

Regression Linéaire

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

2025-03-04

Contents

1	Introduction	3
2	Revue empirique	3
3	Analyse univariée.	6
3.1	Cement	6
3.2	Blast Furnace Slag	6
3.3	Fly Ash	6
3.4	Water	6
3.5	Superplasticizer	6
3.6	Coarse Aggregate	7
3.7	Fine Aggregate	7
3.8	Age	7
3.9	Concrete Compressive Strength	7
4	Analyse bivariée.	8
4.1	Quantité de ciment et résistance du béton.	8
4.2	Additif chimique et eau.	9

4.3	L'age, une variable importante.	9
4.4	blast_furnace et résistance du béton.	10
5	Analyse Multivariée	16
5.1	Histogramme des valeurs propres	17
5.2	Analyse sur les 2 Premiers Axes	18
5.3	Interprétation du Troisième Axe	19
5.4	Implications pour la Modélisation de la Résistance du Béton	20
6	Modélisation Linéaire	20
7	Annexe	29
7.1	Analyse univariée	29
7.2	Analyse bivariée	38
7.3	Analyse multivariée	40

1 Introduction

(en cours)

2 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 superplastifiant 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.

3 Analyse univariée.

3.1 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/m³. Cette répartition suggère que différentes formulations de béton nécessitent des proportions de ciment variées, influençant potentiellement la résistance finale du béton.

3.2 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.

3.3 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.

3.4 Water

La répartition est relativement normale avec un pic autour de 200 kg/m³, 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.

3.5 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.

3.6 Coarse Aggregate

Le coarse aggregate suit une distribution centrée autour de 950-1000 kg/m³, 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.

3.7 Fine Aggregate

La distribution est également centrée autour de 750-800 kg/m³, 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 influence sa capacité de mise en œuvre et son imperméabilité.

3.8 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.

3.9 Concrete Compressive Strength

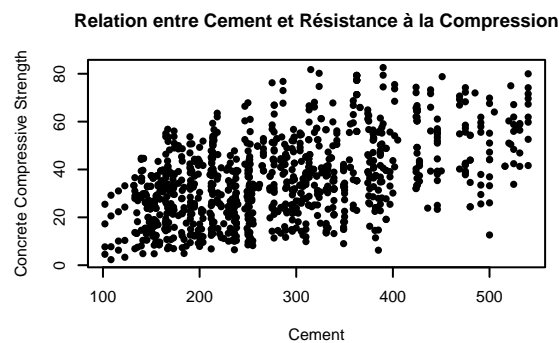
La concrete compressive strength suit une distribution proche d'une loi normale, avec un pic autour de 40 MPa. Cela indique une variabilité contrôlée de la résistance du béton. La concrete compressive strength étant un paramètre clé, cette distribution montre que la plupart des mélanges atteignent une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

4 Analyse bivariée.

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.

4.1 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 :



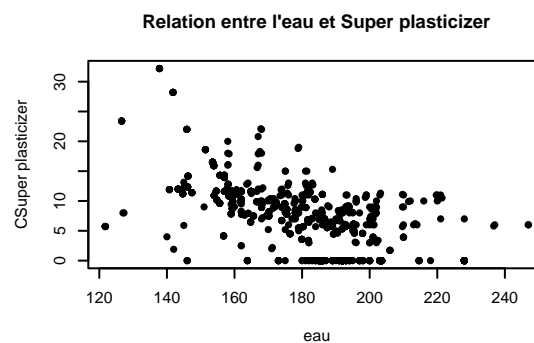
Cette visualisation des données nous donne une première impression du lien potentiel entre ces deux variables. En étudiant plus précisément ces deux variables avec un test de corrélation de Pearson, on se rend compte qu'il y a un lien significatif : ces deux variables sont positivement liées.

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 ρ 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 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.



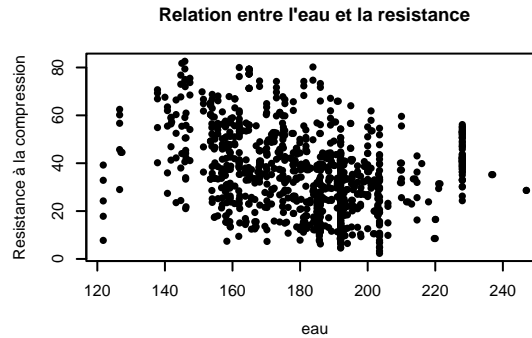
Pour avoir une certitude, nous effectuons alors un test de corrélation de Kendall, qui nous indique, premièrement, qu'il existe un lien significatif entre ces deux variables et, deuxièmement, que le coefficient de corrélation τ étant de **-0.53** (annexe), la négativité de la liaison est prouvée.

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.

4.3 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.

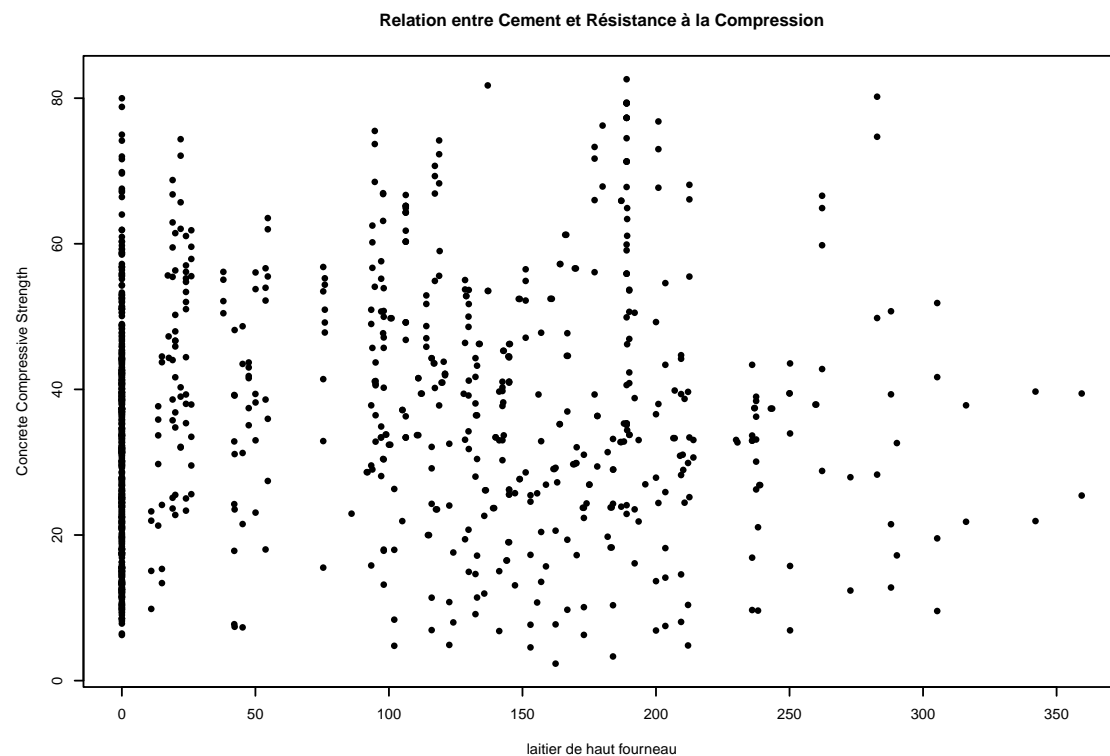


Cette première visualisation ne nous permet pas de dire s'il y aurait en réalité un quelconque lien entre ces deux variables. En conséquence, il est nécessaire de réaliser un test de corrélation.

Ici, le test réalisé nous permet de voir que notre hypothèse est infirmée, car il y a en effet 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 τ de **0,449** (annexe). Les valeurs croissent significativement ensemble, mais il y aura tout de même quelques exceptions.

4.4 blast_furnace et résistance du béton.

graph1 ; On



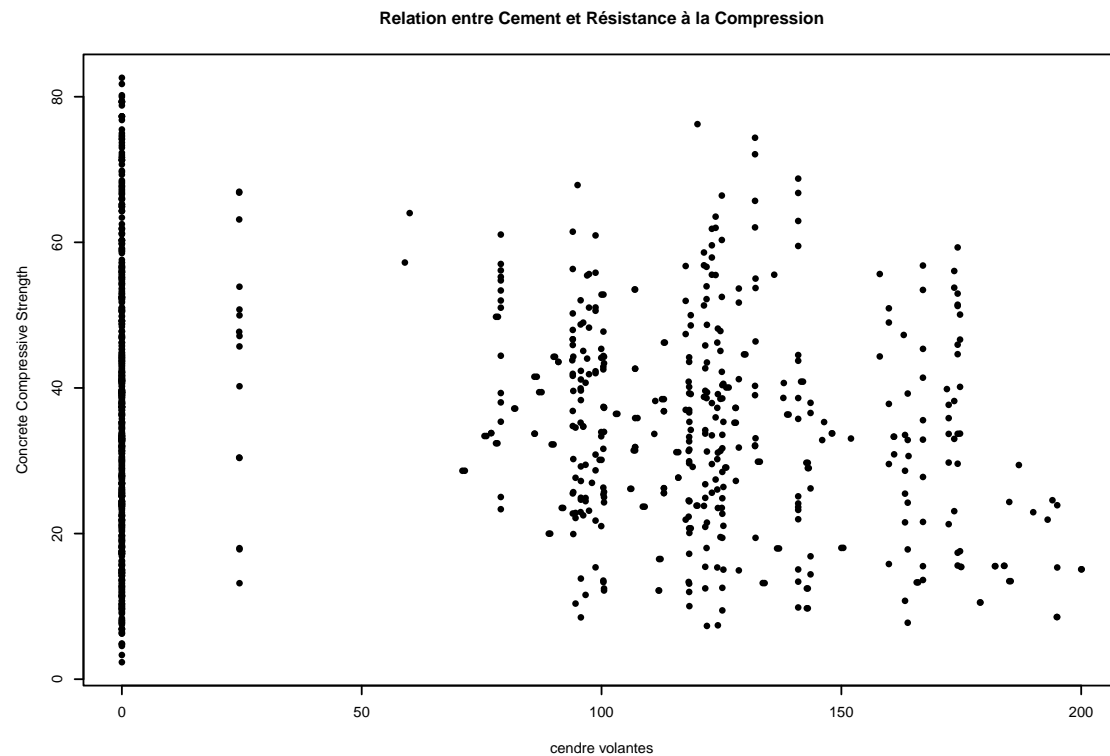
```
test1 ;
```

```
cor.test(bdd$blast_furnace, bdd$y_concrete_compressive, method = "spearman")
```

```
## Warning in cor.test.default(bdd$blast_furnace, bdd$y_concrete_compressive, :  
## Impossible de calculer la p-value exacte avec des ex-aequos
```

```
##  
## Spearman's rank correlation rho  
##  
## data: bdd$blast_furnace and bdd$y_concrete_compressive  
## S = 152531172, p-value = 1.581e-07  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## 0.1624734
```

-> Significatif mais (tau plutôt proche de zero et positif donc lien positif)

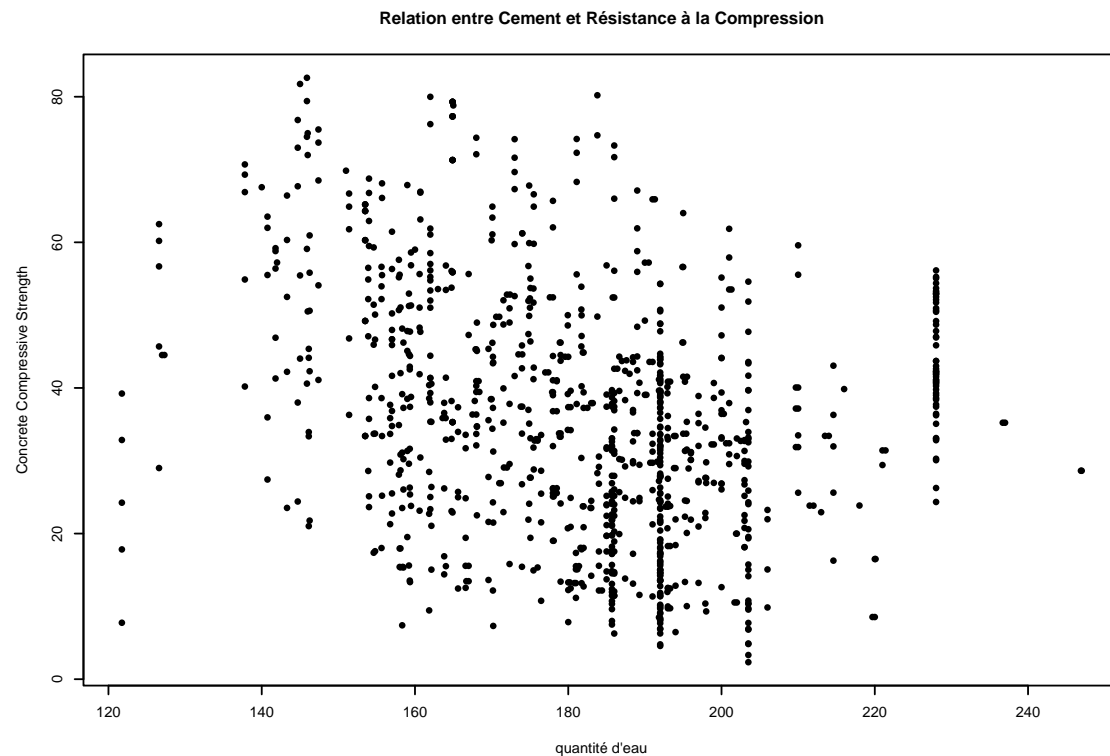


```
cor.test(bdd$fly_ash, bdd$y_concrete_compressive, method = "spearman")
```

```
## Warning in cor.test.default(bdd$fly_ash, bdd$y_concrete_compressive, method =
## "spearman"): Impossible de calculer la p-value exacte avec des ex-aequos
```

```
##
## Spearman's rank correlation rho
##
## data: bdd$fly_ash and bdd$y_concrete_compressive
## S = 196318556, p-value = 0.01233
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.07795675
```

-> pvalue a 0.01 donc seuil de significativité et pour le tau on est a -0.07 donc inversement proportionnel

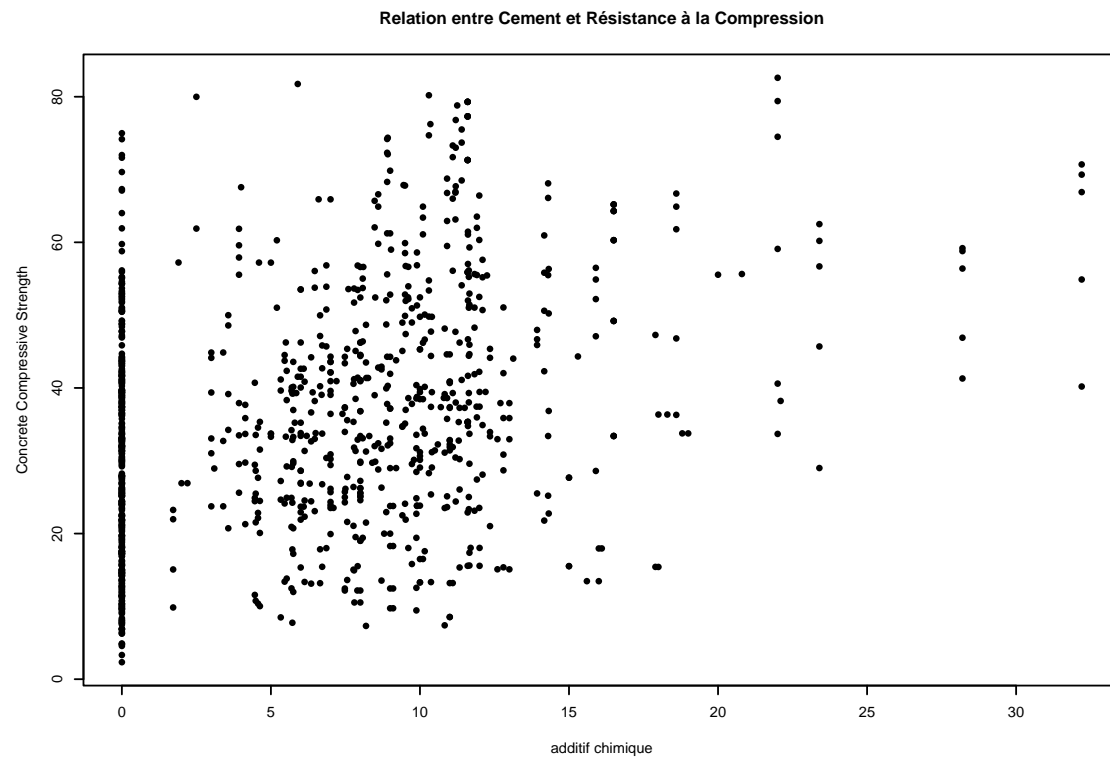


```
cor.test(bdd$water, bdd$y_concrete_compressive, method = "spearman")
```

```
## Warning in cor.test.default(bdd$water, bdd$y_concrete_compressive, method =
## "spearman"): Impossible de calculer la p-value exacte avec des ex-aequos
```

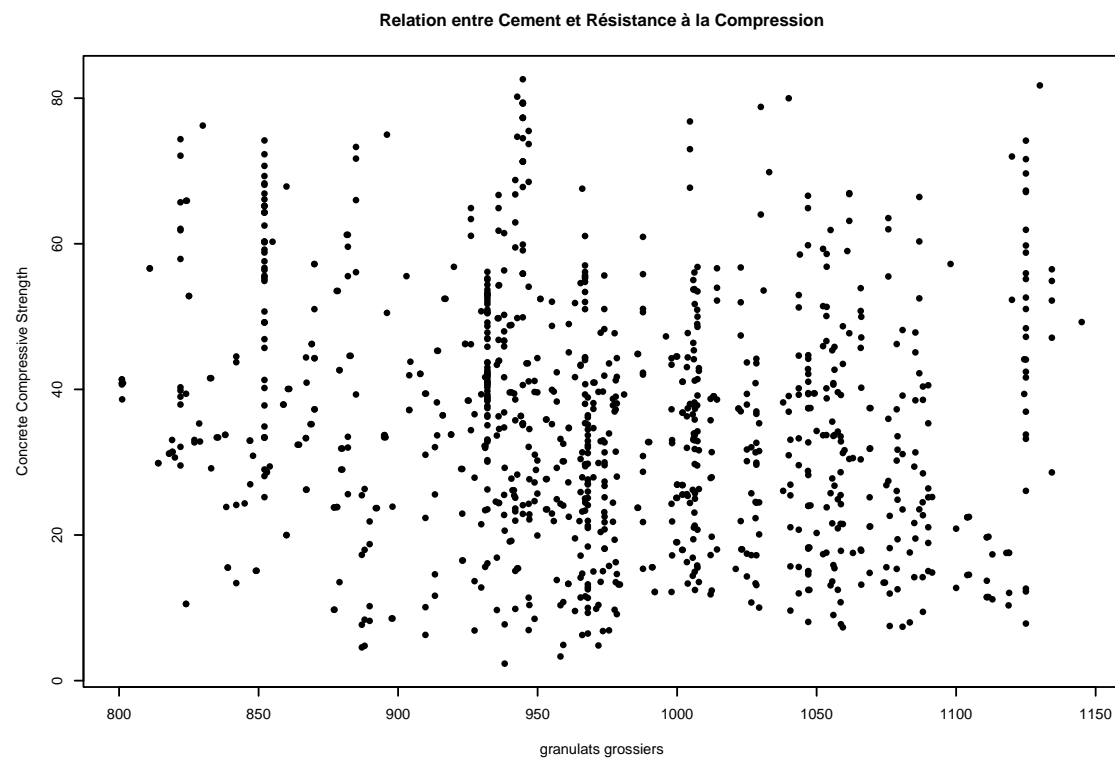
```
##
## Spearman's rank correlation rho
##
## data: bdd$water and bdd$y_concrete_compressive
## S = 238281768, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.3083707
```

-> significatif car $< 2.2e-16$ donc et inversement proportionnel : moins il y a d'eau, plus la résistance est importante



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

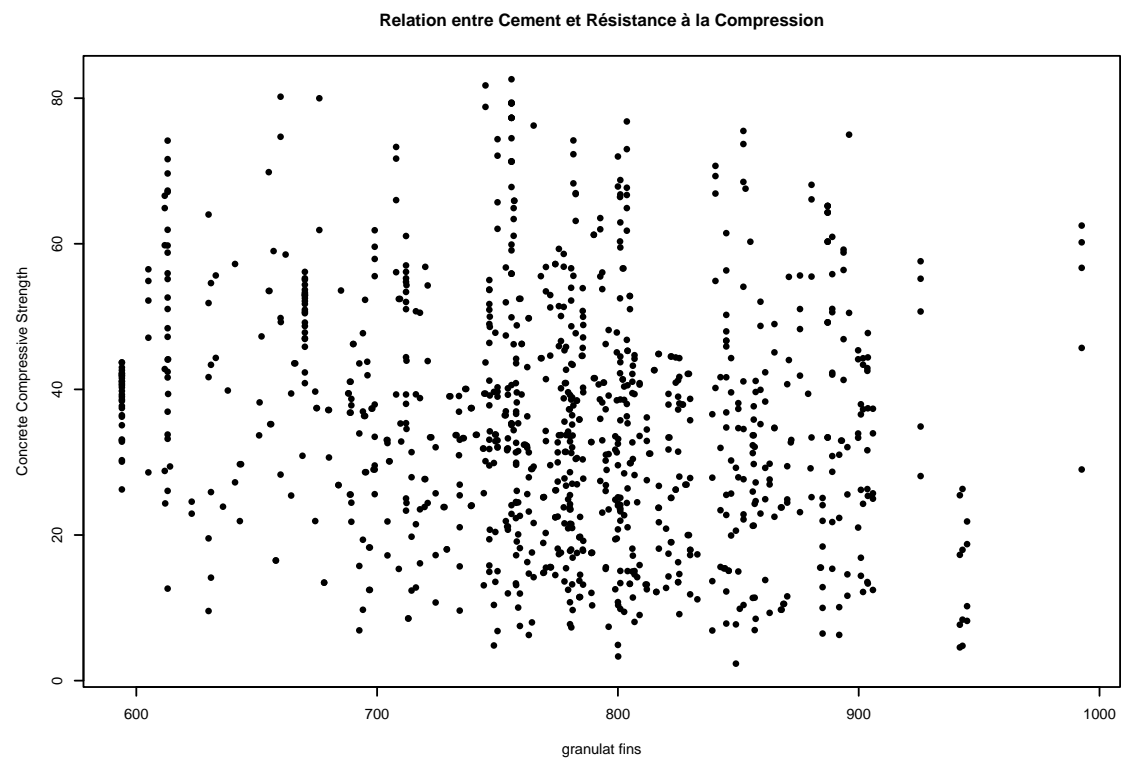
```
##
##  Kendall's rank correlation tau
##
## data:  bdd$super_plast and bdd$y_concrete_compressive
## z = 11.39, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.2490572
```



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

```
##  
## Kendall's rank correlation tau  
##  
## data: bdd$coarse and bdd$y_concrete_compressive  
## z = -5.915, p-value = 3.318e-09  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
##      tau  
## -0.1237223
```

-> lien significatif, tau a -0.123 donc fort lien négatif



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

```
##
##  Kendall's rank correlation tau
##
## data:  bdd$fine_aggr and bdd$y_concrete_compressive
## z = -5.8569, p-value = 4.717e-09
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1222597
```

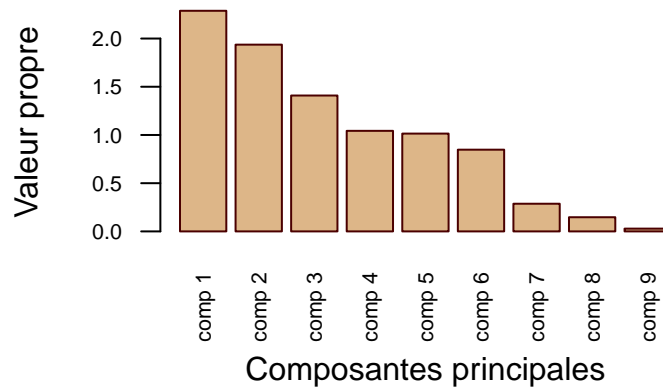
-> lien n significatif, tau négatif donc moins il y a de granulat fin, plus la resistance est grande

5 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.

5.1 Histogramme des valeurs propres

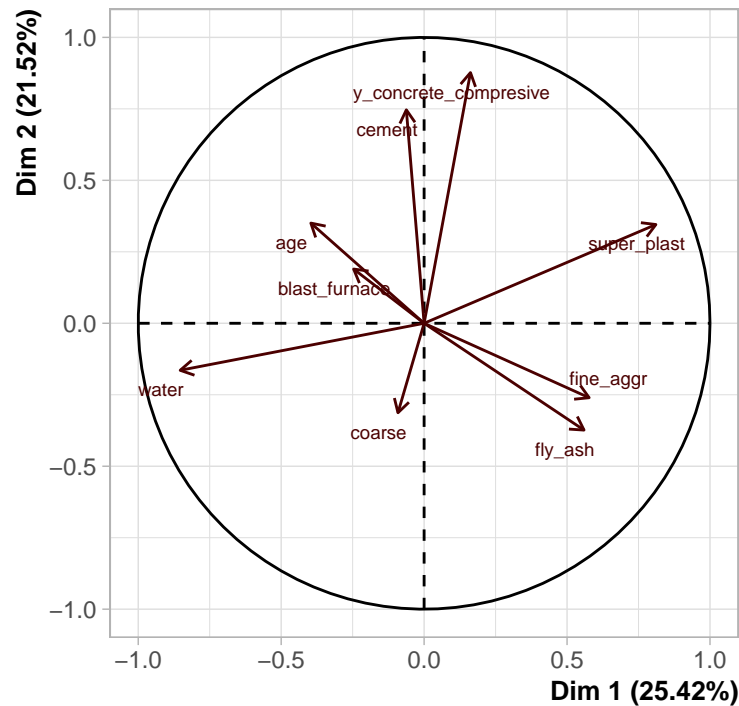


Le graphique illustre l’histogramme des valeurs propres issu de l’ACP, permettant d’identifier le nombre d’axes principaux à retenir.

- 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.

5.2 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 **46,94 %** de la variance totale des données (**25,42 % pour Dim 1 et 21,52 % pour Dim 2**).

5.2.1 Interprétation du Premier Axe (Dim 1 : 25,42 %)

Le premier axe distingue les formulations riches en eau (**water**) et l'âge du béton (**age**) de celles utilisant des superplastifiants (**super_plast**), des cendres volantes (**fly_ash**), et une proportion plus élevée de granulats fins (**fine_aggr**).

- **water** (31,82 % de contribution, $\cos^2 = 0.73$) est la 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** (28,74 % de contribution, $\cos^2 = 0.66$) est projeté dans la direction opposée à **water**, traduisant son rôle de réduction du besoin en eau, favorisant ainsi la résistance.

- **fly_ash** (13,68 % de contribution, $\cos^2 = 0.31$) et **fine_aggr** (14,57 % de contribution, $\cos^2 = 0.33$) sont également bien projetés sur cet axe, suggérant que ces composants influencent la résistance du béton.
- **age** (6,86 % de contribution, $\cos^2 = 0.16$) est aussi bien représenté sur Dim 1, confirmant que la résistance du béton augmente avec le temps de durcissement.
- **y_concrete_compressive** (1,15 % de contribution, $\cos^2 = 0.03$) est légèrement projetée sur Dim 1, mais son faible \cos^2 indique qu'elle est mieux représentée sur un autre axe.

5.2.2 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 de granulats grossiers (**coarse**).

- **y_concrete_compressive** (39,70 % de contribution, $\cos^2 = 0.77$) est la variable la plus influente sur Dim 2. Cela signifie que la résistance à la compression est fortement expliquée par cette dimension.
- **cement** (28,78 % de contribution, $\cos^2 = 0.56$) est bien représenté sur Dim 2, indiquant que la quantité de ciment influence fortement la résistance mécanique.
- **coarse** (5,05 % de contribution, $\cos^2 = 0.10$) est projeté dans la direction opposée à **cement** et **y_concrete_compressive**, suggérant que les formulations contenant une proportion élevée de granulats grossiers sont associées à une résistance plus faible.

5.3 Interprétation du Troisième Axe

5.3.1 Dim(1-3) : 45 % de la Variance Totale (Voir annexe : Partie E)

- **cement** est très bien représenté ($\cos^2 = 0.88$) et contribue en grande partie à la construction de cet axe (66 %).
- **fly_ash** contribue à hauteur de 17 % à la construction de cet axe et est bien représenté avec $\cos^2 = 0.22$.
- **y_concrete_compressive** suit la même direction que **cement**, suggérant une influence positive sur la résistance à la compression.

5.3.2 Dim(2-3) : 35 % de la Variance Totale (Voir annexe : Partie D)

- Contrairement à la précédente interprétation, on remarque que **la combinaison de cement et fly_ash augmente la résistance du béton**.
- Ce contraste sera analysé plus précisément après la construction du modèle.

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, cendres volantes et granulats fins.
2. **Axe 2 (Dim 2)** : Opposition entre le laitier de haut fourneau et les granulats grossiers.
3. **Axe 3 (Dim 3)** : Influence du lien `cement-fly_ash` sur la résistance à la compression du béton.

5.4 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` étant bien projeté sur Dim 1, son interaction avec `super_plast` pourrait être pertinente pour examiner leur effet combiné sur la compacité et la résistance.
- `age` étant bien représenté sur Dim 1, une interaction avec `water` pourrait être étudiée pour voir comment l'évolution du béton avec le temps est influencée par l'eau.
- `y_concrete_compressive` étant principalement projeté sur Dim 2, un modèle linéaire devrait inclure `cement` comme facteur explicatif principal.
- L'opposition entre `coarse` et `cement` sur Dim 2 suggère qu'un effet combiné de ces deux variables pourrait influencer la résistance mécanique.

6 Modélisation Linéaire

Table 1: Tableau des résultats de la régression linéaire

Coefficient	Estimate	Std_Error	t_value	Pr
(Intercept)	-23.163756	26.588421	-0.871	0.383851
cement	0.119785	0.008489	14.110	< 2e-16
blast_furnace	0.103847	0.010136	10.245	< 2e-16
fly_ash	0.087943	0.012585	6.988	5.03e-12
water	-0.150298	0.040179	-3.741	0.000194
super_plast	0.290687	0.093460	3.110	0.001921
coarse	0.018030	0.009394	1.919	0.055227
fine_aggr	0.020154	0.010703	1.883	0.059968
age	0.114226	0.005427	21.046	< 2e-16

Interprétation des p-values

Les résultats de la régression montrent que :

- **Intercept** : La p-value est de **0.38**, ce qui est **supérieur à 0.05**. Nous retenons donc l'**hypothèse H0**, selon laquelle cette valeur n'est **pas significative**.
- **Variables 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. Toutefois, elles deviennent **significatives pour un seuil de 10 %**.

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 diverses 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 0.119785, cela signifie que si le ciment augmente d'une unité alors la résistance à la compression augmente moins proportionnellement que l'augmentation du ciment (0.12 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 0.12 unité de la variable cible. Puis, on s'intéresse à la variable fly_ash (cendres volantes) qui est positivement corrélée avec la variable endogène

(0.087943). Cela pourrait être traduit comme étant une augmentation moins proportionnelle de la résistance comparé aux cendres volantes. De plus, nous remarquons une relation négative entre la variable `water` et la résistance à la compression. cette relation est de l'ordre de -0.15 c'est-à-dire que l'augmentation de l'eau d'une unité entraîne une baisse de la résistance de 0.15. ce qui rejoint notre conclusion de l'analyse multivariée. Par ailleurs, la relation de notre variable cible avec la variable `super_plast` est de 0.29, qui est la relation la plus importante dans notre modèle. cette relation est positive, ce qui signifie que 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 `reponse`, Ce qui confirme les résultats des revues qui stipulent qu'avec la résistance augmente avec le temps.

Test de Durbin-Watson

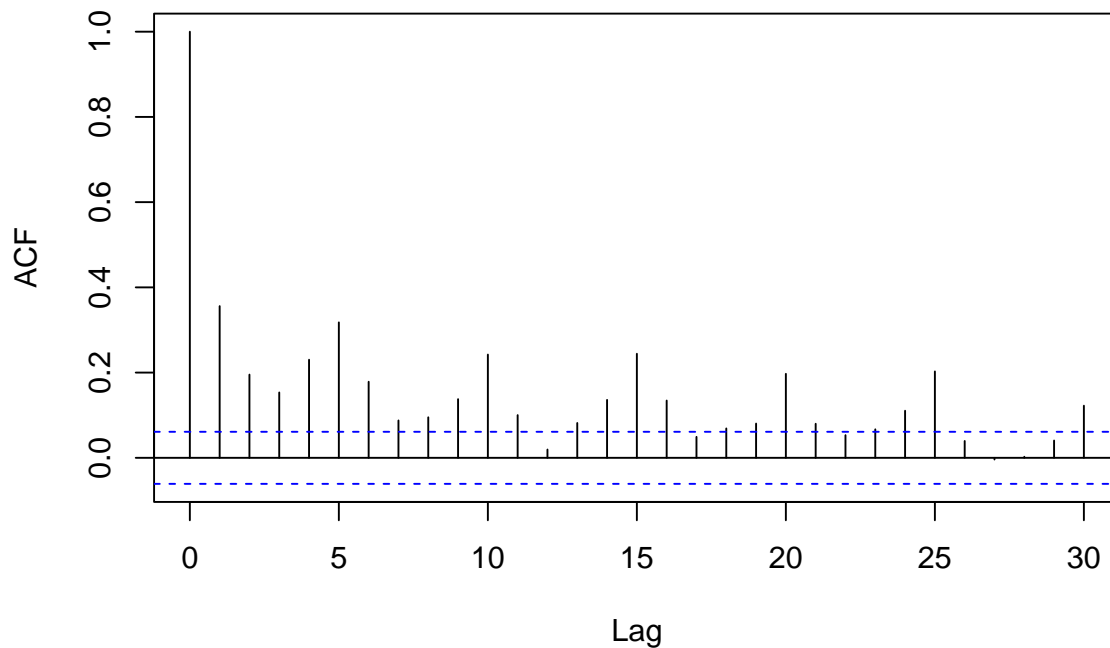
Table 2: Durbin-Watson Test Results

Description	Valeur
DW	1.2815
p-value	< 2.2e-16
Alternative hypothesis	true autocorrelation is not 0

- **Durbin-Watson statistic (DW) : 1.2815**
- **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)	140.25
Degrés de liberté (df)	8
p-value	$< 2.2e-16$

- **Breusch-Pagan statistic (BP) : 140.25**
- **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)	0.99532
p-value	0.002986

- Shapiro-Wilk statistic (W) : **0.99532**

- p-value : **0.002986**

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.

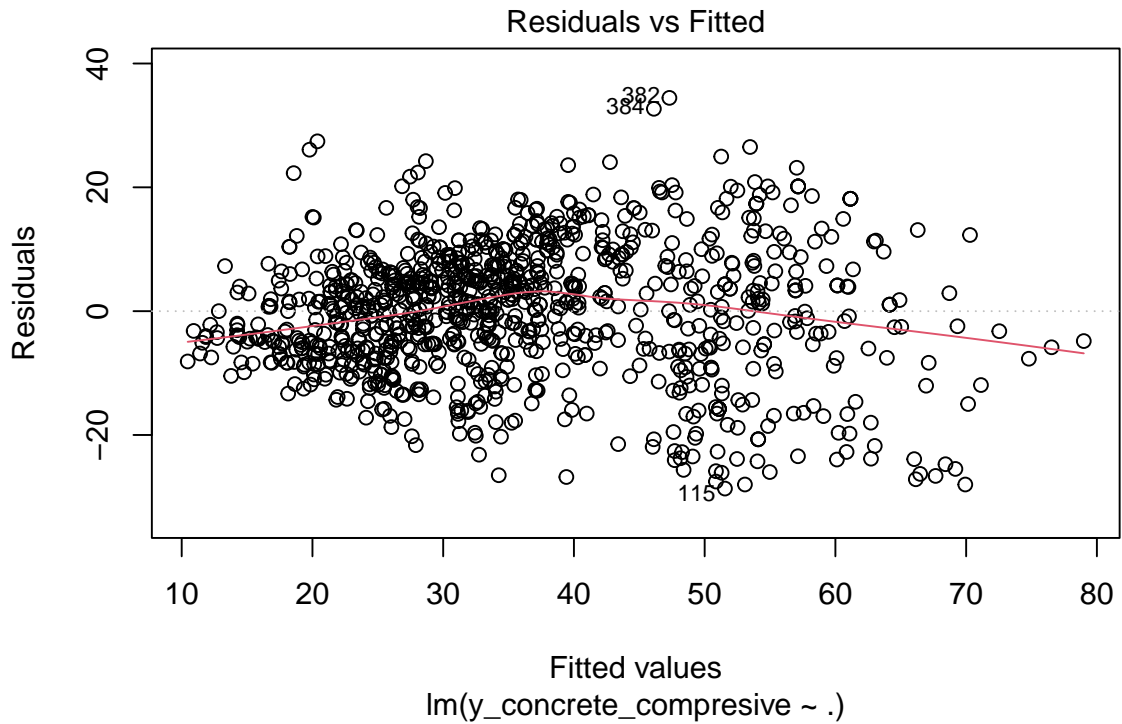
Par conséquent, **les erreurs ne suivent pas une distribution normale**.

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

```
## # A tibble: 6 x 6
```

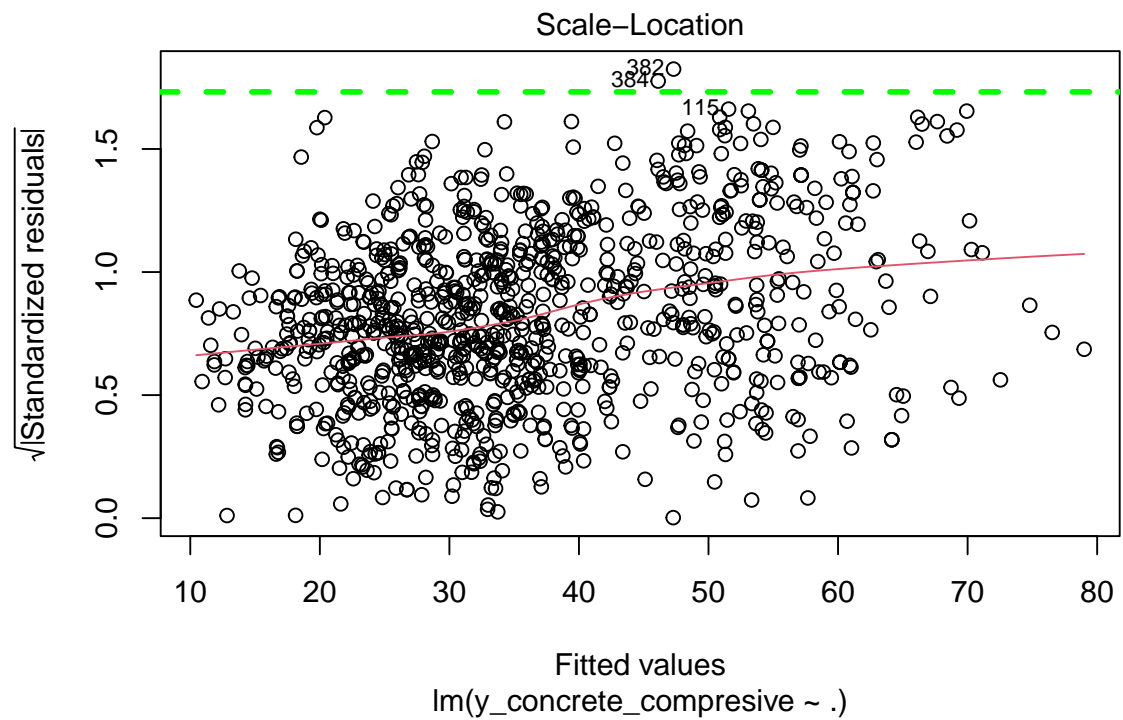
```
##   .fitted .resid   .hat .sigma .cooksd .std.resid
##   <dbl>  <dbl>   <dbl> <dbl>   <dbl>    <dbl>
## 1   53.5   26.5  0.0137  10.4  0.0102     2.57
## 2   53.7    8.14 0.0129  10.4 0.000902     0.788
## 3   56.8  -16.5  0.0170  10.4 0.00495    -1.60
## 4   67.7  -26.6  0.0287  10.4 0.0221    -2.60
## 5   60.9  -16.6  0.0306  10.4 0.00923    -1.62
## 6   26.9   20.2  0.00694  10.4 0.00294     1.95
```

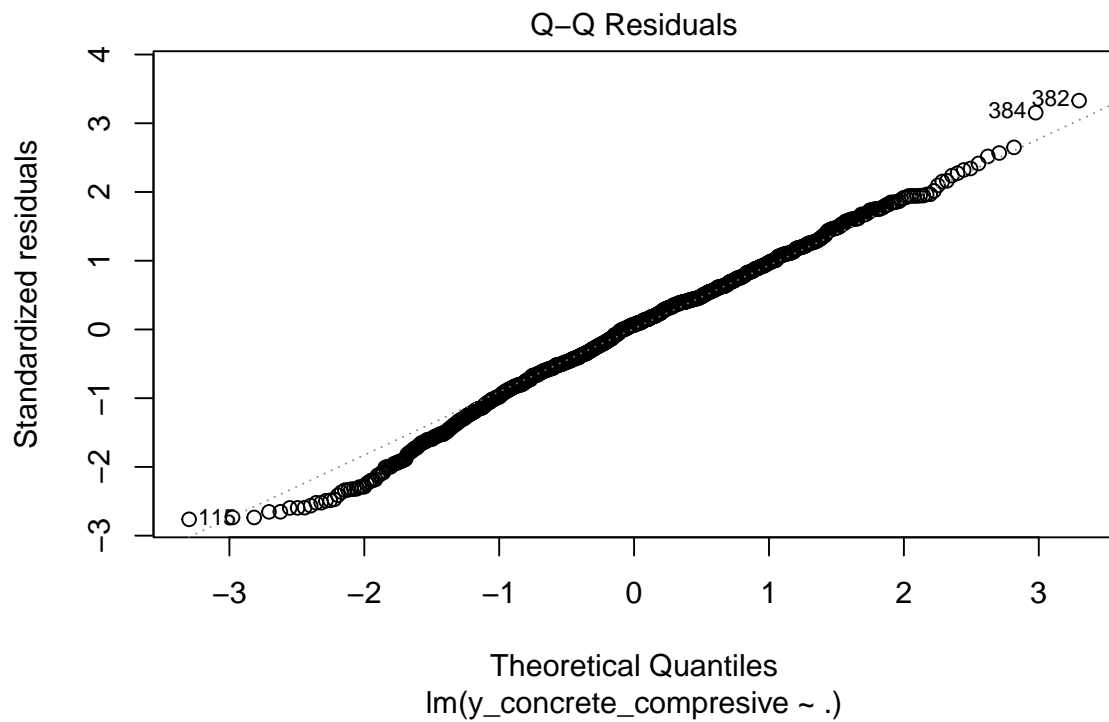
D'après ce plot, il **n'existe aucune linéarité** entre la variable réponse et les variables explicatives.

La **courbe rouge n'est pas horizontale**, ce qui indique que la relation **n'est pas linéaire**.

Les observations **382, 384 et 115** sont les **bétons présentant les plus grandes valeurs en termes de résidus**.

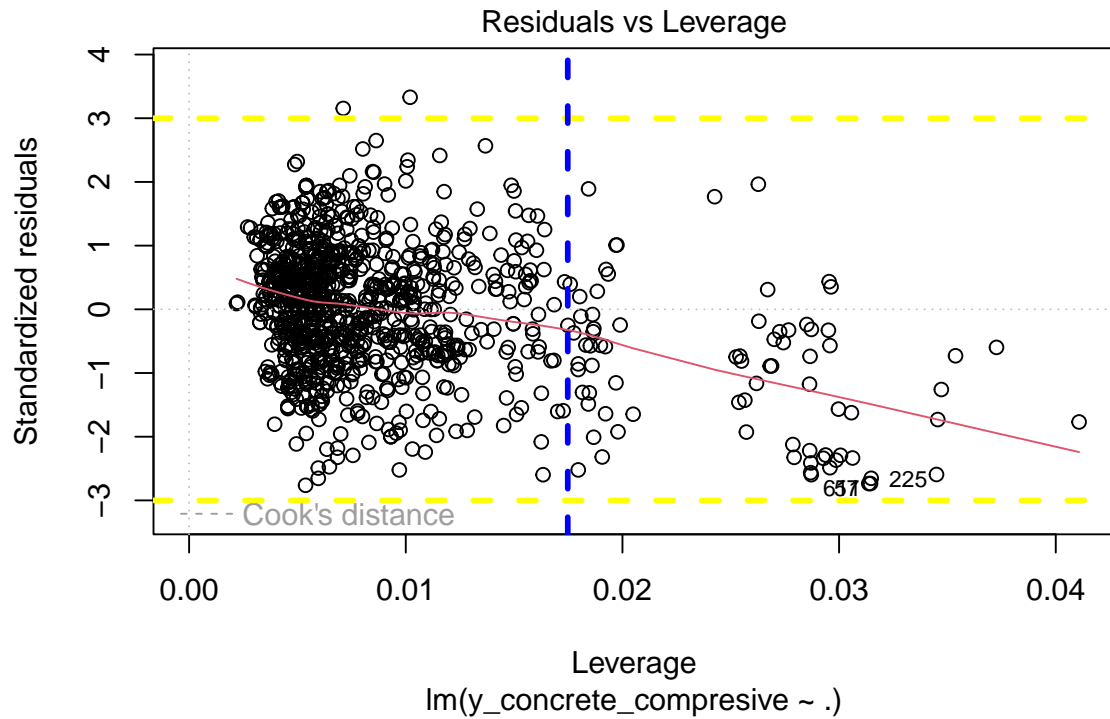


Dans notre cas, la **courbe rouge n'est pas horizontale** et ne sépare pas les points de manière homogène de part et d'autre. Cela indique que **les erreurs ne sont pas homogènes**, ce qui confirme la conclusion du **test de Breusch-Pagan**.



Ce plot de l'**hypothèse de normalité** montre que plusieurs points **ne suivent pas la ligne**, bien qu'ils en soient proches.

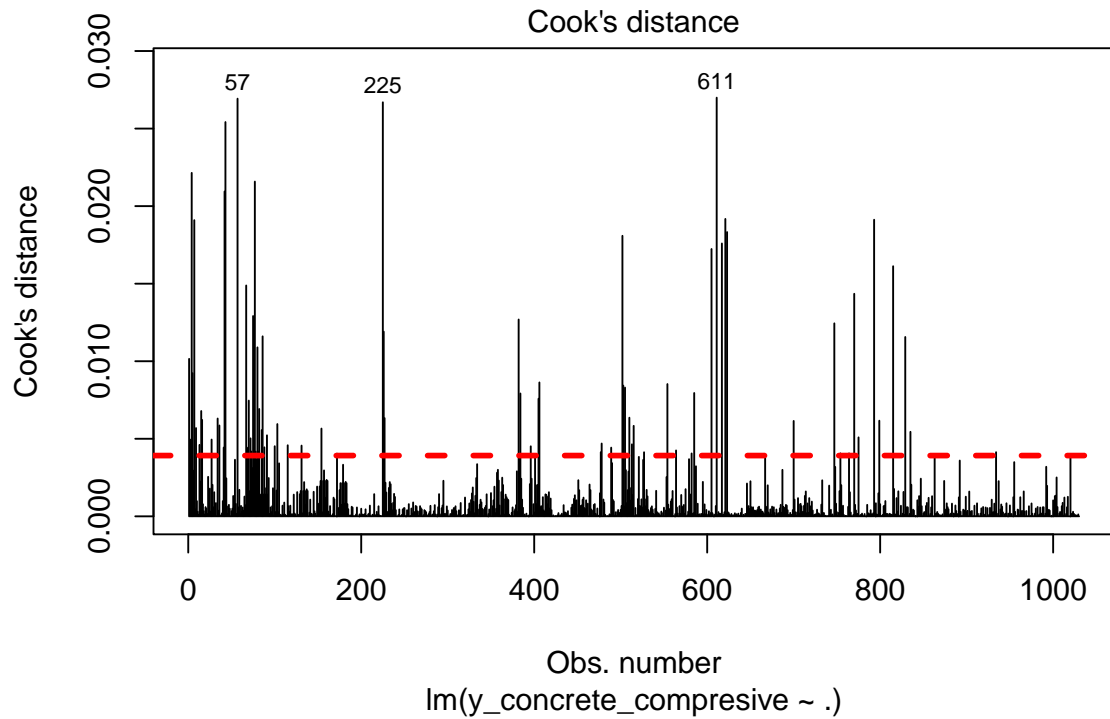
Cela confirme les résultats du **test de normalité de Shapiro-Wilk**, indiquant que les erreurs **ne suivent pas une distribution normale**.



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**.



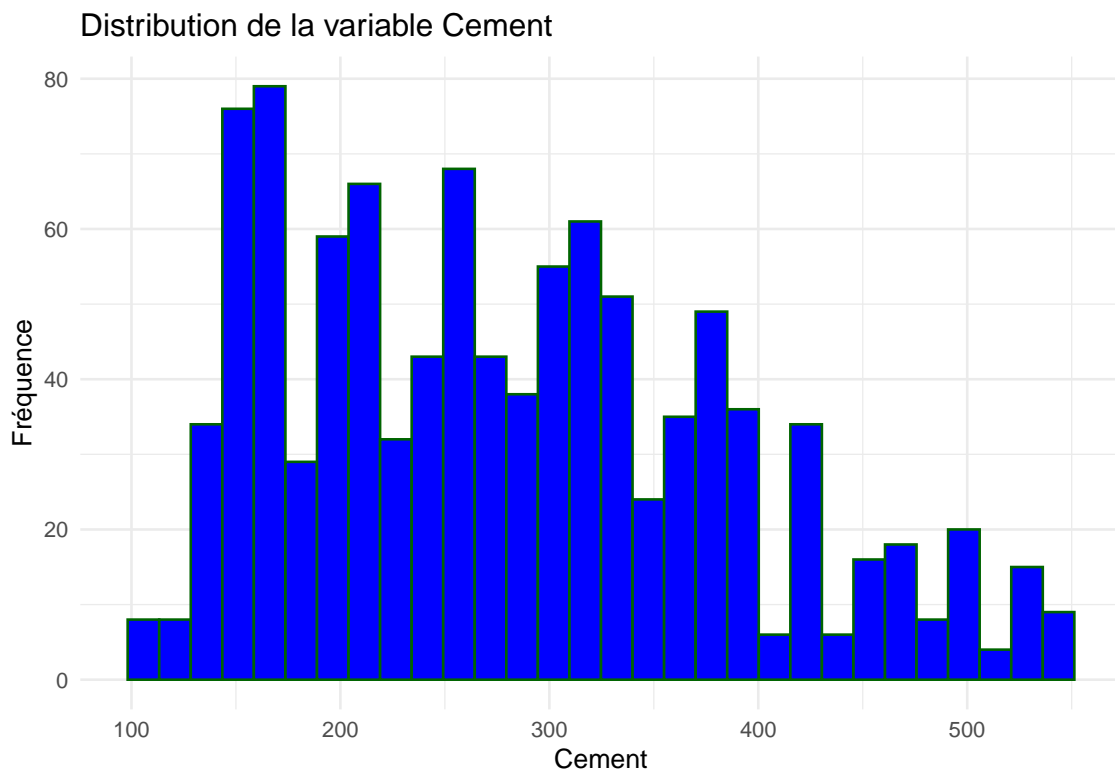
Suite à l'observation des résultats donnés par le graphique, on remarque que **plusieurs points dépassent la ligne** formée par les **pointillés rouges**, qui représente le seuil à partir duquel un point est **considéré comme influent**.

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

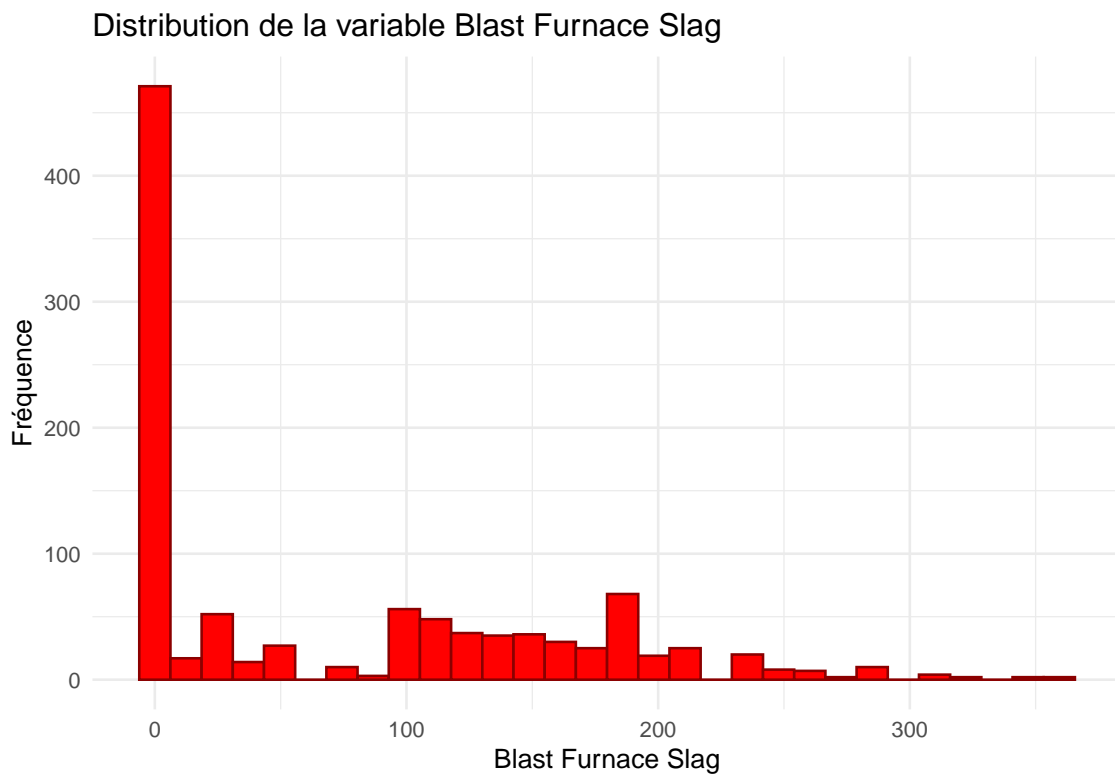
7 Annexe

7.1 Analyse univariée

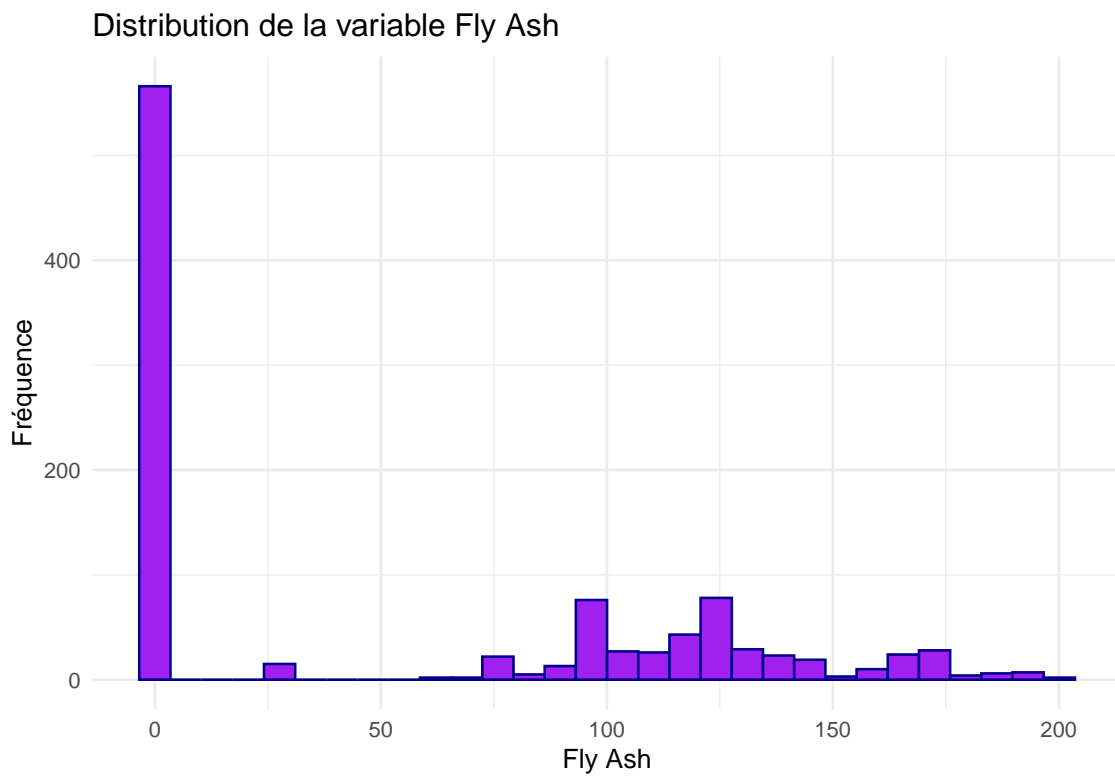
Histogramme pour la variable `Cement`



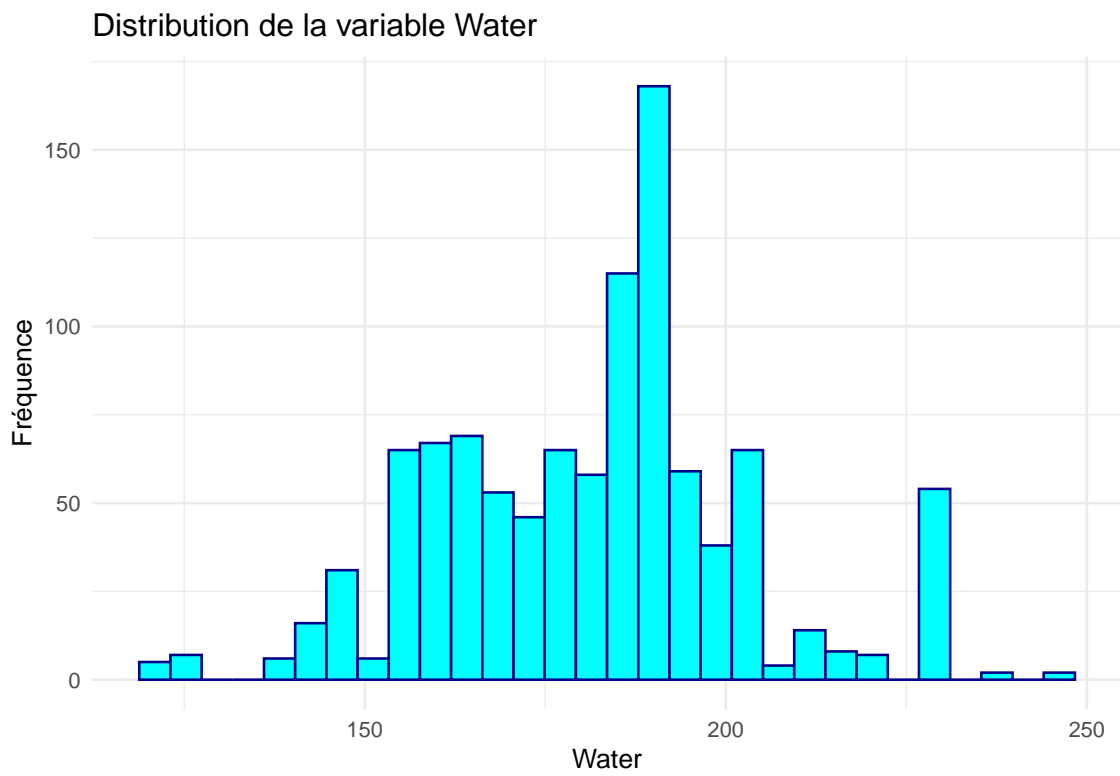
Histogramme pour la variable Blast Furnace Slag



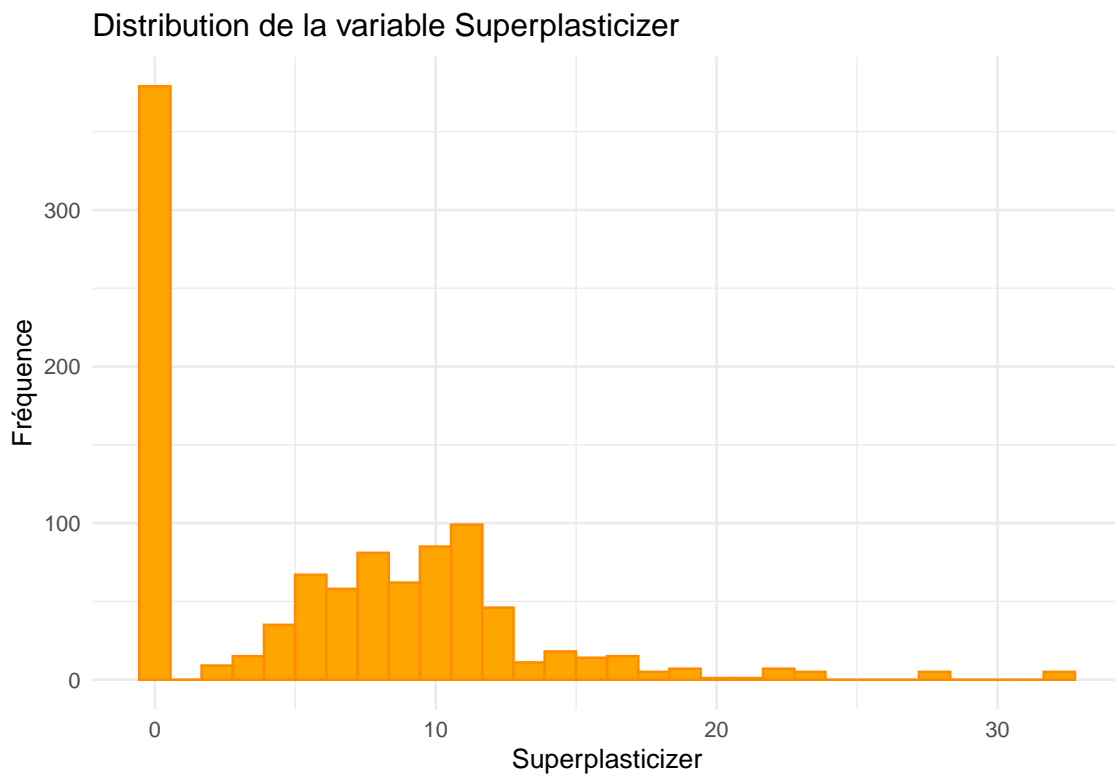
Histogramme pour la variable Fly Ash



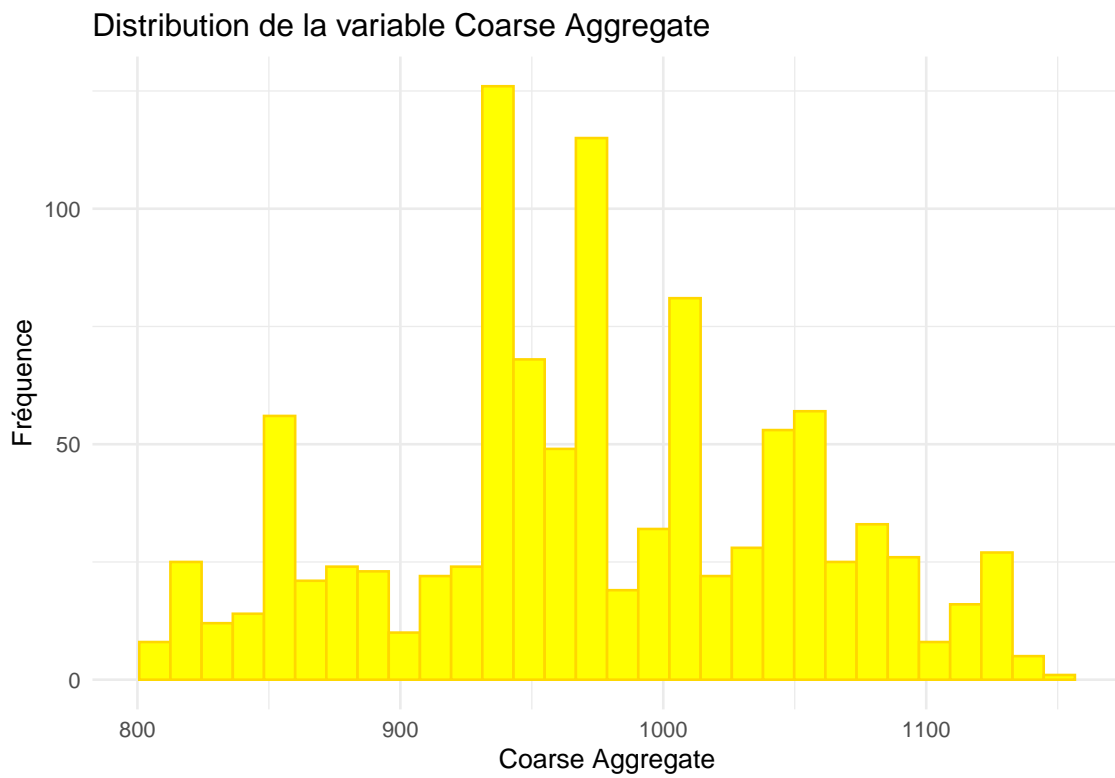
Histogramme pour la variable Water



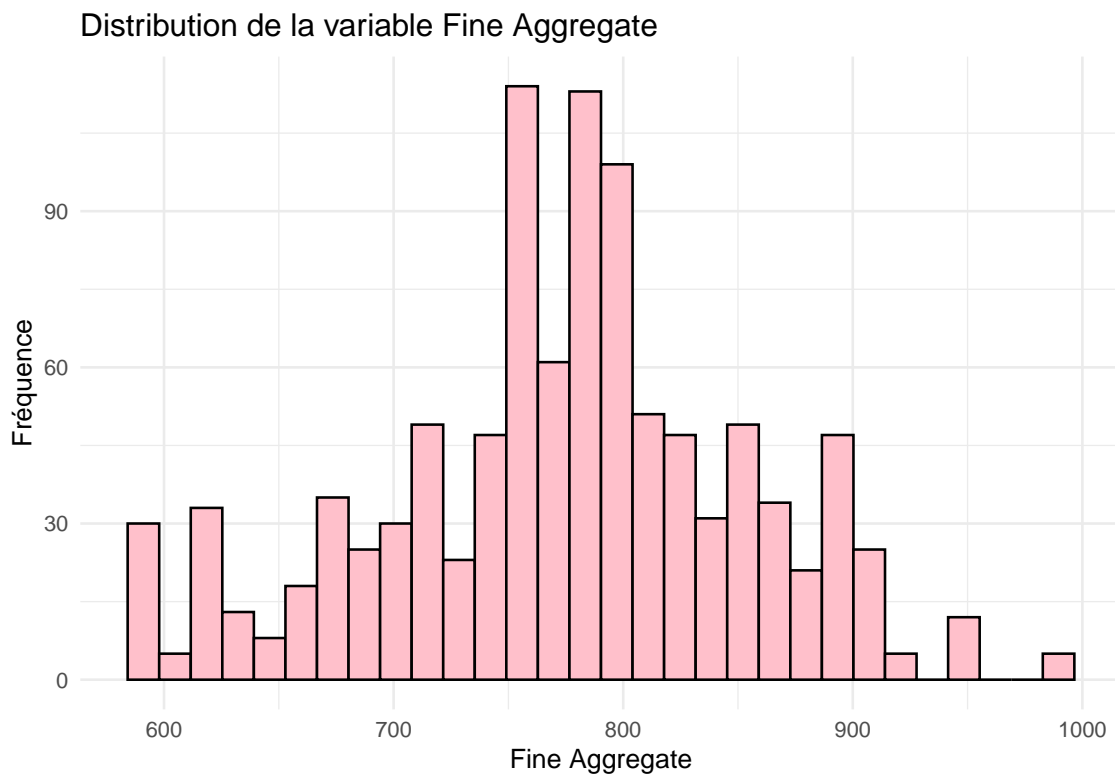
Histogramme pour la variable Superplasticizer



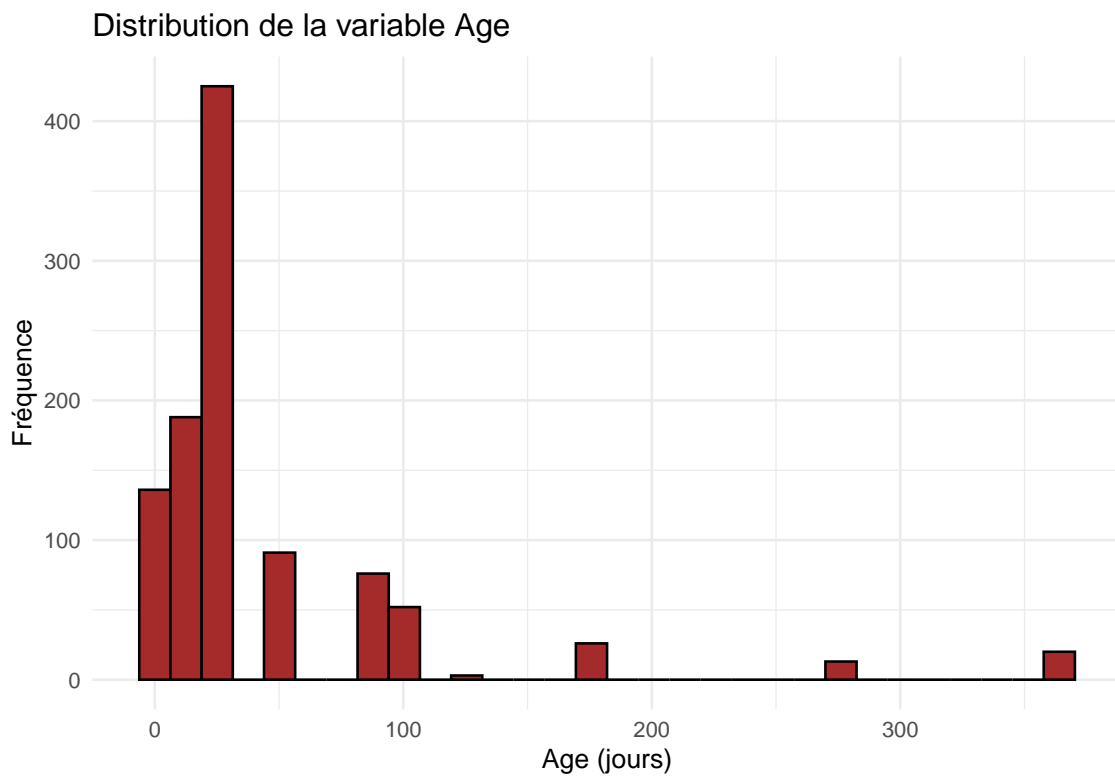
Histogramme pour la variable Coarse Aggregate



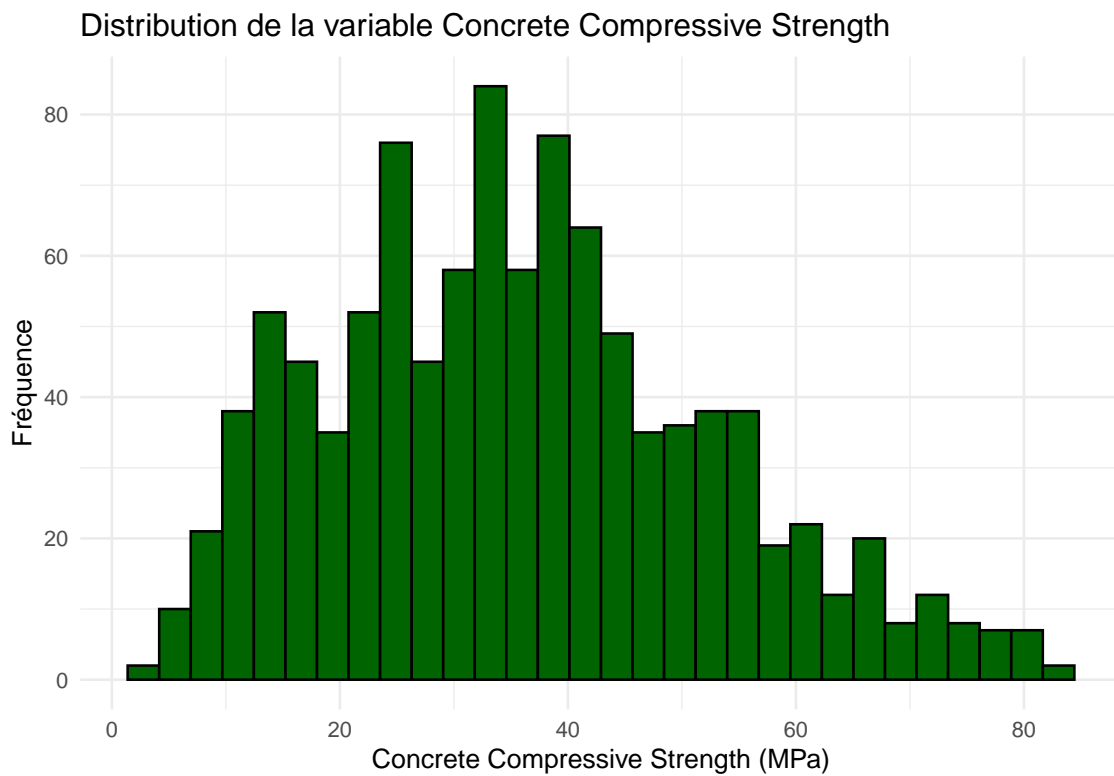
Histogramme pour la variable Fine Aggregate



Histogramme pour la variable Age

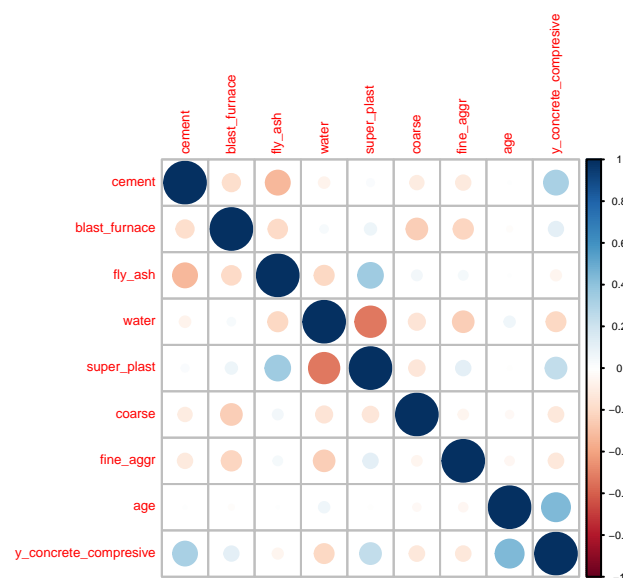


Histogramme pour la variable Concrete Compressive Strength



7.2 Analyse bivariee

Matrice de corrélation



test corrélation 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 corrélation 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 corrélation 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
```

7.3 Analyse multivariée

7.3.1 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	0.1686558	28.780302	1.293835e+01	9.5954664	2.990868e-01
## blast_furnace	2.6566581	1.857785	4.885852e+01	0.5815925	1.314741e+01
## fly_ash	13.6760355	7.204202	3.911677e-02	36.0802387	5.179921e+00
## water	31.8191398	1.395419	1.447269e+00	0.2201554	8.766687e+00
## super_plast	28.7359223	6.161874	3.532860e+00	2.7508660	1.368220e-01
## coarse	0.3655176	5.054896	3.019502e+01	4.9126729	2.975316e+01
## fine_aggr	14.5682217	3.499218	1.522624e-04	27.8593901	1.478283e+01
## age	6.8599228	6.341809	2.876323e+00	12.9252642	2.793407e+01
## y_concrete_compressive	1.1499265	39.704494	1.123899e-01	5.0743538	9.380241e-06

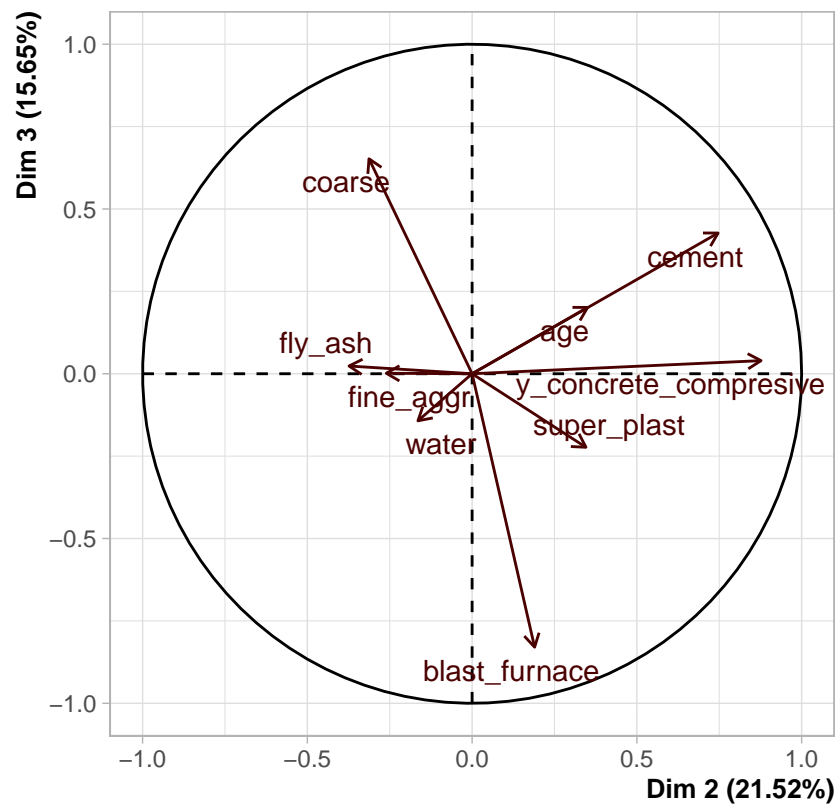
7.3.2 Tableau des \cos^2 des variables sur


```
# 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
## cement          0.003858358 0.55733497 1.822917e-01 0.100060379
## blast_furnace    0.060776682 0.03597629 6.883804e-01 0.006064777
## fly_ash          0.312868284 0.13951048 5.511263e-04 0.376240426
## water            0.727930233 0.02702251 2.039094e-02 0.002295754
## super_plast      0.657395101 0.11932563 4.977538e-02 0.028685702
## coarse           0.008361991 0.09788884 4.254254e-01 0.051228767
## fine_aggr        0.333278934 0.06776290 2.145264e-06 0.290514397
## age              0.156935267 0.12281010 4.052526e-02 0.134783114
## y_concrete_compressive 0.026307004 0.76888362 1.583491e-03 0.052914756
##               Dim.5
## cement          3.033202e-03
## blast_furnace    1.333351e-01
## fly_ash          5.253239e-02
## water            8.890773e-02
## super_plast      1.387586e-03
## coarse           3.017429e-01
## fine_aggr        1.499208e-01
## age              2.832946e-01
## y_concrete_compressive 9.513012e-08
```

7.3.3 partie D

```
plot.PCA(
  res_pca,
  axes = c(2, 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.quanti.sup = "#0000FF", # Couleur pour la variable quantitative supplém
  label = "all",
  title = "",
  addgrid.col = "#DDB688"   # Couleur de la grille
)
```



7.3.4 Partie E

```
# Représentation des variables sur le plan factoriel (axes 1 et 3)
plot.PCA(
  res_pca,
  axes = c(1, 3),           # On se concentre sur les 2 premiers axes
  choix = "var",            # Afficher les variables dans le plan factoriel
  col.var = "#4B0000",      # Couleur des variables alignée au style
  col.quanti.sup = "#0000FF", # Couleur pour la variable quantitative supplémentaire
  label = "all",
  title = "",
  addgrid.col = "#DDB688"   # Couleur de la grille
)
```

