

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 symétriques pour Gas et pour Oil :

```
ICprevGas<-matrix(0,100,3)
ICprevOil<-matrix(0,100,3)
ICprevGas[,1]<-prevGas$fit
q1_inf<-0.1
q1_sup<-0.15
ICprevGas[,2]<-ICprevGas[,1]+qt(q1_inf,df=prevGas$df)*sqrt(prevGas$se.fit^2 + prevGas$residual.scale^2)
ICprevGas[,3]<-ICprevGas[,1]+qt(1-q1_sup,df=prevGas$df)*sqrt(prevGas$se.fit^2 + prevGas$residual.scale^2)
colnames(ICprevGas)=c("fit", "lwr", "upr")
q2_inf<-0.15
q2_sup<-0.15
ICprevOil[,2]<-ICprevOil[,1]+qt(q1_inf,df=prevOil$df)*sqrt(prevOil$se.fit^2 + prevOil$residual.scale^2)
ICprevOil[,3]<-ICprevOil[,1]+qt(1-q1_sup,df=prevOil$df)*sqrt(prevOil$se.fit^2 + prevOil$residual.scale^2)
colnames(ICprevOil)=c("fit", "lwr", "upr")
```

Et puis on va changer des valeurs de [q1_inf, 1-q1_sup], [q2_inf, 1- q2_sup], pour définir la confiance 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 obtient 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 :

```
prevlogGas <- predict(reglogGas,new=Depth.new,se.fit = TRUE)
prevlogoil <- predict(reglogoil,new=Zone.new,se.fit = TRUE)
icprevGas<-matrix(0,100,3)
icprevoil<-matrix(0,100,3)
icprevlogGas[,1]<-prevlogGas$fit
q1_inf<-0.15
q1_sup<-0.10
icprevlogGas[,2]<-icprevlogGas[,1]+qt(q1_inf,df=prevlogGas$df)*sqrt(prevlogGas$se.fit^2 + prevlogGas$residual.scale^2)
icprevlogGas[,3]<-icprevlogGas[,1]+qt(1-q1_sup,df=prevlogGas$df)*sqrt(prevlogGas$se.fit^2 + prevlogGas$residual.scale^2)
colnames(icprevGas)=c("fit","lwr","upr")
icprevlogoil[,1]<-prevlogoil$fit
q2_inf<-0.10
q2_sup<-0.15
icprevlogoil[,2]<-icprevlogoil[,1]+qt(q2_inf,df=prevlogoil$df)*sqrt(prevlogoil$se.fit^2 + prevlogoil$residual.scale^2)
icprevlogoil[,3]<-icprevlogoil[,1]+qt(1-q2_sup,df=prevlogoil$df)*sqrt(prevlogoil$se.fit^2 + prevlogoil$residual.scale^2)
colnames(icprevoil)=c("fit","lwr","upr")
```

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									

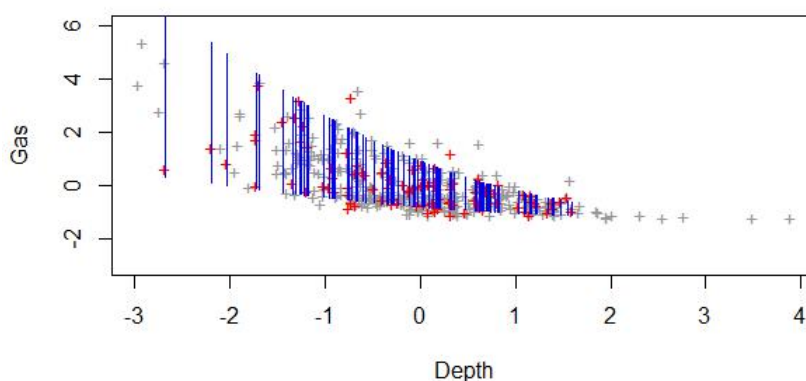
Et puis on fait une boucle pour tester dans le condition meilleur dans les tests et change dans les intervalles $q1_inf=0.15\pm0.05$, $q1_sup=0.10\pm0.05$, $q2_inf=0.00+0.05$, $q1_sup=0.20\pm0.05$ avec pas=0.01,et on obtenir le minimum score :

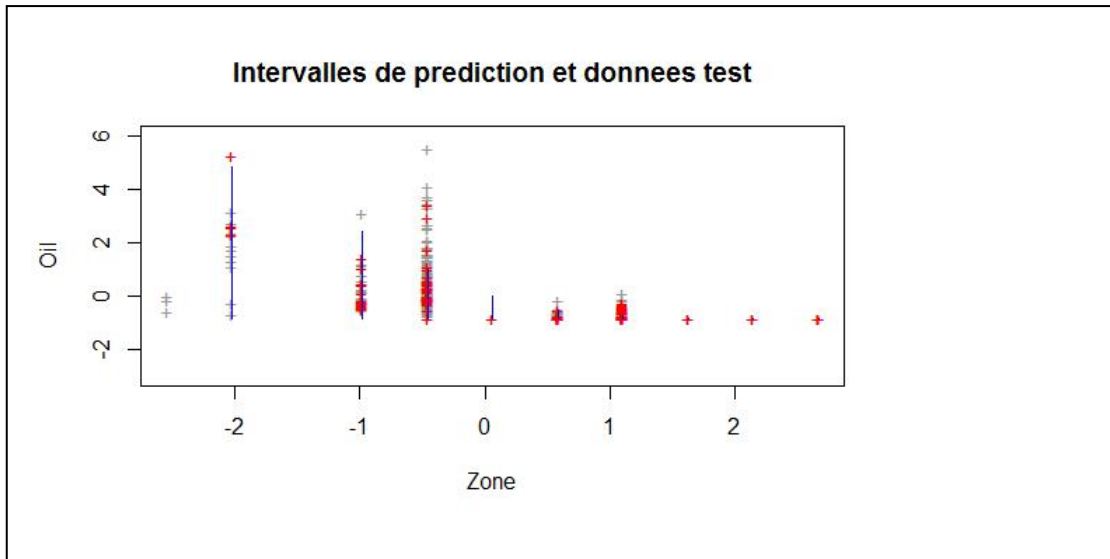
q1_inf	0.12
q1_sup	0.09
q2_inf	0.00
q2_sup	0.23
score	5.02

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 :

Intervalles de prediction et donnees test





Question 3 :

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

En utilisant la première transformation on peut obtenir un modèle :

Supposant on a m prédicteurs ($p_i, i=1, \dots, m$) pour Gas et même pour Oil, donc on a des modèles :

$$\text{Gas1} = \sum_1^m \alpha_i p_i$$

$$\text{Oil1} = \sum_1^m \alpha_i' p_i$$

$$\text{Gas2} = \sum_1^m \gamma_i \exp(\beta_{i1} p_i + \beta_{i2} p_i^2) + \text{seuil.G}_i$$

$$\text{Oil2} = \sum_1^m \gamma_i' \exp(\beta_{i1}' p_i + \beta_{i2}' p_i^2) + \text{seuil.O}_i$$

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

$$\text{Gas} = \theta_1 \text{Gas1} + \theta_2 \text{Gas2}$$

$$\text{Oil} = \theta_1' \text{Gas1} + \theta_2' \text{Gas2}$$

Donc finalement un modèle comme :

$$\text{Gas} = \sum_1^m \alpha_i p_i + \sum_1^m \gamma_i \exp(\beta_{i1} p_i + \beta_{i2} p_i^2) + \rho \text{seuil.G}$$

$$\text{Oil} = \sum_1^m \alpha_i' p_i + \sum_1^m \gamma_i' \exp(\beta_{i1}' p_i + \beta_{i2}' p_i^2) + \rho' \text{seuil.O}$$

$\alpha_i, \gamma_i, \beta_{i1}, \beta_{i2}, \rho, \alpha_i', \gamma_i', \beta_{i1}', \beta_{i2}', \rho'$ 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 modèle réduit de Gas de 3 variables :

"Erosion_PPLS..ft.", "Shot_Density..shots.ft.", "Shot_Total"

De même, on a la modèle réduit 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'après ces variables, on construit le modèles de Gas et le modèle 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 même méthode comme des questions précédentes, 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

