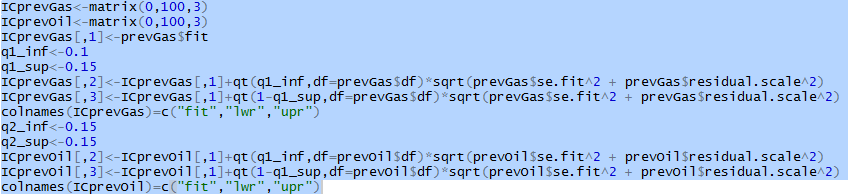
**Question 1:**

**Améliorer le score obtenu en se basant sur des intervalles de prédiction non symétriques contrairement à ceux calculés par défaut avec le modèle linéaire.**

On utilise des codes pour calculer des intervalles non symetriques pour Gas et pour Oil :



Et puis on va changer des valeurs de [q1\_inf, 1-q1\_sup], [q2\_inf, 1- q2\_sup], pour definir la confidence des intervalles.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q1\_inf | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| q1\_sup | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| q2\_inf | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 | 0.15 |
| q2\_sup | 0.15 | 0.15 | 0.15 | 1.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.25 |
| score | 4.99 | 5.07 | 6.83 | 5.21 | 5.20 | 5.22 | 6.47 | 5.09 | 4.90 | 4.82 |
| q1\_inf | 0.15 |  |  |  |  |  |  |  |  |  |
| q1\_sup | 0.15 |  |  |  |  |  |  |  |  |  |
| q2\_inf | 0.15 |  |  |  |  |  |  |  |  |  |
| q2\_sup | 0.30 |  |  |  |  |  |  |  |  |  |
| score | 5.08 |  |  |  |  |  |  |  |  |  |

Et puis on fait une boucle pour tester dans le condition meilleur dans les tests et change q1\_inf=0.15±0.05, q1\_sup=0.15±0.05, q2\_inf=0.15±0.05, q1\_sup=0.25±0.05 avec pas=0.01,et on obtenir le minimum score :

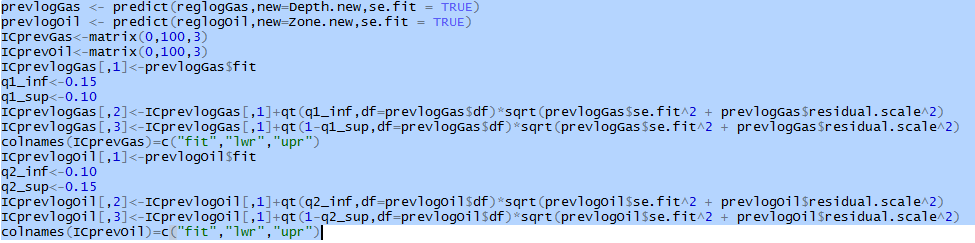
|  |  |
| --- | --- |
| q1\_inf | 0.12 |
| q1\_sup | 0.13 |
| q2\_inf | 0.15 |
| q2\_sup | 0.24 |
| score | 4.35 |

A cette condition, l’intervalle de confiance pour Gas est [0.12 ;0.87] confiance=0.75 et pour Oil [0.15 ;0.76] confiance=0.61

**Question 2 :**

**On a également considéré une transformation simple des variables à prédire (passage au log après translation des valeurs) et en ajoutant un terme quadratique pour la réponse prédite (régression linéaire multiple). A nouveau, obtenir le meilleur score possible en jouant sur la forme des intervalles de prédiction.**

On fait de meme comme question 1 et utilise des intervalles non symetriques de prediction pour obtenir le meilleur score :



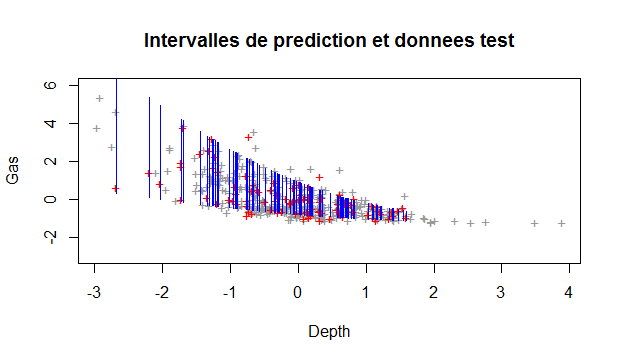
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q1\_inf | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| q1\_sup | 0.15 | 0.15 | 0.15 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |
| q2\_inf | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.05 | 0.00 | 0.00 | 0.00 |
| q2\_sup | 0.15 | 0.15 | 0.15 | 1.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 |
| score | 6.26 | 6.27 | 6.46 | 6.22 | 6.61 | 6.16 | 5.91 | 5.51 | 6.17 | 5.22 |
| q1\_inf | 0.15 |  |  |  |  |  |  |  |  |  |
| q1\_sup | 0.10 |  |  |  |  |  |  |  |  |  |
| q2\_inf | 0.00 |  |  |  |  |  |  |  |  |  |
| q2\_sup | 0.25 |  |  |  |  |  |  |  |  |  |
| score | 5.32 |  |  |  |  |  |  |  |  |  |

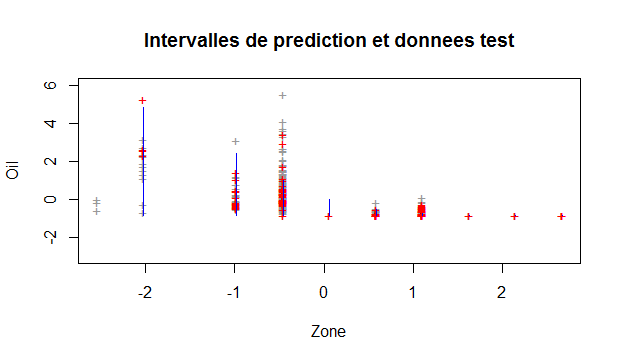
Et puis on fait une boucle pour tester dans le condition meilleur dans les tests et change dans les intervalles q1\_inf=0.15±0.05, q1\_sup=0.10±0.05, q2\_inf=0.00 +0.05, q1\_sup=0.20±0.05 avec pas=0.01,et on obtenir le minimum score :

|  |  |
| --- | --- |
| q1\_inf | 0.12 |
| q1\_sup | 0.09 |
| q2\_inf | 0.00 |
| q2\_sup | 0.23 |
| score | 5.02 |

A cette condition, l’intervalle de confiance pour Gas est [0.12 ;0.91] confiance=0.79 et pour Oil [0.00 ;0.77] confiance=0.77

Et des images sont affichees comme :





**Question 3 :**

**Considérer un modèle linéaire complet utilisant les réponses transformées et tous les prédicteurs.**

En utilisant le premiere transformation on peut obtenir un modele :

Supposant on a m predicteurs (pi ,i=1,...,m) pour Gas et meme pour Oil, donc on a des modeles :

Donc on a finalement modele de Gas et modele de Oil :

Donc finalement une modele comme :

sont tous coefficients.

**Question 4：**

**Essayer d’améliorer le résultat obtenu en essayant de construire un modèle réduit (on pourra utiliser le test de Fisher de comparaison de modèles emboîtés, utiliser la fonction anova de R).**

On a la modele reduit de Gas de 3 variables :

"Erosion\_PPLS..ft.", "Shot\_Density..shots.ft.", "Shot\_Total"

De meme, on a la modele reduit de Oil de 10 variables :

"Pressure\_PPLS..PSI.", "GR\_PPLS..API.", "Heat\_Flow..W.m2.", "Avg\_Treating\_Pressure..KPa.", "Max\_Treating\_pressure..KPa.", "Min\_Treating\_Pressure..KPa.", "Max\_Rate\_Slurry..bpm.", "ShutInPressure\_Fil..KPa.", "Shot\_Density..shots.ft.", "Shot\_Total"

d’apres ces variables, on construit le modeles de Gas et le modele de Oil, et puis on change des intervalles de confiance et calcule le score.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q1\_inf | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| q1\_sup | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.25 | 0.30 | 0.25 | 0.25 | 0.25 |
| q2\_inf | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.25 |
| q2\_sup | 0.15 | 0.15 | 0.15 | 1.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| score | 5.18 | 5.62 | 5.48 | 5.51 | 5.17 | 5.01 | 5.24 | 5.31 | 4.77 | 6.14 |
| q1\_inf | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |  |  |  |  |  |
| q1\_sup | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |  |  |  |  |  |
| q2\_inf | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |  |  |  |  |  |
| q2\_sup | 0.10 | 0.20 | 0.25 | 0.30 | 0.35 |  |  |  |  |  |
| score | 4.99 | 4.70 | 4.58 | 4.57 | 4.92 |  |  |  |  |  |

On utilise la meme methode comme des questions precedents , et on obtient :

|  |  |
| --- | --- |
| q1\_inf | 0.15 |
| q1\_sup | 0.20 |
| q2\_inf | 0.20 |
| q2\_sup | 0.28 |
| score | 4.18 |

A cette condition, l’intervalle de confiance pour Gas est [0.15 ;0.80] confiance=0.65 et pour Oil [0.20 ;0.72] confiance=0.52

Question 5:

Apres de construction de chaque matrice pour analyser, on d’abord utilise commande princomp pour obtenir des composantes principales de chaque matrice.

Et puis on choisisse des valeurs propres superieure a 1. Pour gas, il y 10 comps >1, puis on peut construire la regression de gas avec ses 10 premiere colonnes de projections.

Et puis on fait de meme pour la matrice de Oil, et on trouve qu’il y a 11 comps>1, et puis on peut construire la regression de oil avec ses 11 premieres colonnes de projections.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q1\_inf | 0.15 | 0.10 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |  |  |
| q1\_sup | 0.15 | 0.15 | 0.15 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |  |  |
| q2\_inf | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 | 0.15 | 0.15 |  |  |
| q2\_sup | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.10 | 0.20 |  |  |
| score | 4.00 | 3.59 | 3.51 | 3.47 | 3.64 | 3.99 | 3.64 | 3.66 |  |  |

Et puis on teste autour les intervalles qu’on obtient le meilleur score dans les tests precedents.

|  |  |
| --- | --- |
| q1\_inf | 0.06 |
| q1\_sup | 0.12 |
| q2\_inf | 0.15 |
| q2\_sup | 0.17 |
| score | 3.32 |

Donc d’apres cette modele, on obtient un score plus petit.

Pour calculer des intervalle