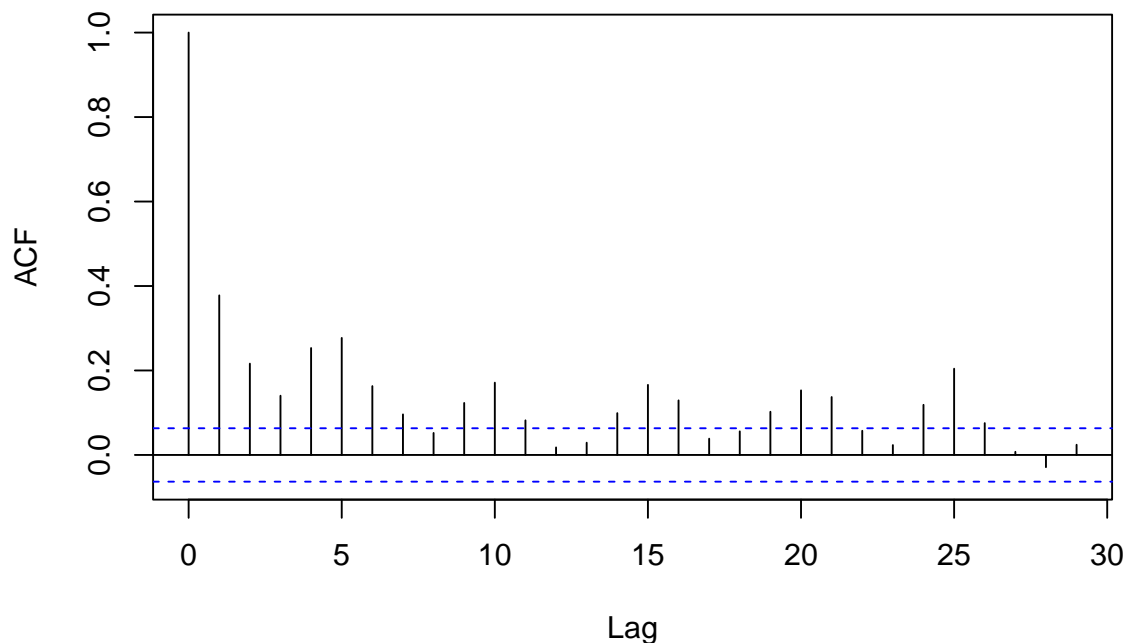
**Table 2:** Durbin-Watson Test Results

| Description            | Valeur                        |
|------------------------|-------------------------------|
| DW                     | 1.235                         |
| p-value                | $< 2.2e-16$                   |
| Alternative hypothesis | true autocorrelation is not 0 |

- **Durbin-Watson statistic (DW) : 1.235**
- **p-value :  $< 2.2e-16$**
- **Hypothèse alternative** : L'autocorrélation des erreurs est **différente de 0**.

La p-value étant inférieure à 0.05, nous rejetons l'**hypothèse H0** et acceptons l'**hypothèse H1**,  
ce qui signifie que **les erreurs sont fortement corrélées**.

### Autocorrélation des erreurs



On remarque que plusieurs barres sortent de l'intervalle de confiance, ce qui vient soutenir la conclusion  
du test de Durbin-Watson selon laquelle les erreurs sont **autocorrélées**.

## Test de Breusch-Pagan

**Table 3:** Breusch-Pagan Test Results

| Description                  | Valeur    |
|------------------------------|-----------|
| Breusch-Pagan statistic (BP) | 137.02    |
| Degrés de liberté (df)       | 8         |
| p-value                      | < 2.2e-16 |

- Breusch-Pagan statistic (BP) : **137.02**
- Degrés de liberté (df) : **8**
- p-value : **< 2.2e-16**

D'après les résultats du **test de Breusch-Pagan**, nous acceptons l'**hypothèse H1**,

ce qui signifie que **l'erreur est hétéroscédastique**. Cela indique qu'il existe une **dépendance**

entre la variance des erreurs et les **variables explicatives**.

## Test de Normalité de Shapiro-Wilk

**Table 4:** Shapiro-Wilk Normality Test Results

| Description                | Valeur    |
|----------------------------|-----------|
| Shapiro-Wilk statistic (W) | 0.99324   |
| p-value                    | 0.0002368 |

- Shapiro-Wilk statistic (W) : **0.99324**
- p-value : **0.0002368**

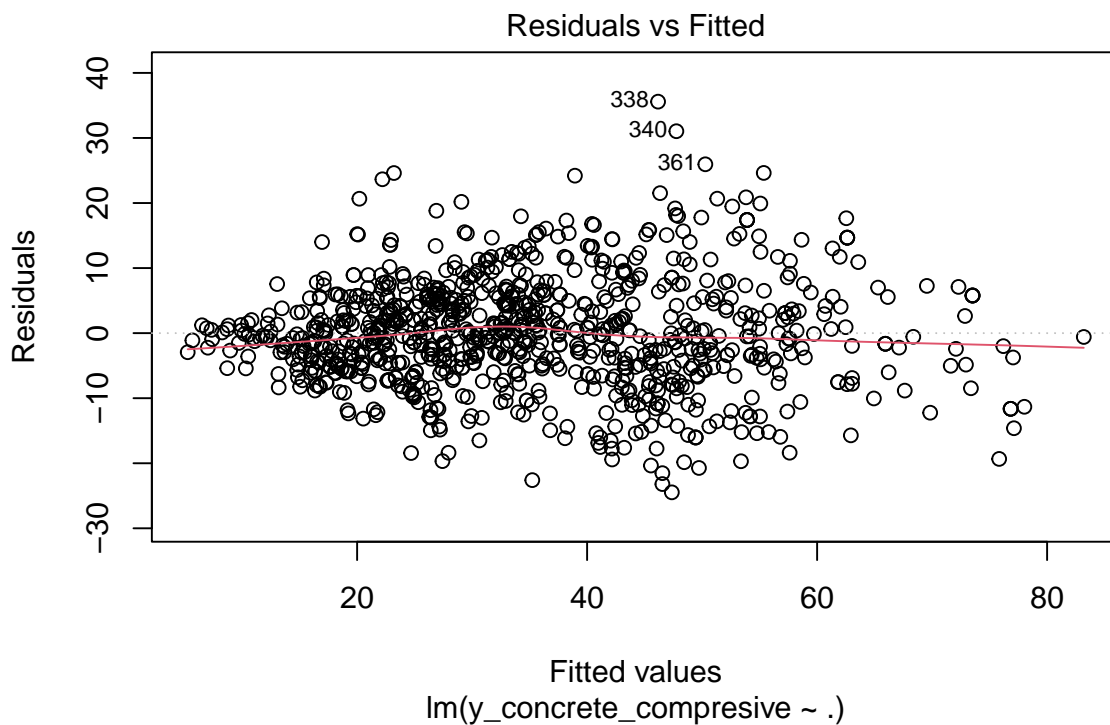
D'après les résultats du **test de Shapiro-Wilk**, la p-value est **inférieure au seuil de 5 %**.

Nous rejetons donc l'**hypothèse nulle (H0)** de normalité des erreurs.

Par conséquent, **les erreurs ne suivent pas une distribution normale**.

**Table 5:** A tibble: 6 x 6

| .fitted | .resid | .hat    | .sigma | .cooksd  | .std.resid |
|---------|--------|---------|--------|----------|------------|
| 53.5    | 26.50  | 0.01370 | 10.4   | 0.010200 | 2.570      |
| 53.7    | 8.14   | 0.01290 | 10.4   | 0.000902 | 0.788      |
| 56.8    | -16.50 | 0.01700 | 10.4   | 0.004950 | -1.600     |
| 67.7    | -26.60 | 0.02870 | 10.4   | 0.022100 | -2.600     |
| 60.9    | -16.60 | 0.03060 | 10.4   | 0.009230 | -1.620     |
| 26.9    | 20.20  | 0.00694 | 10.4   | 0.002940 | 1.950      |

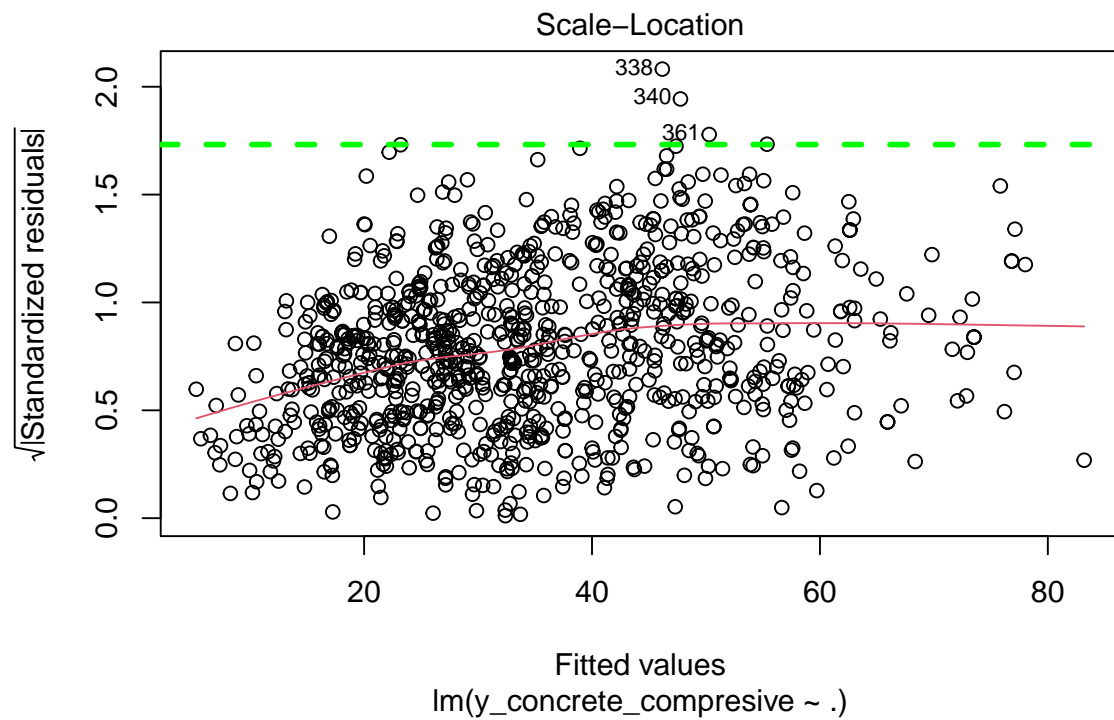


Le nuage de points montre une dispersion relativement homogène autour de zéro, mais la courbe rouge (loess) présente une légère forme incurvée, ce qui suggère une possible non-linéarité dans la relation entre les variables explicatives et la variable réponse.

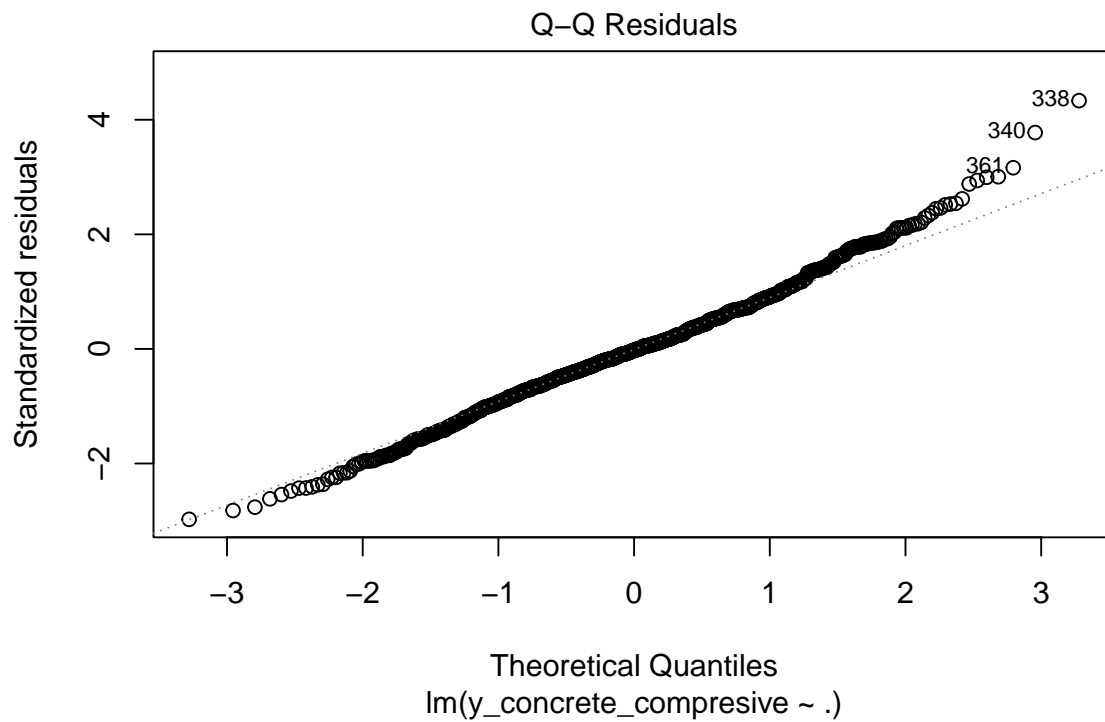
Il n'y a pas de motif très marqué, mais la présence de cette courbe légèrement incurvée pourrait indiquer que certaines transformations ou interactions pourraient améliorer le modèle.

Les observations 338, 340 et 361 présentent les résidus les plus élevés, mais ne semblent pas influentes ou aberrantes à ce stade.



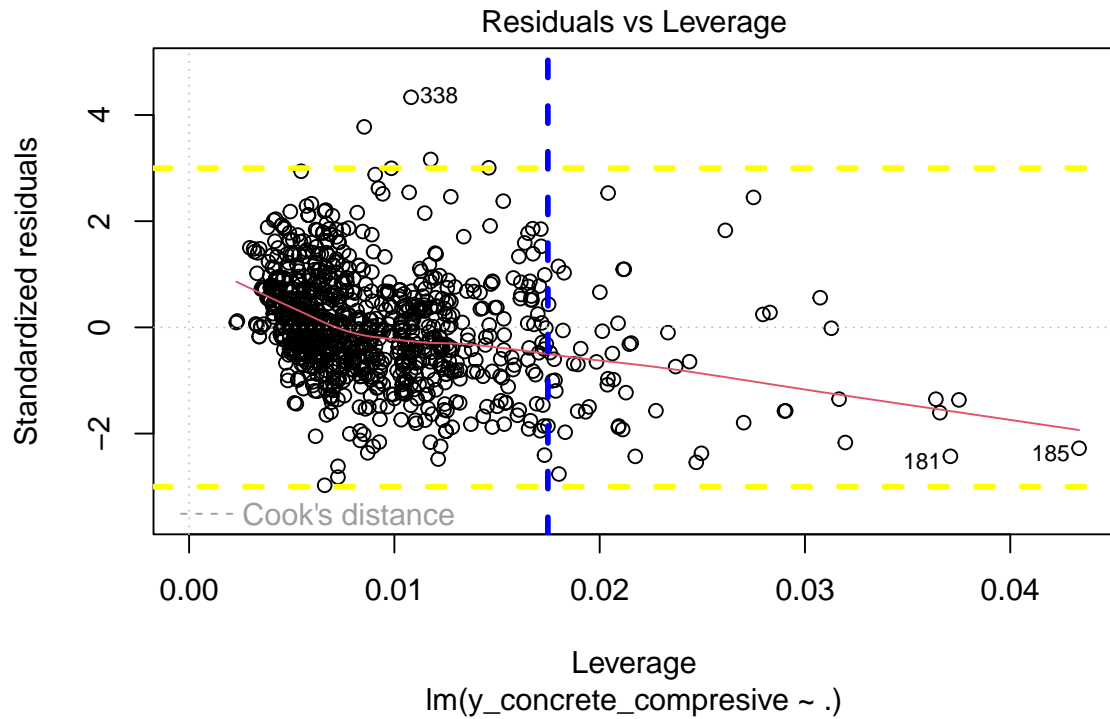


Dans notre cas, la **courbe rouge n'est pas horizontale** et ne sépare pas les points de manière homogène de part et d'autre. Cela indique que **les erreurs ne sont pas homogènes**, ce qui confirme la conclusion du **test de Breusch-Pagan**.



Le Q-Q plot montre que la majorité des résidus sont proches de la droite théorique, ce qui suggère une approximation raisonnable à la normalité. Cependant, quelques valeurs extrêmes dans la queue droite (observations 338, 340 et 361) s'en écartent légèrement.

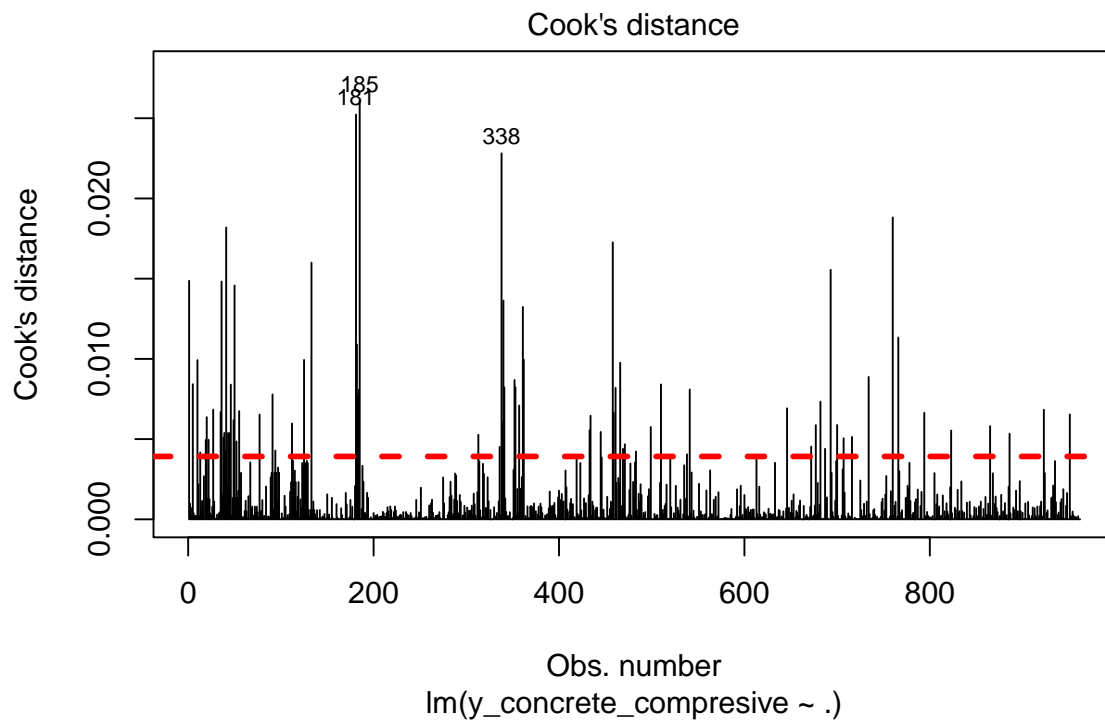
Bien que visuellement la normalité semble globalement respectée, le test de Shapiro-Wilk rejette l'hypothèse de normalité au seuil de 5 %, indiquant que les résidus ne suivent pas une loi normale.



Tous les points situés à droite de la droite verticale bleue sont des **points leviers extrêmes**,  
c'est-à-dire des **valeurs extrêmes** des variables explicatives.

De plus, tous les points en dehors de l'intervalle formé par les pointillés **jaunes** sont des **outliers**,  
c'est-à-dire des **valeurs extrêmes** de la variable cible.

Dans notre cas, nous avons **deux outliers** et **plusieurs points leviers extrêmes**.



Suite à l'observation des résultats donnés par le graphique, on remarque que **plusieurs points dépassent la ligne** formée par les **pointillés rouges**, qui représente le seuil à partir duquel un point est **considéré comme influent**.

Les **bétons 57, 225 et 611** sont les **plus influents**.

## 5.4 Selection du Meilleur modèle

L'objectif de cette partie est de choisir le modèle de regression linéaire qui prédise au mieux la variable que l'on veut expliquer, c'est à dire ici la résistance à la compression du béton. On peut alors regarder premièrement la selection du meilleur modèle en utilisant le critère BIC

Selon le critère aic, le premier modèle qui a été sélectionné par l'algorithme exhaustive est le suivant :  $y\_concrete\_compressive \sim 1 + cement + blast\_furnace + fly\_ash + water + super\_plast + age$ .

selon la deuxième méthode qui est l'algorithme génétique, on a :  $y\_concrete\_compressive \sim y\_concrete\_compressive \sim 1 + cement + blast\_furnace + fly\_ash + water + super\_plast + age$  les deux methodes nous donnent le meme modèle.

On peut alors voir que les variables **coarse** et **fine\_aggr** on été retirées du modèle par manque de significativité et impact sur la valeur de l'erreur empirique.

On peut se demander si l'utilisation d'un autre critère pour sélectionner le meilleur modèle peut changer le modèle optimal obtenu précédemment. Essayons le critère AIC.

La encore, le même modèle `cement + blast_furnace + fly_ash + water + super_plast + age` ressort, ce qui était assez prévisible car ces deux critères sont très similaires.

Selon le critère aic, le premier modèle qui a été sélectionné par l'algorithme exhaustive est le suivant : `y_concrete_compressive ~ 1 + cement + blast_furnace + fly_ash + water + super_plast`

selon la deuxième méthode qui est l'algorithme génétique, on a : `y_concrete_compressive ~ 1 + cement + blast_furnace + fly_ash + water + super_plast + age` les deux méthodes nous donnent le même modèle.

```
## Start:  AIC=4065.85
## y_concrete_compressive ~ cement + blast_furnace + fly_ash + water +
##      super_plast + coarse + fine_aggr + age
##
##              Df Sum of Sq    RSS    AIC    F value    Pr(>F)
## - fine_aggr      1         0 64868 4063.9     0.0036    0.9522
## - coarse          1         2 64869 4063.9     0.0279    0.8674
## <none>                        64867 4065.9
## - super_plast     1       1058 65926 4079.4    15.5318 8.706e-05 ***
## - water           1       1986 66853 4092.8    29.1421 8.494e-08 ***
## - fly_ash         1       2032 66899 4093.5    29.8192 6.053e-08 ***
## - blast_furnace   1       6824 71691 4160.0   100.1476 < 2.2e-16 ***
## - cement          1      17190 82058 4289.8   252.2890 < 2.2e-16 ***
## - age             1      75088 139955 4802.8  1101.9937 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=4063.86
## y_concrete_compressive ~ cement + blast_furnace + fly_ash + water +
##      super_plast + coarse + age
##
##              Df Sum of Sq    RSS    AIC    F value    Pr(>F)
## - coarse          1         11 64878 4062.0     0.1557    0.6932
## <none>                        64868 4063.9
## + fine_aggr       1         0 64867 4065.9     0.0036    0.9522
## - super_plast     1       1082 65949 4077.8    15.8930 7.215e-05 ***
## - water           1       5245 70112 4136.6    77.0511 < 2.2e-16 ***
```

```
## - fly_ash      1      5304  70172 4137.4   77.9273 < 2.2e-16 ***
## - blast_furnace 1      26163  91030 4387.5  384.3664 < 2.2e-16 ***
## - cement       1      63929 128797 4721.0  939.2137 < 2.2e-16 ***
## - age          1      75321 140188 4802.4 1106.5715 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=4062.01
## y_concrete_compressive ~ cement + blast_furnace + fly_ash + water +
##      super_plast + age
##
##              Df Sum of Sq    RSS    AIC    F value    Pr(>F)
## <none>                64878 4062.0
## + coarse              1         11  64868 4063.9      0.1557    0.6932
## + fine_aggr           1          9  64869 4063.9      0.1314    0.7170
## - super_plast         1       1149  66027 4076.9     16.8912 4.299e-05 ***
## - fly_ash             1       5359  70237 4136.3     78.8040 < 2.2e-16 ***
## - water               1       6844  71722 4156.4    100.6397 < 2.2e-16 ***
## - blast_furnace       1      28681  93559 4411.8    421.7378 < 2.2e-16 ***
## - cement              1      68363 133241 4751.6   1005.2373 < 2.2e-16 ***
## - age                 1      75394 140272 4801.0   1108.6315 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Start:  AIC=5422.78
## y_concrete_compressive ~ 1
##
##              Df Sum of Sq    RSS    AIC    F value    Pr(>F)
## + age          1      72105 198590 5127.1   348.1968 < 2.2e-16 ***
## + cement       1      62942 207754 5170.5   290.5420 < 2.2e-16 ***
## + super_plast  1      44529 226166 5252.1   188.8142 < 2.2e-16 ***
## + water        1      34144 236551 5295.2   138.4248 < 2.2e-16 ***
## + coarse       1       7440 263256 5398.0    27.1011 2.362e-07 ***
## + blast_furnace 1       5874 264821 5403.7    21.2710 4.527e-06 ***
## + fine_aggr    1       5602 265093 5404.7    20.2670 7.563e-06 ***
## + fly_ash      1       1436 269259 5419.7      5.1145  0.02395 *
## <none>                270695 5422.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=5127.12
```

```
## y_concrete_compressive ~ age
##
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## + cement      1      67528 131063 4729.8 493.593 < 2.2e-16 ***
## + super_plast  1      38724 159867 4920.7 232.050 < 2.2e-16 ***
## + water        1      31044 167547 4965.8 177.502 < 2.2e-16 ***
## + coarse       1       8459 190132 5087.3  42.620 1.077e-10 ***
## + fine_aggr    1       7676 190914 5091.2  38.518 8.066e-10 ***
## + blast_furnace 1       7419 191171 5092.5  37.180 1.560e-09 ***
## + fly_ash      1       2732 195859 5115.8  13.363 0.0002706 ***
## <none>                                198590 5127.1
## - age          1      72105 270695 5422.8 348.197 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:   AIC=4729.75
## y_concrete_compressive ~ age + cement
##
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## + super_plast  1      33652  97410 4446.6 330.6147 < 2.2e-16 ***
## + blast_furnace 1      25164 105898 4526.9 227.4110 < 2.2e-16 ***
## + water        1      23078 107985 4545.6 204.5271 < 2.2e-16 ***
## + coarse       1       4325 126738 4699.5  32.6565 1.468e-08 ***
## + fly_ash      1       2651 128412 4712.1  19.7565 9.825e-06 ***
## + fine_aggr    1       1090 129973 4723.7   8.0269 0.004706 **
## <none>                                131063 4729.8
## - cement       1      67528 198590 5127.1 493.5929 < 2.2e-16 ***
## - age          1      76691 207754 5170.5 560.5706 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:   AIC=4446.58
## y_concrete_compressive ~ age + cement + super_plast
##
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## + blast_furnace 1      21785  75625 4205.3 275.3942 < 2.2e-16 ***
## + fine_aggr     1       4221  93190 4406.0  43.2971 7.737e-11 ***
## + water         1       2283  95127 4425.8  22.9430 1.933e-06 ***
## + fly_ash       1       1592  95819 4432.8  15.8797 7.263e-05 ***
## + coarse        1        511  96899 4443.5   5.0409 0.02498 *
```

```

## <none>                                97410 4446.6
## - super_plast      1      33652 131063 4729.8 330.6147 < 2.2e-16 ***
## - cement           1      62457 159867 4920.7 613.5993 < 2.2e-16 ***
## - age              1      70891 168301 4970.1 696.4630 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:   AIC=4205.31
## y_concrete_compressive ~ age + cement + super_plast + blast_furnace
##
##              Df Sum of Sq    RSS    AIC  F value    Pr(>F)
## + water        1      5388  70237 4136.3   73.2545 < 2.2e-16 ***
## + fly_ash       1      3903  71722 4156.4   51.9651 1.147e-12 ***
## + coarse        1       723  74902 4198.1    9.2160 0.002464 **
## <none>          75625 4205.3
## + fine_aggr     1        73  75552 4206.4    0.9179 0.338282
## -blast_furnace  1     21785  97410 4446.6  275.3942 < 2.2e-16 ***
## - super_plast   1     30273 105898 4526.9  382.6917 < 2.2e-16 ***
## - age           1     74591 150216 4862.8  942.9251 < 2.2e-16 ***
## - cement        1     78288 153913 4886.2  989.6569 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:   AIC=4136.29
## y_concrete_compressive ~ age + cement + super_plast + blast_furnace +
##   water
##
##              Df Sum of Sq    RSS    AIC  F value    Pr(>F)
## + fly_ash       1      5359  64878 4062.0   78.8040 < 2.2e-16 ***
## + fine_aggr     1      1497  68741 4117.6   20.7740 5.839e-06 ***
## <none>          70237 4136.3
## + coarse        1        66  70172 4137.4    0.8909 0.3455
## - water         1      5388  75625 4205.3   73.2545 < 2.2e-16 ***
## - super_plast   1      7738  77976 4234.7  105.2161 < 2.2e-16 ***
## - blast_furnace  1     24890  95127 4425.8  338.4222 < 2.2e-16 ***
## - age           1     74879 145117 4831.6 1018.1138 < 2.2e-16 ***
## - cement        1     76726 146963 4843.8 1043.2183 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```



```
## Step: AIC=4062.01
## y_concrete_compressive ~ age + cement + super_plast + blast_furnace +
## water + fly_ash
##
##           Df Sum of Sq    RSS    AIC    F value    Pr(>F)
## <none>                64878 4062.0
## + coarse             1         11 64868 4063.9      0.1557      0.6932
## + fine_aggr          1          9 64869 4063.9      0.1314      0.7170
## - super_plast        1       1149 66027 4076.9     16.8912 4.299e-05 ***
## - fly_ash            1       5359 70237 4136.3     78.8040 < 2.2e-16 ***
## - water              1       6844 71722 4156.4    100.6397 < 2.2e-16 ***
## - blast_furnace      1      28681 93559 4411.8    421.7378 < 2.2e-16 ***
## - cement             1      68363 133241 4751.6   1005.2373 < 2.2e-16 ***
## - age                1     75394 140272 4801.0   1108.6315 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A ce niveau, nous avons utilisé le critère Fisher. nous avons utilisé deux méthodes à savoir l'algorithme bidirectionnel descendante puis l'algorithme bidirectionnel ascendant. avec la première méthode, on a :  $y_{\text{concrete\_compressive}} \sim \text{cement} + \text{blast\_furnace} + \text{fly\_ash} + \text{water} + \text{super\_plast} + \text{age}$ . La deuxième méthode quant à elle nous donne le modèle suivant :  $y_{\text{concrete\_compressive}} \sim \text{age} + \text{cement} + \text{super\_plast} + \text{blast\_furnace} + \text{water} + \text{fly\_ash}$ . Avec les deux méthodes, nous avons pratiquement les memes résultats. De plus, les deux autres critères à savoir bic et aic ont donné les memes résultats avec les deux algorithmes à savoir : l'exhaustif et génétique.

### Erreur théorique

On va appliquer ici les méthode pour estimer l'erreur théorique d'un modèle : on compare les modèles générés par le critère bic, aic et Fisher.

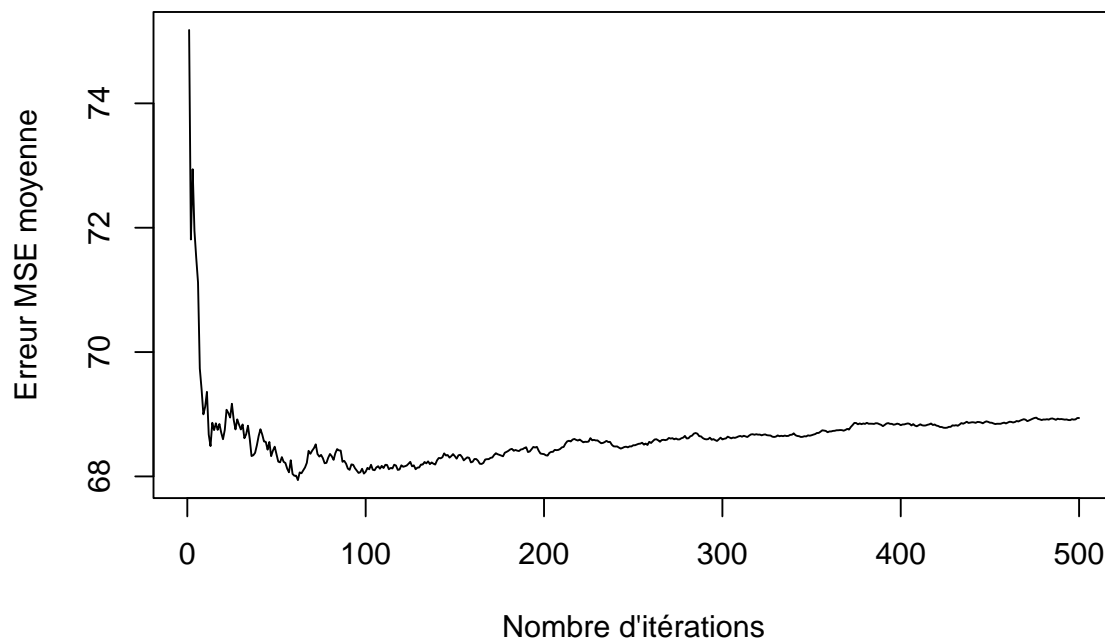
#### 5.4.0.1 Méthode 1 : estimation de l'erreur par la méthode apprentissage/validation

```
## Warning: le package 'ISLR' a été compilé avec la version R 4.4.2
```

```
## [1] 72.58354
```

```
## [1] 71.78797
```

### Convergence de l'erreur de prévision



```
## [1] "Erreur de prévision finale: 68.9399"
```

On voit ici que pour la méthode d'apprentissage validation, le modèle retenu avec les AIC/BIC/Fisher a une erreur théorique qui est inférieure à celle du modèle complet c'est-à-dire que ce modèle est plus performant que le modèle complet. modèle retenu avec les critères AIC/BIC/Fisher : “y\_concrete\_compressive ~ age + cement + super\_plast + blast\_furnace + water + fly\_ash”. En répétant la méthode plusieurs fois (dans ce cas 500 fois), on a eu une erreur théorique inférieure à celui des deux méthodes. Malgré les inconvénients de cette méthode, en répétant la méthode plusieurs fois (la loi de grands nombres), l'erreur théorique tend vers l'erreur théorique optimale. On regarde à présent la méthode de validation croisée K-fold.

#### 5.4.0.2 Méthode 2 : estimation de l'erreur par la K-fold CV

```
## Erreur estimée par validation croisée ( 10 -fold) : 109.4825
```

On trouve une erreur plus importante que celle de la méthode précédente. L'erreur théorique de cette méthode est de 109.43 contre 69.24 pour la méthode apprentissage/validation. L'estimation de l'erreur théorique de prévision pour la méthode

apprentissage/validation dépend de la partition aléatoire, c'est moins précis car on n'utilise qu'une partie des données pour entraîner le modèle.

**5.4.0.3 Sélection du meilleur modèle** Pour la sélection du modèle, on s'est appuyée sur trois approches complémentaires :

1. les critères statistiques (AIC/BIC/Fisher),
2. l'analyse descriptive (notamment l'ACP),
3. la littérature sur les propriétés du béton.

#### 1. Sélection par AIC/BIC/Fisher

Parmi les huit variables initiales, six ont été retenues : *age*, *cement*, *fly\_ash*, *water*, *super\_plast* et *blast\_furnace*.

Les variables *fine\_aggr* et *coarse* ont été écartées selon les critères AIC/BIC.

#### 2. Apport de l'ACP

L'ACP a permis d'identifier des interactions pertinentes qu'on a testé :

- Opposition *water* / *super\_plast* (Dim 1) → test de l'interaction *water* \* *super\_plast*.
- Projection de *fly\_ash* et *fine\_aggr* sur Dim 2 → test de *cement* \* *fine\_aggr* pour explorer leur effet combiné sur la compacité et la résistance (non retenue).
- Bonne projection de *age* (Dim 4) → ajout de l'interaction *age* \* *cement*.
- Opposition entre *coarse* et *super\_plast* sur Dim 2 → interaction testée malgré l'exclusion initiale de *coarse* avec AIC, mais non retenue car non significative et perturbant l'ajustement du modèle.

#### 3. Apport de la littérature

Des interactions basées sur la littérature ont été intégrées :

- *age* \* *blast\_furnace* : le laitier améliore la résistance surtout après plusieurs jours.
- *age* \* *fly\_ash* : les cendres volantes renforcent le béton avec le temps.
- *water*<sup>2</sup> : effet non linéaire du rapport eau/ciment.

#### Modèle final retenu

Variables principales : *age*, *cement*, *blast\_furnace*, *fly\_ash*, *water* (centrée pour réduire la colinéarité avec *water*<sup>2</sup>), *super\_plast*

Interactions retenues :

- *age* \* *cement*
- *age* \* *blast\_furnace*
- *age* \* *fly\_ash*
- *water\_c* \* *super\_plast*
- *water\_c*<sup>2</sup>

Les résultats ci-dessous correspondent au modèle final ajusté, incluant les interactions retenues à partir de l'ACP, de la littérature et des critères AIC/BIC.

```
##
## Call:
## glm(formula = y_concrete_compressive ~ age * cement + water_c +
##       water_c2 + super_plast * water_c + blast_furnace + fly_ash +
##       age * blast_furnace + age * fly_ash, data = bdd_final)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -9.3399340   1.8114560  -5.156 3.07e-07 ***
## age             0.0928123   0.0443613   2.092 0.036687 *
## cement         0.0968491   0.0047159  20.537 < 2e-16 ***
## water_c       -0.2952547   0.0274903 -10.740 < 2e-16 ***
## water_c2       0.0016766   0.0005175   3.240 0.001239 **
## super_plast    0.4431133   0.0778203   5.694 1.65e-08 ***
## blast_furnace  0.0543417   0.0055726   9.752 < 2e-16 ***
## fly_ash        0.0144983   0.0089183   1.626 0.104351
## age:cement     0.0004085   0.0001145   3.569 0.000377 ***
## water_c:super_plast 0.0179500  0.0031173   5.758 1.15e-08 ***
## age:blast_furnace 0.0008896  0.0001324   6.720 3.13e-11 ***
## age:fly_ash    0.0009277   0.0001949   4.760 2.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 62.92863)
##
## Null deviance: 270695  on 960  degrees of freedom
## Residual deviance: 59719  on 949  degrees of freedom
## AIC: 6721.6
##
## Number of Fisher Scoring iterations: 2
```

#### 5.4.0.4 Comparaison des performances par validation croisée (10-fold)

```
## Erreur estimée par validation croisée (10-fold) - modèle final : 64.5623
```

Le modèle initial (basé uniquement sur les critères AIC/BIC) présente une erreur moyenne de 109.48.

Le modèle final, construit en combinant les critères AIC/BIC, les résultats de l'ACP et les apports de la littérature, a une erreur de 64.56.

#### 5.4.0.5 R<sup>2</sup> ajusté

## R<sup>2</sup> ajusté du modèle : 0.7768

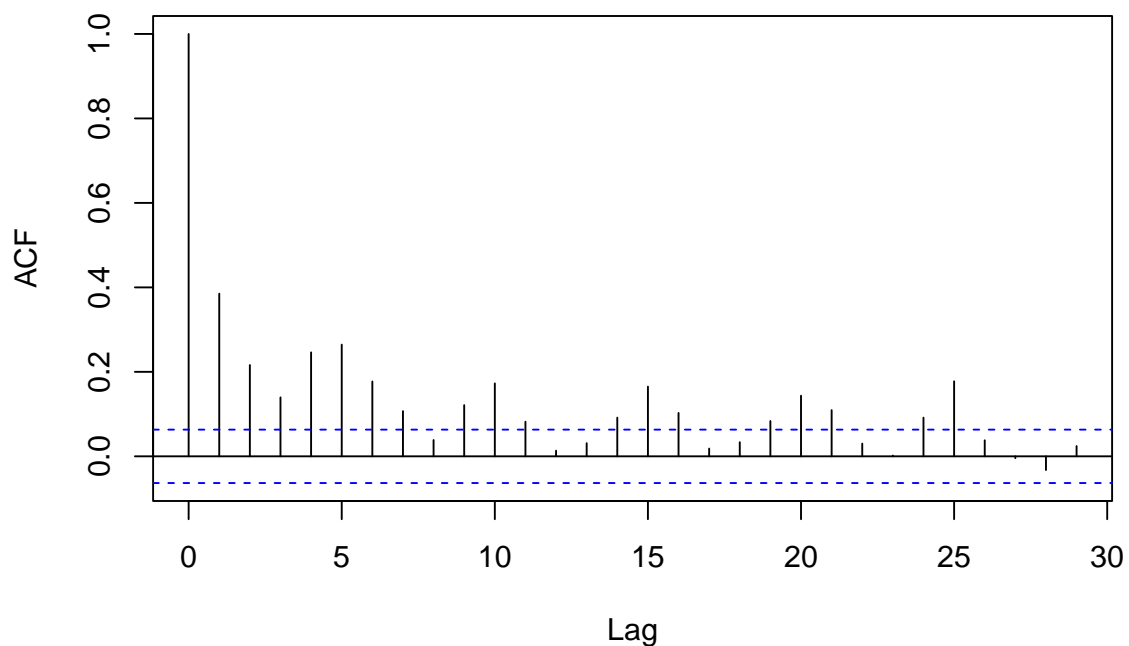
Le R<sup>2</sup> ajusté du modèle final est de 0.7768, ce qui signifie que le modèle explique environ 78 % de la variabilité observée dans la résistance à la compression.

### 5.5 Diagnostic du modèle final retenu

#### 5.5.1 Test de Durbin-Watson

```
##  
## Durbin-Watson test  
##  
## data: modele_complet  
## DW = 1.2214, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

#### Autocorrélation des erreurs – modèle final



Le statistique de Durbin-Watson est égal à 1.22, avec une p-value très petite ( $< 2.2e-16$ ). Cela signifie que les erreurs du modèle sont autocorrélées.

### 5.5.2 Test de Breusch-Pagan

```
##  
## studentized Breusch-Pagan test  
##  
## data: modele_complet  
## BP = 150.28, df = 11, p-value < 2.2e-16
```

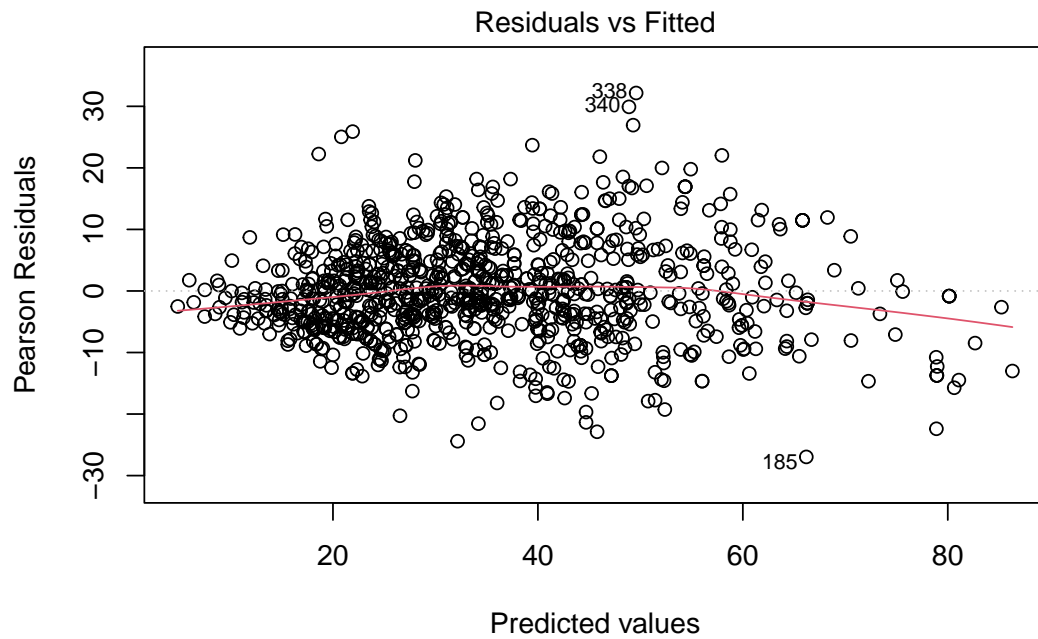
Le test de Breusch-Pagan donne une statistique de 150.28 avec 11 degrés de liberté, et une p-value inférieure à  $2.2e-16$ . Cela signifie que les erreurs du modèle ne sont pas homogènes : il y a hétéroscédasticité.

### 5.5.3 Test de normalité de Shapiro-Wilk

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(modele_complet)  
## W = 0.99088, p-value = 1.14e-05
```

Le test de Shapiro-Wilk donne une statistique de  $W = 0.99088$  avec une p-value =  $1.14e-05$ . Cette p-value étant inférieure à 0.05, on rejette l'hypothèse de normalité des résidus.

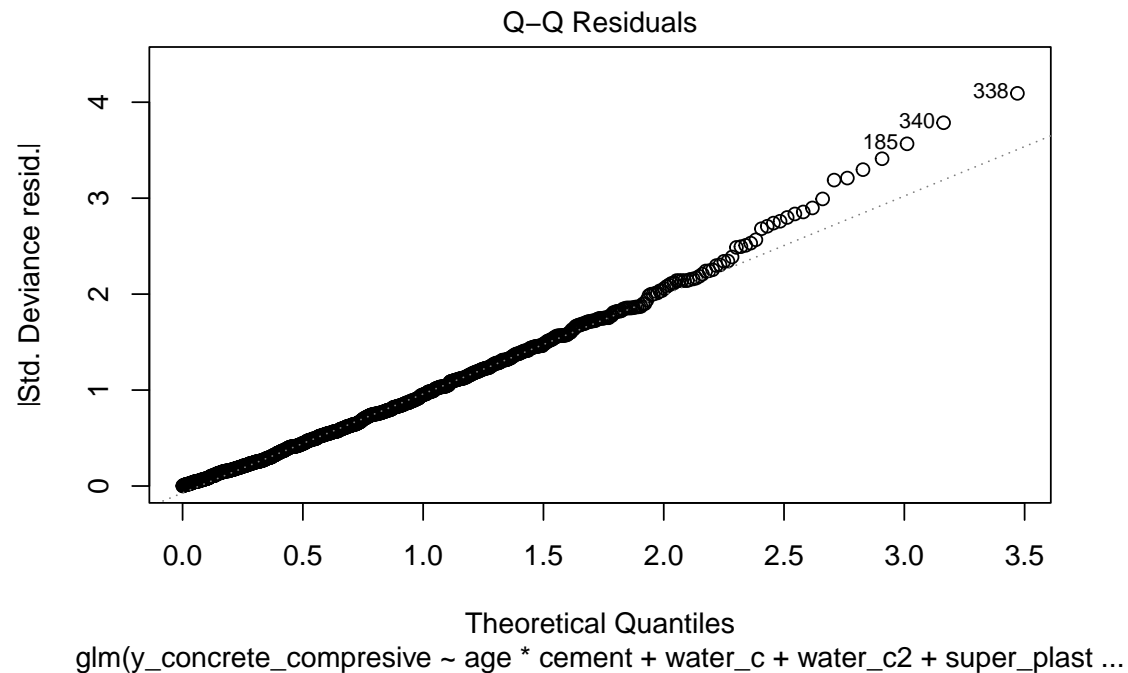
#### 5.5.4 Graphique : Résidus vs Valeurs ajustées



`glm(y_concrete_compressive ~ age * cement + water_c + water_c2 + super_plast ...`

Le nuage de points est globalement centré autour de zéro, mais la courbe rouge montre une légère courbure. Cela suggère que la relation n'est pas parfaitement linéaire, malgré l'ajout d'interactions et de transformations.

### 5.5.5 Graphique : Q-Q plot



Les observations 338, 340 et 185 s'écartent de la droite dans la partie haute, ce qui montre quelques résidus extrêmes. Cela rejoint le résultat du test de Shapiro-Wilk, qui avait déjà indiqué que les résidus ne sont pas normaux.

## 6 Statistique en Grande dimension

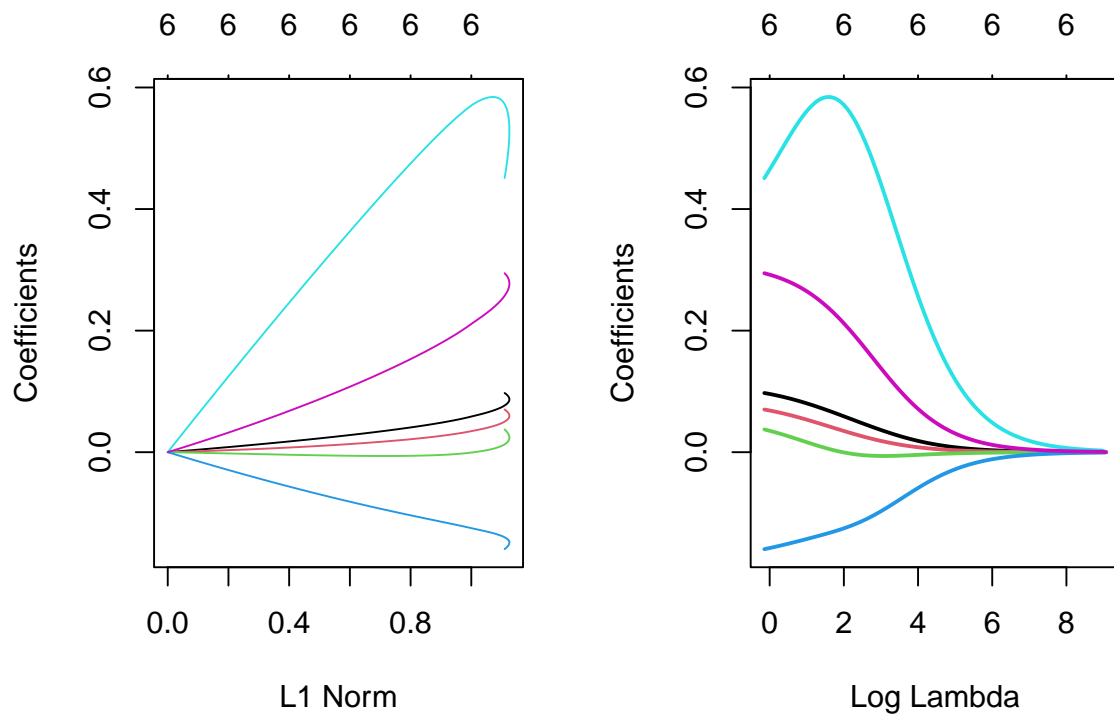
### 6.1 methode ridge

cette methode nous permet de contraindre les coefficients de telle sorte à avoir un équilibre juste entre le biais et la variance.

```
## Le chargement a nécessité le package : Matrix
```

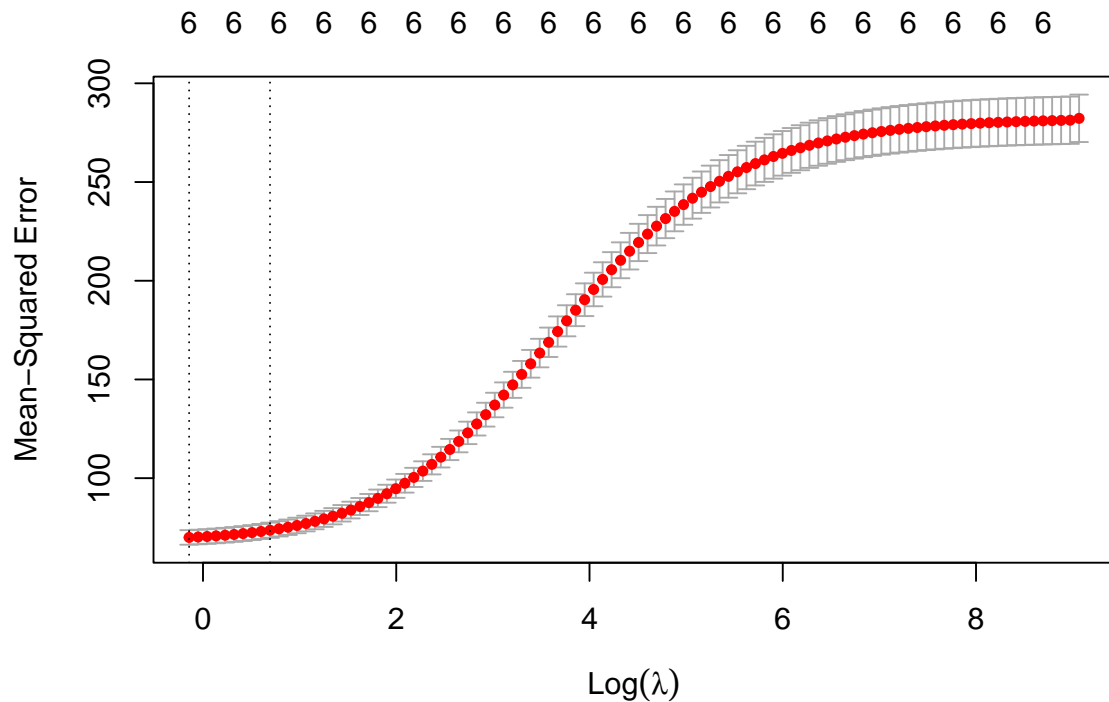
```
## Loaded glmnet 4.1-8
```





```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per  
## fold
```

```
## [1] "Meilleur lambda: 0.8662047727533"
```



```
## [1] "Erreur MSE minimale: 69.9738363782654"
```

```
## [1] "Coefficients optimaux:"
```

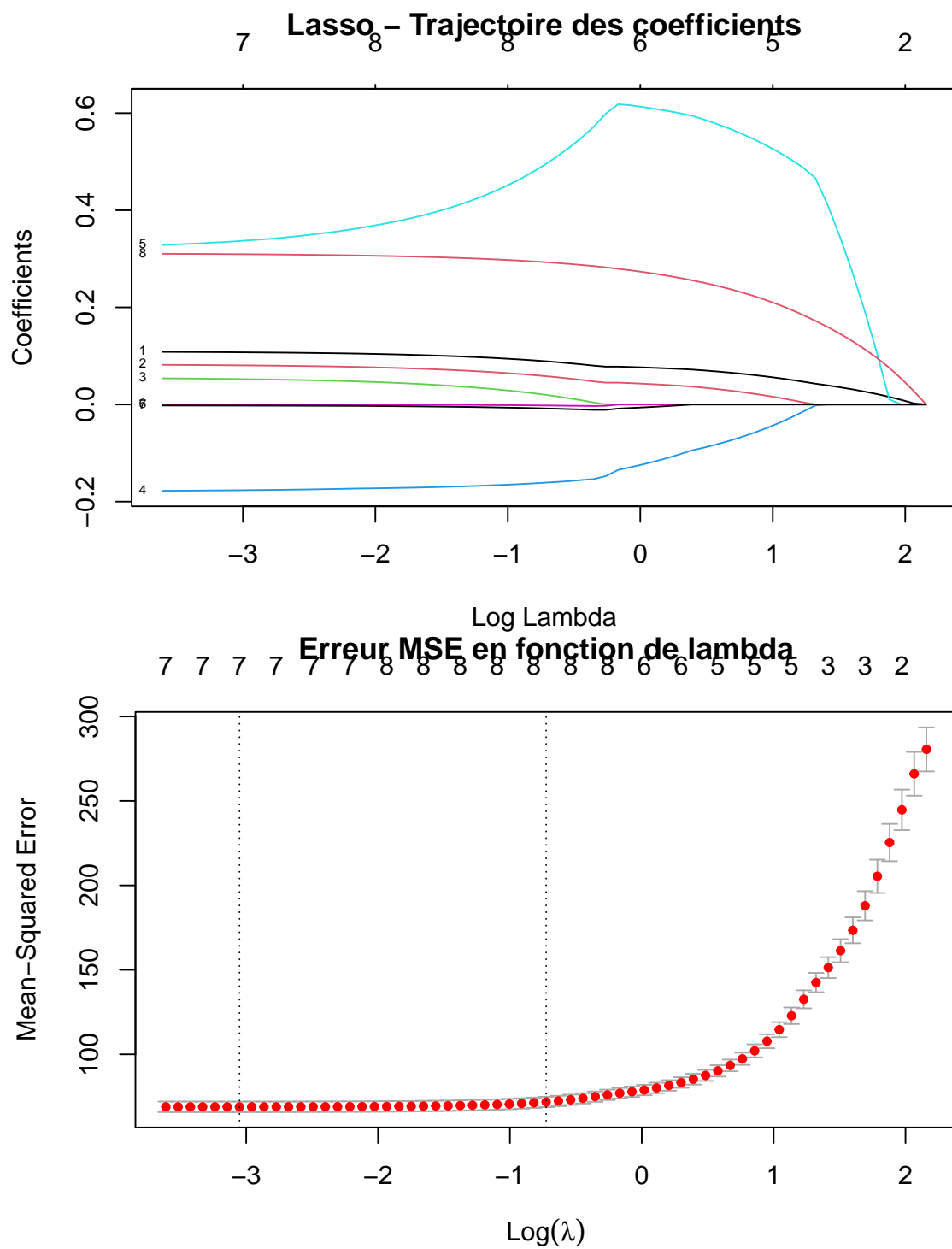
```
## 7 x 1 sparse Matrix of class "dgCMatrix"
```

```
##              s1
## (Intercept) 17.14454013
## cement      0.09734465
## blast_furnace 0.07021218
## fly_ash      0.03744530
## water       -0.15953543
## super_plast  0.45084306
## age         0.29440887
```

On remarque que l'erreur théorique est de 69.97 avec le meilleur lambda qui est de 0.86 .

## 6.2 methode lasso

cette deuxième nous permet de selectionner les variables à inserer dans le modèle tout en contraignant également les coefficients de telle sorte qu'il aii équilibre parfait entre biais et variance.



## Lambda optimal: 0.04731451

```
## 9 x 1 sparse Matrix of class "dgCMatrix"
##                               s1
## (Intercept)    17.752885121
## cement        0.107597256
## blast_furnace  0.080643114
## fly_ash        0.052423318
## water         -0.176721943
## super_plast    0.336140535
## coarse         .
## fine_aggr     -0.002259712
## age           0.309551398

##
## Variables sélectionnées par Lasso:

## [1] "cement"          "blast_furnace" "fly_ash"        "water"
## [5] "super_plast"     "fine_aggr"     "age"

##
## Erreur MSE minimale: 68.88931
```

Pour la méthode de lasso qui est une méthode de sélection de variable, on remarque à la différence des autres méthodes et critères de sélection, elle ajoute `fine_aggr`. le modèle donné par cette méthode est la suivante : “`y_concrete_compressive`” = “`cement`” + “`blast_furnace`” + “`fly_ash`” + “`water`” + “`super_plast`” + “`fine_aggr`” + “`age`”. Son erreur théorique (68.88) est inférieure à celle de la méthode ridge (69.97) avec un  $\lambda$  inférieur à celui de la méthode ridge.