

Projet MRR

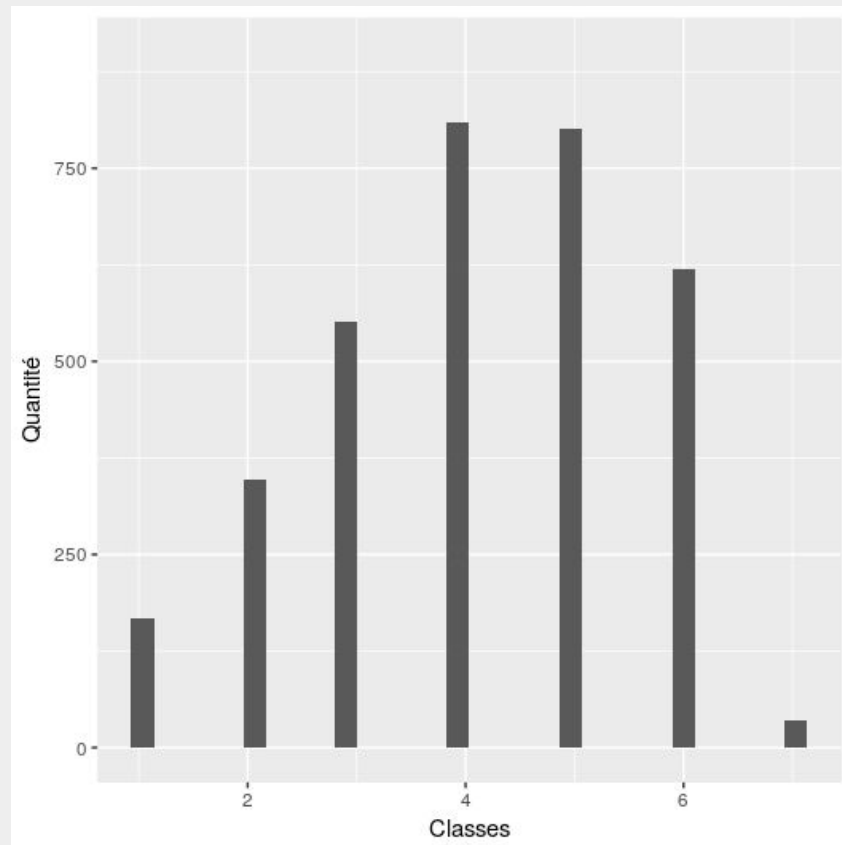
Groupe 16

BELHOMME Hugo,
DI MARTINO Sacha,
FRAISE Aymeric

Notre base de données

“SkillCraft1 Master Table Dataset”

- 3395 parties de Starcraft II
- 3395 joueurs



Nos variables

20 variables :

- Identifiant de la partie
- Informations sur le joueur (4, dont son rang)
- Mesures d'activité en jeu (5)
- Mesures d'actions en jeu (10)

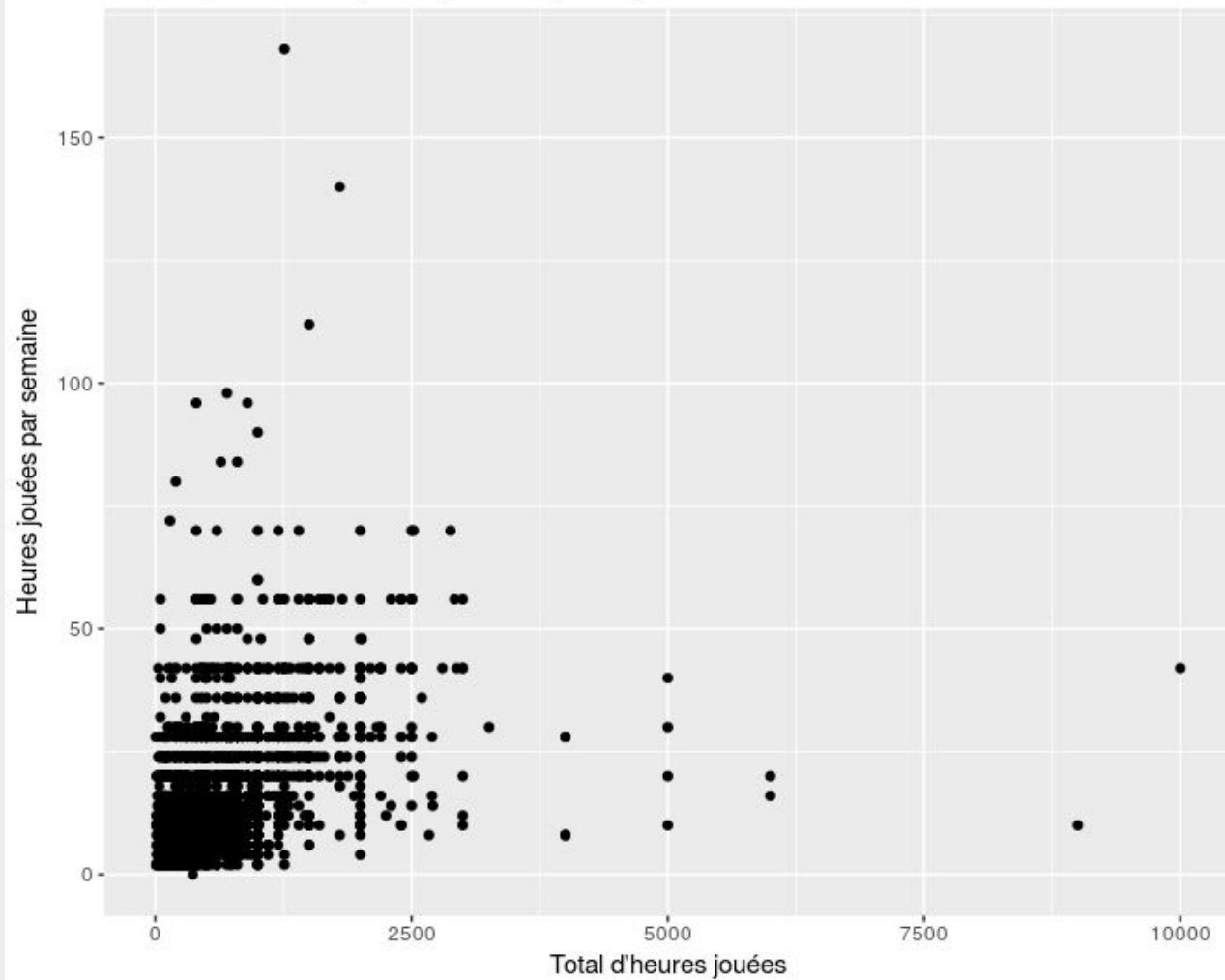
Notre objectif

Estimer le rang d'un joueur à partir des 18 autres variables.

Traitement des données

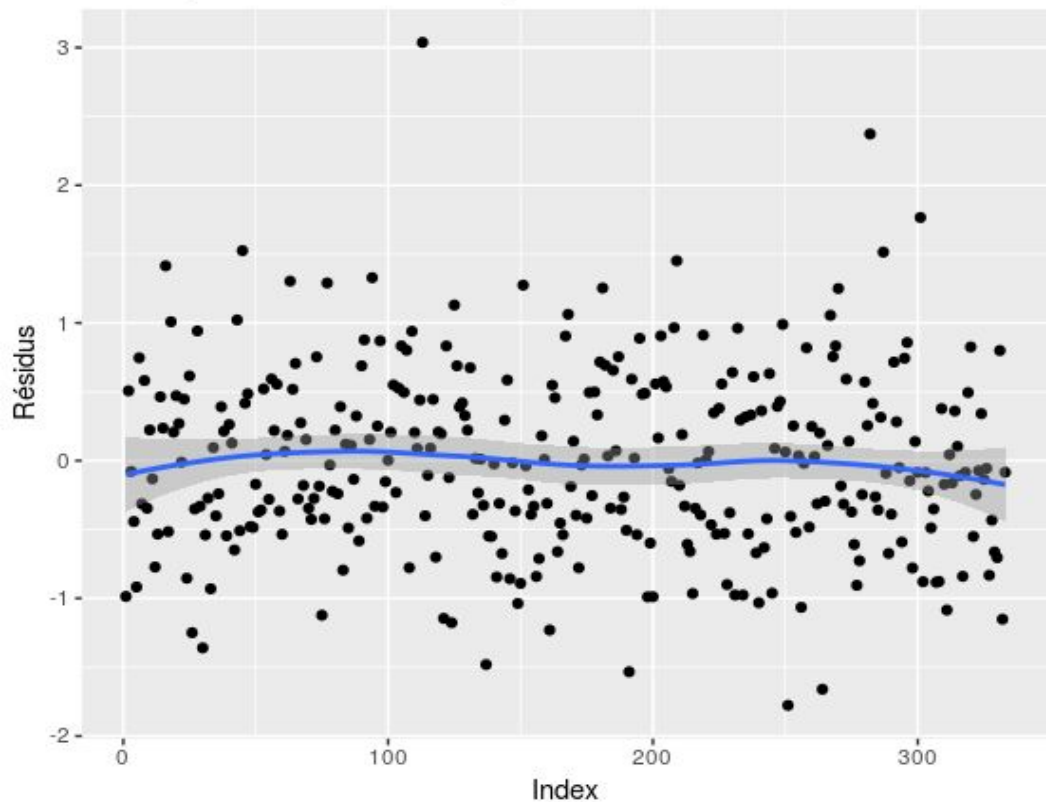
- Données manquantes pour le dernier rang
- Données renseignées par les participants
⇒ perte de 2% des données

Affichage des temps de jeu des participants



Approche par régression linéaire simple

Affichage des résidus sur le jeu d'entraînement



Residuals:

Min	1Q	Median	3Q	Max
-1.7734	-0.6288	0.0434	0.6920	3.1802

Coefficients:

	Est	Std. Error	t value	Pr(> t)
(Intercept)	4.18486	0.01793	233.420	< 2e-16 ***
APM	-0.11176	0.11454	-0.976	0.3293
SelectByHotkeys	0.29928	0.06870	4.357	1.36e-05 ***
AssignToHotkeys	0.20492	0.02343	8.748	< 2e-16 ***
UniqueHotkeys	0.08152	0.02062	3.953	7.89e-05 ***
MinimapAttacks	0.18407	0.01935	9.513	< 2e-16 ***
MinimapRightClicks	0.01336	0.02030	0.658	0.5105
NumberOfPACs	0.29512	0.06664	4.429	9.82e-06 ***
GapBetweenPACs	-0.17060	0.02714	-6.285	3.74e-10 ***
ActionLatency	-0.41918	0.04130	-10.149	< 2e-16 ***
ActionsInPAC	0.07993	0.04971	1.608	0.1079
TotalMapExplored	-0.03297	0.02434	-1.354	0.1758
WorkersMade	0.10786	0.02031	5.310	1.18e-07 ***
UniqueUnitsMade	-0.05498	0.02318	-2.371	0.0178 *
ComplexUnitsMade	0.04111	0.02377	1.730	0.0838 .
ComplexAbilitiesUsed	0.02113	0.02279	0.927	0.3538

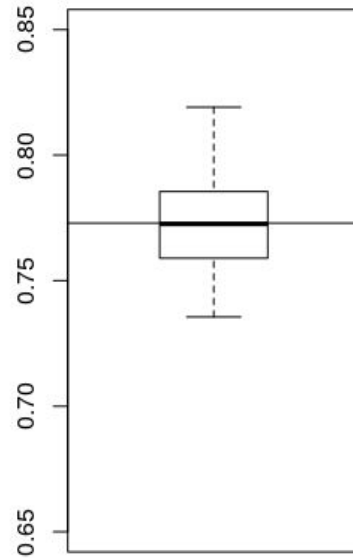
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9903 on 3035 degrees of freedom
Multiple R-squared: 0.5719, Adjusted R-squared: 0.5697
F-statistic: 270.2 on 15 and 3035 DF, p-value: < 2.2e-16

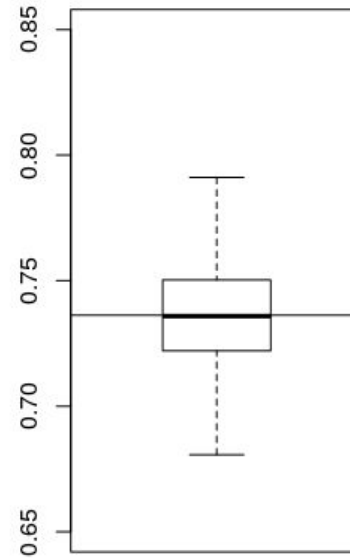
Arrondissement des valeurs

Erreur moyenne absolue

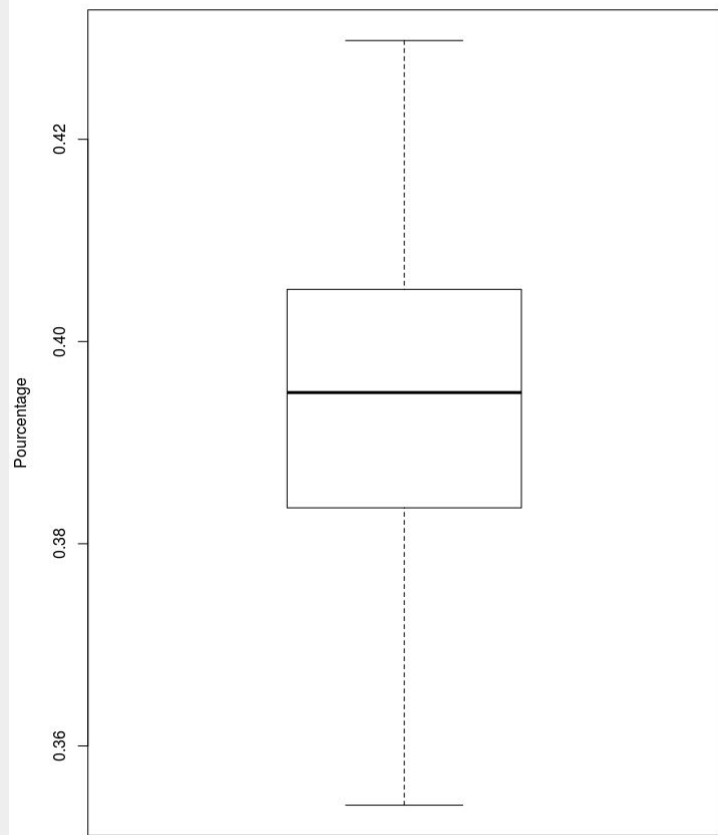
Prédictions continues



Prédictions arrondies

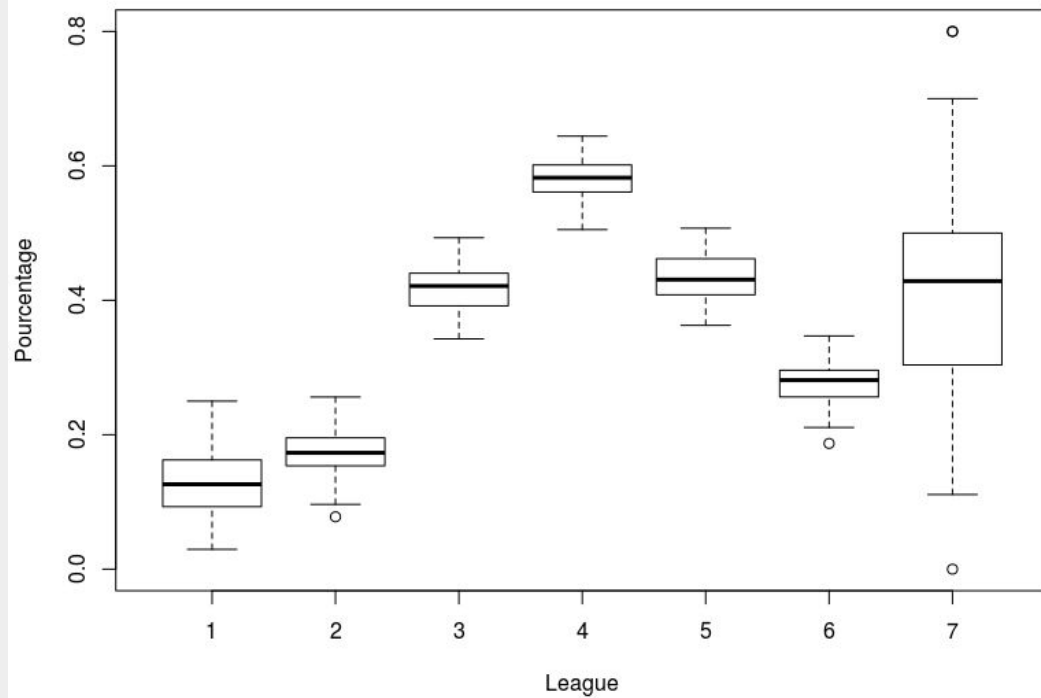


Pourcentages de réponses justes sur 100 régressions linéaires simples

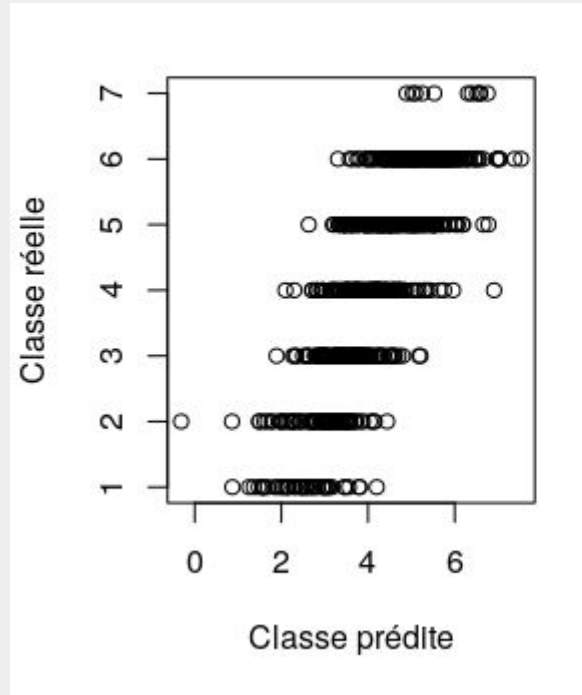


(Pourcentage moyen : 39%)

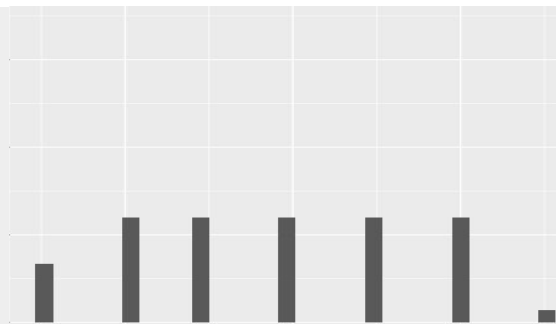
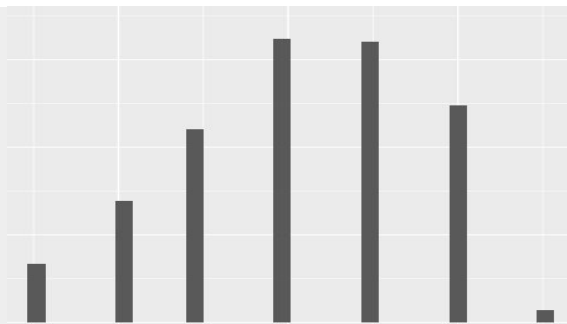
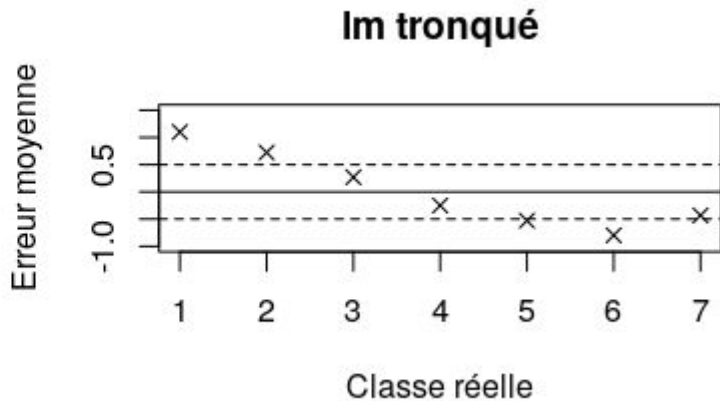
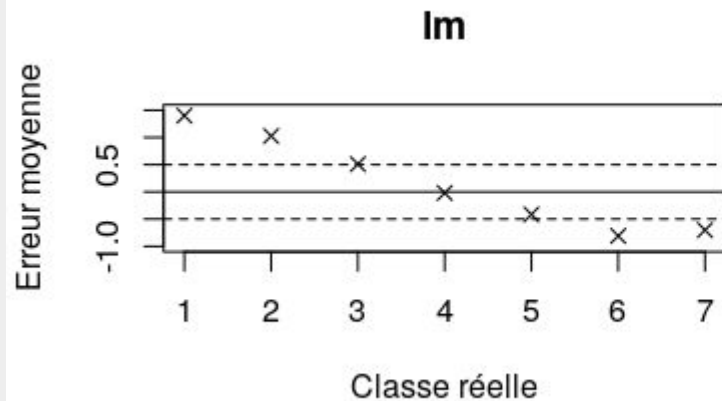
Pourcentage de réponses justes pour chaque classe



Écrasement des prédictions

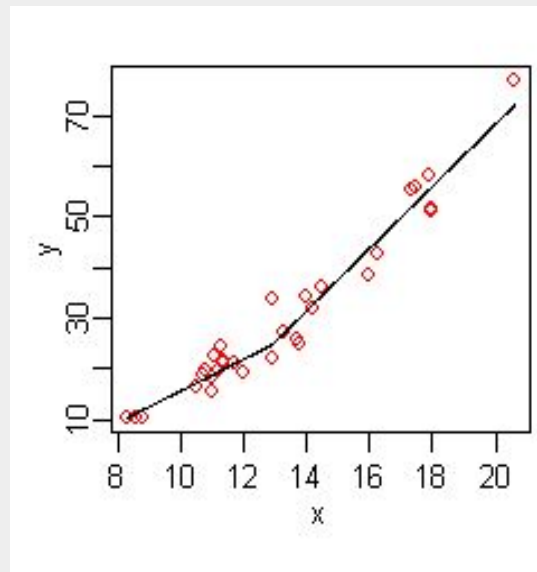
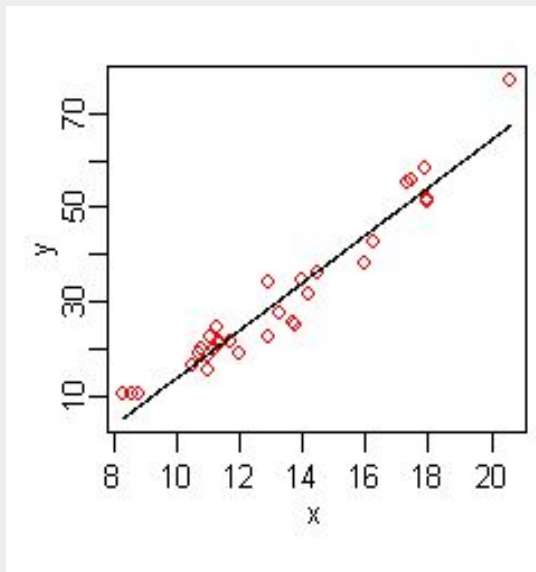


Écrasement des prédictions

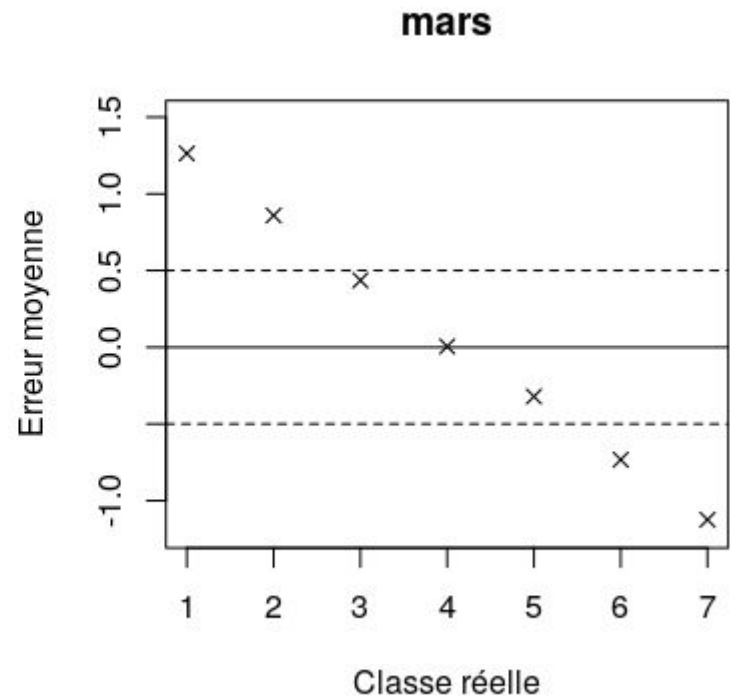
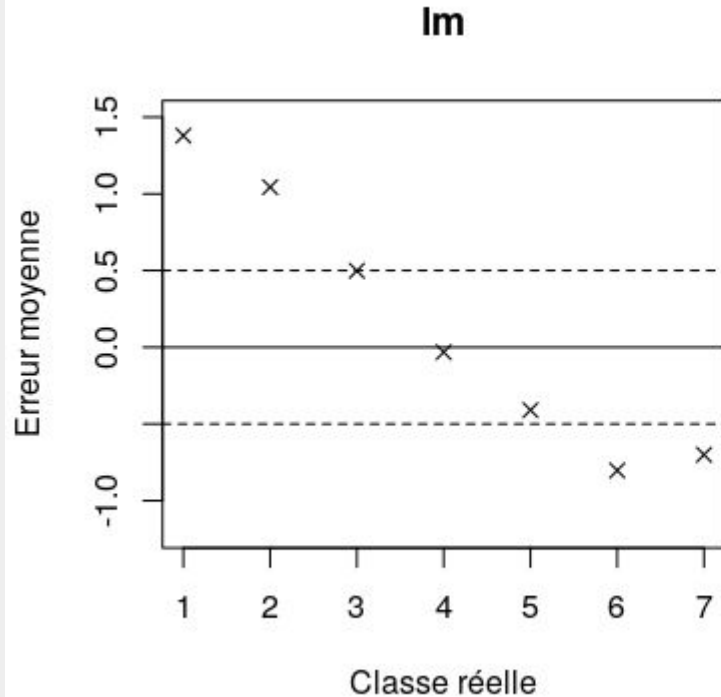


Écrasement des prédictions

MARS
(Multivariate Adaptive Regression Splines)



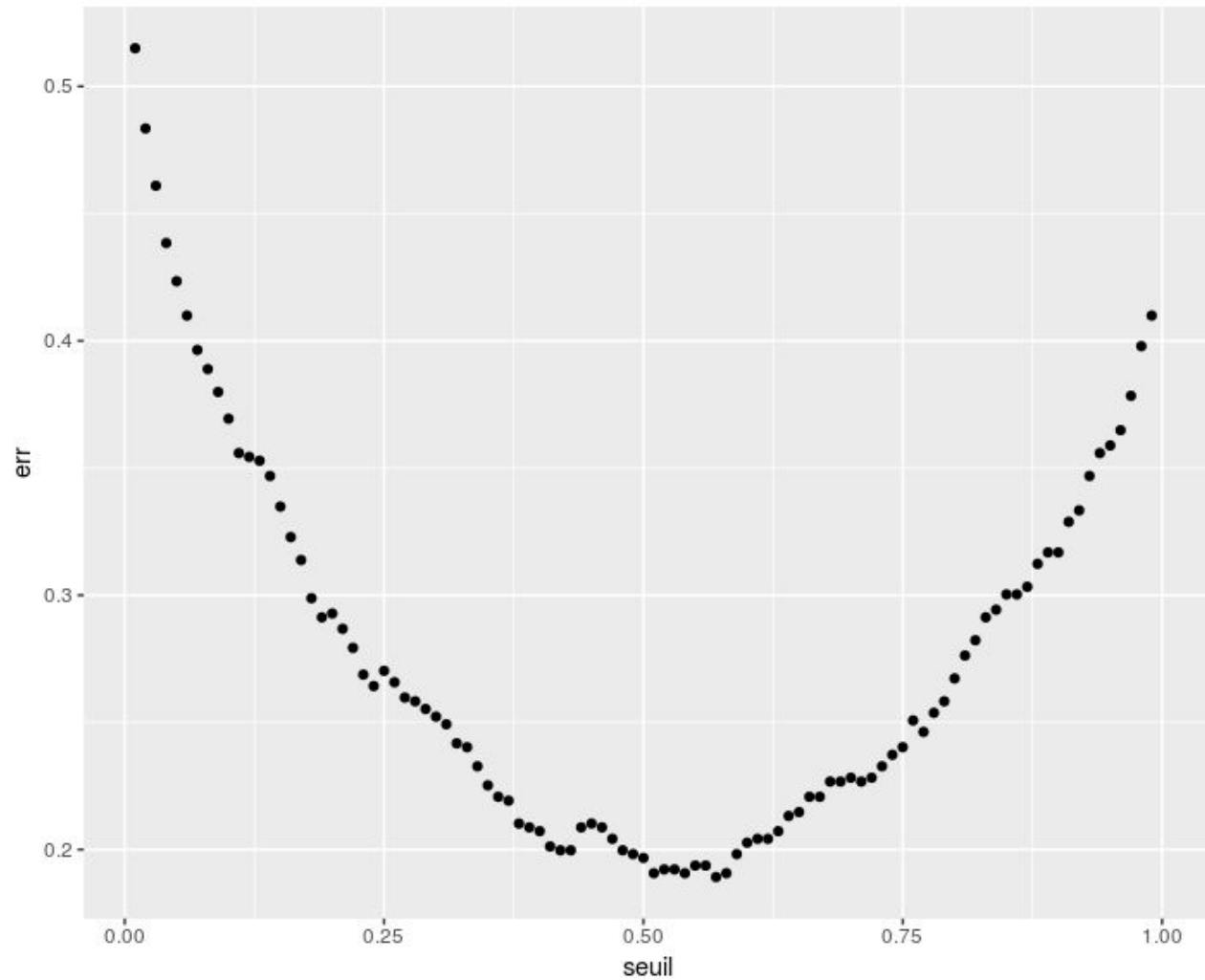
Écrasement des prédictions



Approche par régression logistique

- Séparation des 7 rangs en 2 classes, régression logistique :
 - *LOW* : Bronze + Argent + Or + Platine
 - *HIGH* : Diamant + Master + Grand Master
- Régression linéaire sur chaque classe obtenue

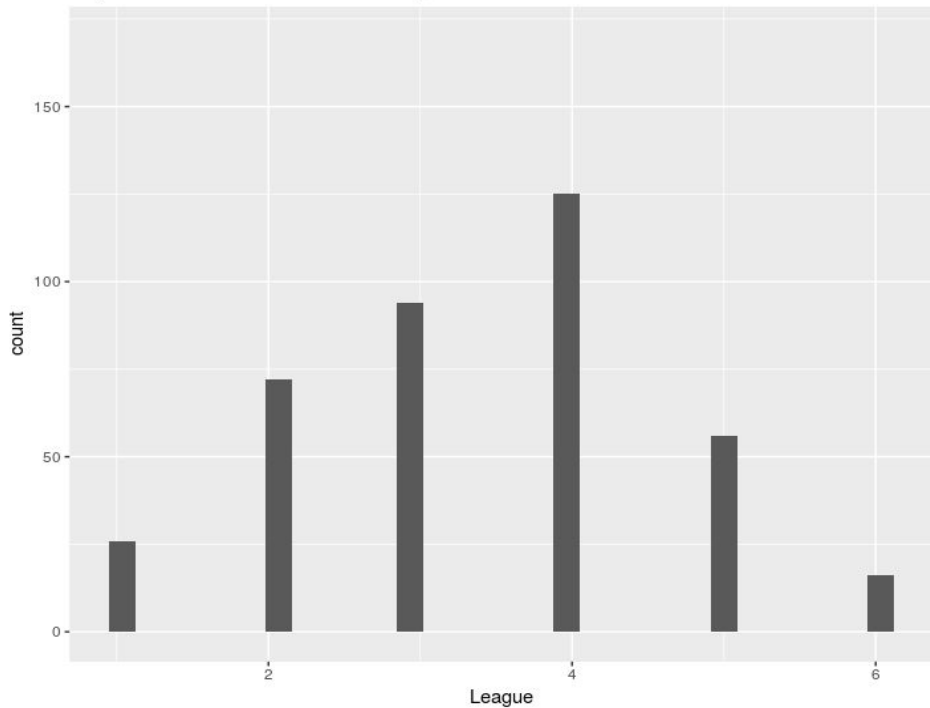
Évolution de l'erreur en fonction du seuil de sélection



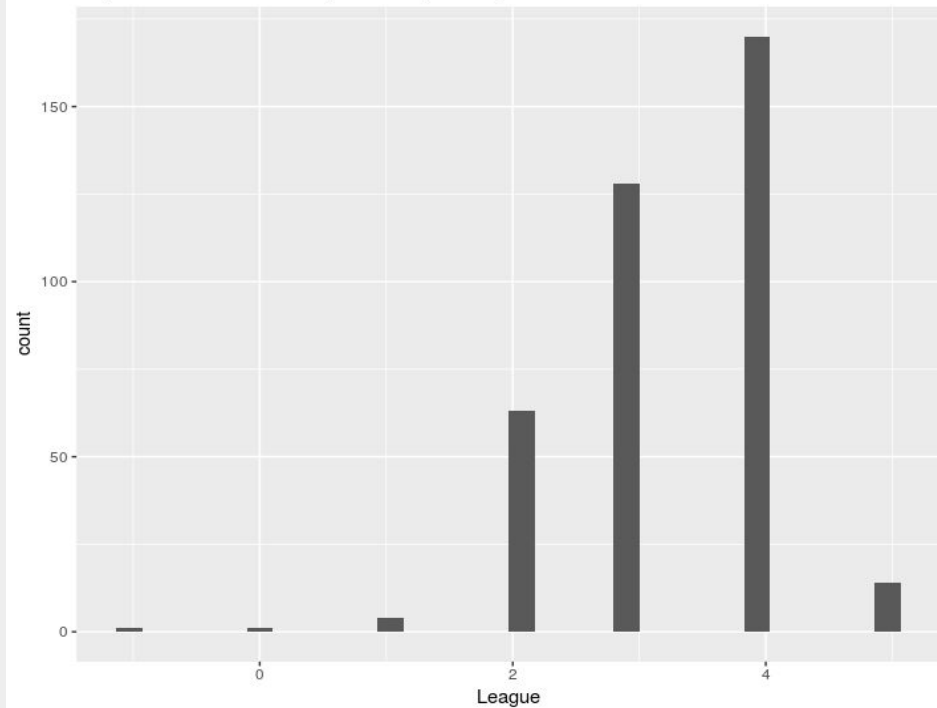
Seuil : [0.44, 0.58]

Erreur moyenne : 17.2%

Répartition des individus dans le jeu de test LOW

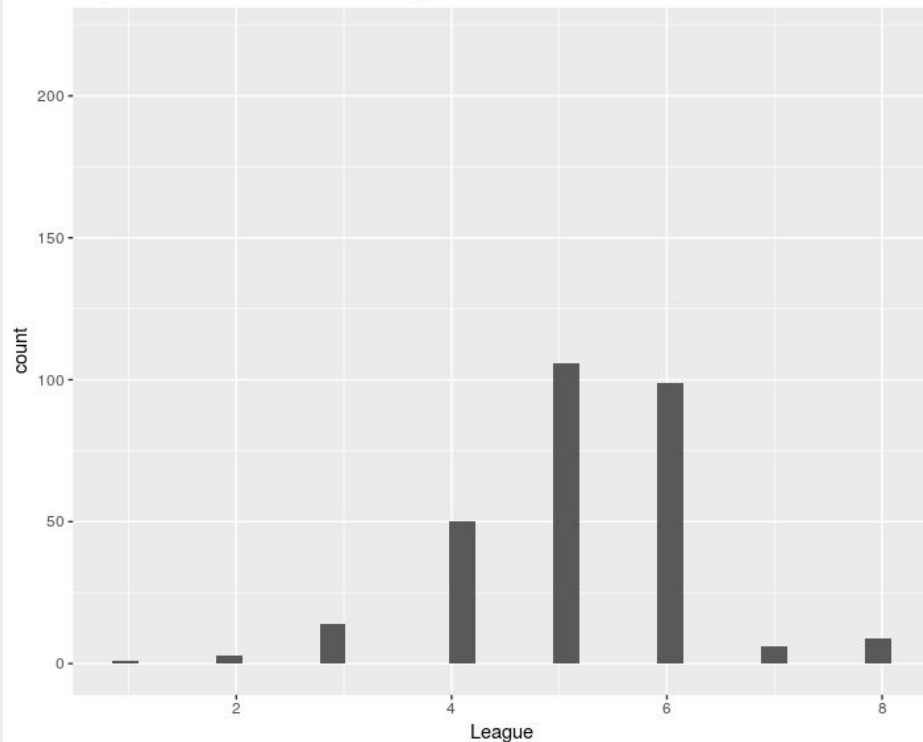


Répartition des valeurs prédites pour le jeu de test LOW

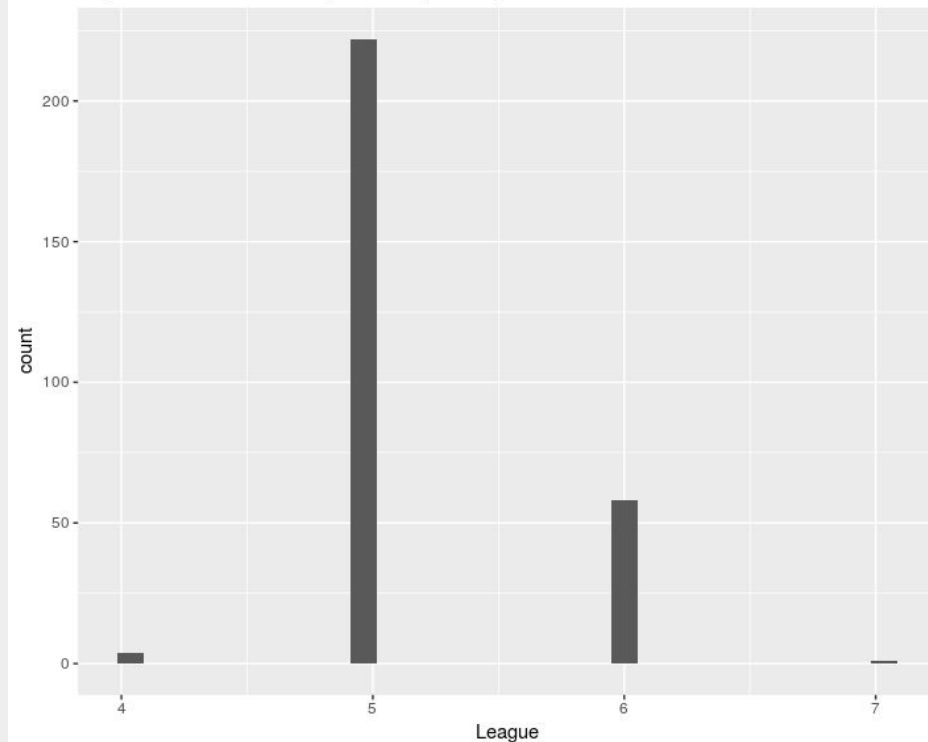


→ Taux d'erreur moyen : 63%

Répartition des individus dans le jeu de test HIGH



Répartition des valeurs prédites pour le jeu de test HIGH



→ Taux d'erreur moyen : 56%

Approche par régression logistique

Impact des variables sur les 2 classes

Coefficients (**LOW**) :

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.818e+00	5.915e-01	9.836	< 2e-16 ***
Age	-6.458e-03	7.545e-03	-0.856	0.392248
HoursPerWeek	5.756e-03	1.906e-03	3.020	0.002585 **
TotalHours	1.531e-04	3.608e-05	4.244	2.37e-05 ***
APM	-9.796e-03	3.113e-03	-3.147	0.001692 **
SelectByHotkeys	7.637e+01	1.754e+01	4.355	1.45e-05 ***
AssignToHotkeys	4.489e+02	1.303e+02	3.445	0.000593 ***
UniqueHotkeys	3.484e-02	1.226e-02	2.843	0.004555 **
MinimapAttacks	5.151e+02	1.196e+02	4.308	1.79e-05 ***
MinimapRightClicks	3.801e+01	6.470e+01	0.587	0.556999
NumberOfPACs	1.925e+02	8.827e+01	2.180	0.029429 *
GapBetweenPACs	-1.348e-02	3.462e-03	-3.894	0.000104 ***
ActionLatency	-2.715e-02	5.504e-03	-4.933	9.33e-07 ***
ActionsInPAC	8.258e-02	5.132e-02	1.609	0.107888
TotalMapExplored	-1.110e-03	4.425e-03	-0.251	0.802018
WorkersMade	1.611e+02	4.756e+01	3.388	0.000730 ***
UniqueUnitsMade	1.207e-02	1.691e-02	0.713	0.475709
ComplexUnitsMade	-2.403e+01	2.623e+02	-0.092	0.927016
ComplexAbilitiesUsed	-2.014e+01	1.000e+02	-0.201	0.840486

Coefficients (**HIGH**) :

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.013e+00	5.487e-01	5.491	4.69e-08 ***
Age	3.891e-03	5.814e-03	0.669	0.503500
HoursPerWeek	-1.094e-02	2.777e-03	-3.938	8.59e-05 ***
TotalHours	1.064e-03	8.552e-05	12.440	< 2e-16 ***
APM	1.025e-02	4.205e-03	2.438	0.014864 *
SelectByHotkeys	1.678e+01	2.768e+01	0.606	0.544602
AssignToHotkeys	1.030e+03	1.917e+02	5.373	8.98e-08 ***
UniqueHotkeys	1.837e-02	1.247e-02	1.473	0.140908
MinimapAttacks	1.432e+03	3.212e+02	4.459	8.85e-06 ***
MinimapRightClicks	1.292e+01	9.375e+01	0.138	0.890411
NumberOfPACs	1.297e+02	1.317e+02	0.985	0.324823
GapBetweenPACs	-6.812e-03	1.956e-03	-3.482	0.000512 ***
ActionLatency	-1.141e-02	2.843e-03	-4.015	6.25e-05 ***
ActionsInPAC	-9.067e-02	5.766e-02	-1.572	0.116047
TotalMapExplored	-1.110e-02	4.831e-03	-2.297	0.021742 *
WorkersMade	2.612e+02	6.721e+01	3.887	0.000106 ***
UniqueUnitsMade	-3.073e-02	1.773e-02	-1.733	0.083262 .
ComplexUnitsMade	6.226e+02	3.698e+02	1.684	0.092460 .
ComplexAbilitiesUsed	9.445e+01	1.535e+02	0.616	0.538302

Approche par régression logistique

Impact des variables sur les 2 classes

Coefficients (*LOW*) :

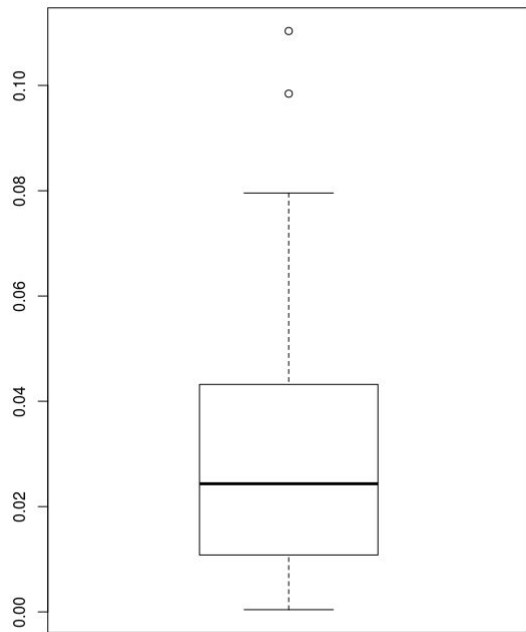
	Estimate	Std. Error	t value	Pr(> t)	
HoursPerWeek	5.756e-03	1.906e-03	3.020	0.002585	**
TotalHours	1.531e-04	3.608e-05	4.244	2.37e-05	***
SelectByHotkeys	7.637e+01	1.754e+01	4.355	1.45e-05	***
AssignToHotkeys	4.489e+02	1.303e+02	3.445	0.000593	***
MinimapAttacks	5.151e+02	1.196e+02	4.308	1.79e-05	***
GapBetweenPACs	-1.348e-02	3.462e-03	-3.894	0.000104	***
ActionLatency	-2.715e-02	5.504e-03	-4.933	9.33e-07	***
WorkersMade	1.611e+02	4.756e+01	3.388	0.000730	***

Coefficients (*HIGH*) :

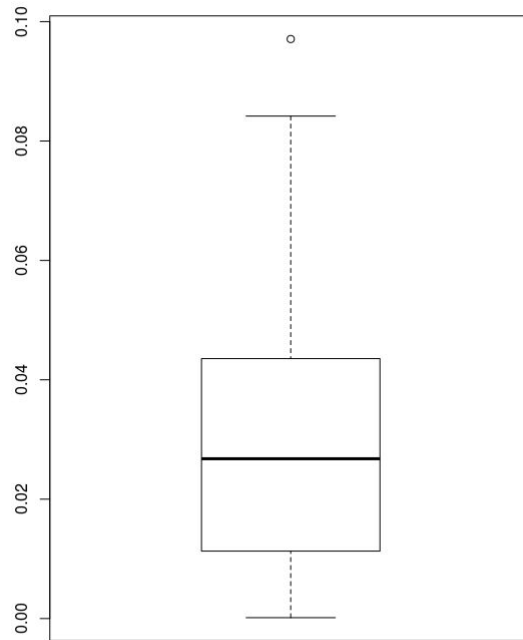
	Estimate	Std. Error	t value	Pr(> t)	
HoursPerWeek	-1.094e-02	2.777e-03	-3.938	8.59e-05	***
TotalHours	1.064e-03	8.552e-05	12.440	< 2e-16	***
SelectByHotkeys	1.678e+01	2.768e+01	0.606	0.544602	
AssignToHotkeys	1.030e+03	1.917e+02	5.373	8.98e-08	***
MinimapAttacks	1.432e+03	3.212e+02	4.459	8.85e-06	***
GapBetweenPACs	-6.812e-03	1.956e-03	-3.482	0.000512	***
ActionLatency	-1.141e-02	2.843e-03	-4.015	6.25e-05	***
WorkersMade	2.612e+02	6.721e+01	3.887	0.000106	***

Approche par régressions backward, forward et both

Boxplot des RMSE de 100 régressions linéaires simples



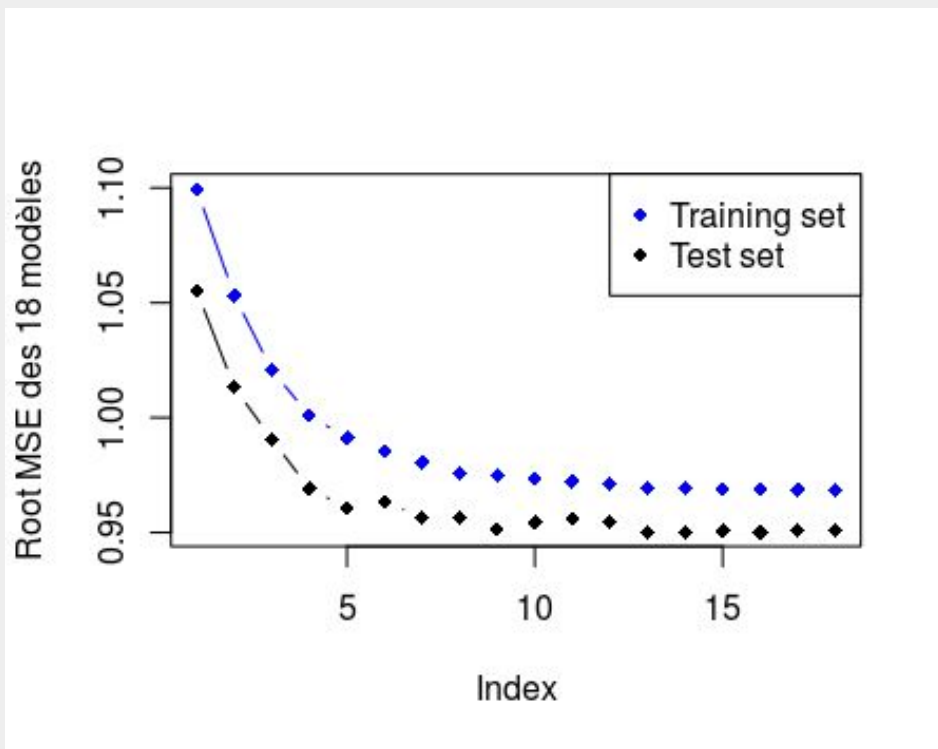
Boxplot des RMSE de 100 régressions backward



3 Variables retirées :

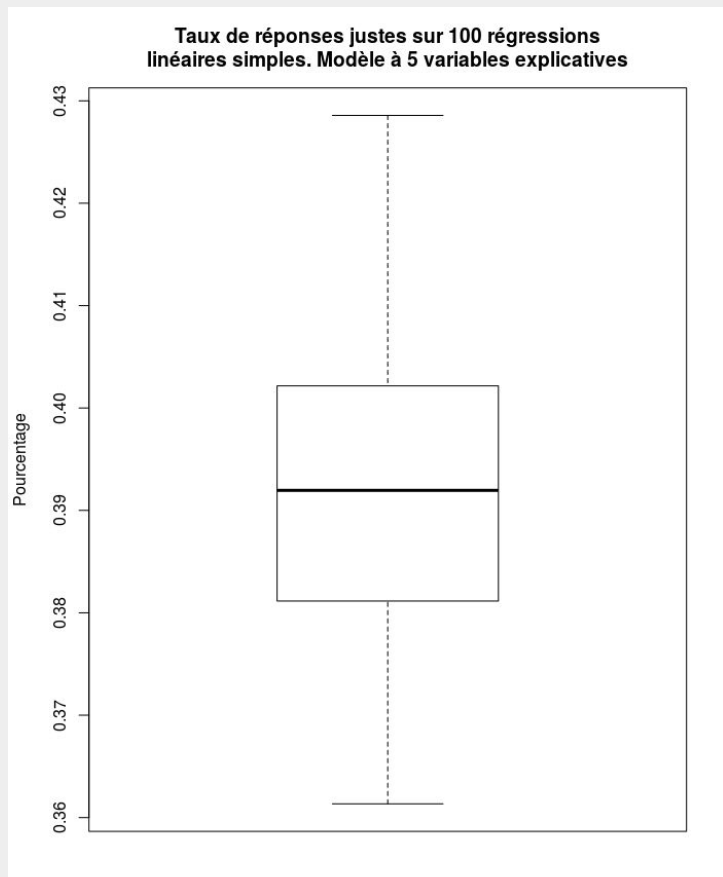
- HoursPerWeek
- MinimapRightClicks
- ComplexAbilitiesUsed

Autre moyen de sélection de variables



- **Fonction regsubsets (librairie 'Leaps')**
- Choix sur la RMSE la plus faible
- Modèles 5 à 18 se valent : on choisit le 5.

Autre moyen de sélection de variables



5 variables retenues :

- TotalHours
- APM
- AssignToHotkeys
- MinimapAttacks
- ActionLatency

→ Pourcentage moyen : 39%

Régression logistique multinomiale

- **Fonction multinom (librairie 'nnet')**
- Modalité de référence : la ligue 4 (Platine)
- 9 variables retenues :
 - HoursPerWeek
 - TotalHours
 - APM
 - AssignToHotkeys
 - UniqueHotkeys
 - MinimapAttacks
 - GapBetweenPACs
 - ActionLatency
 - ActionsInPAC

Régression logistique multinomiale

	1	2	3	4	5	6	7
1	54	30	11	4	0	0	0
2	61	117	73	43	4	0	0
3	33	94	148	89	31	0	0
4	16	98	255	429	237	58	0
5	2	7	56	205	361	229	1
6	1	1	9	40	169	330	32
7	0	0	0	0	0	3	2

Matrice de confusion

Résultats :

- 43% de réponses justes
- 88% de réponses justes à +/- un rang d'écart

Bilan

- Nous n'avons pas réussi à trouver de modèle capable d'estimer précisément le rang d'un joueur.

Bilan

- Nous n'avons pas réussi à trouver de modèle capable d'estimer précisément le rang d'un joueur.
- La régression logistique nous a permis de correctement scinder les joueurs en deux groupes.

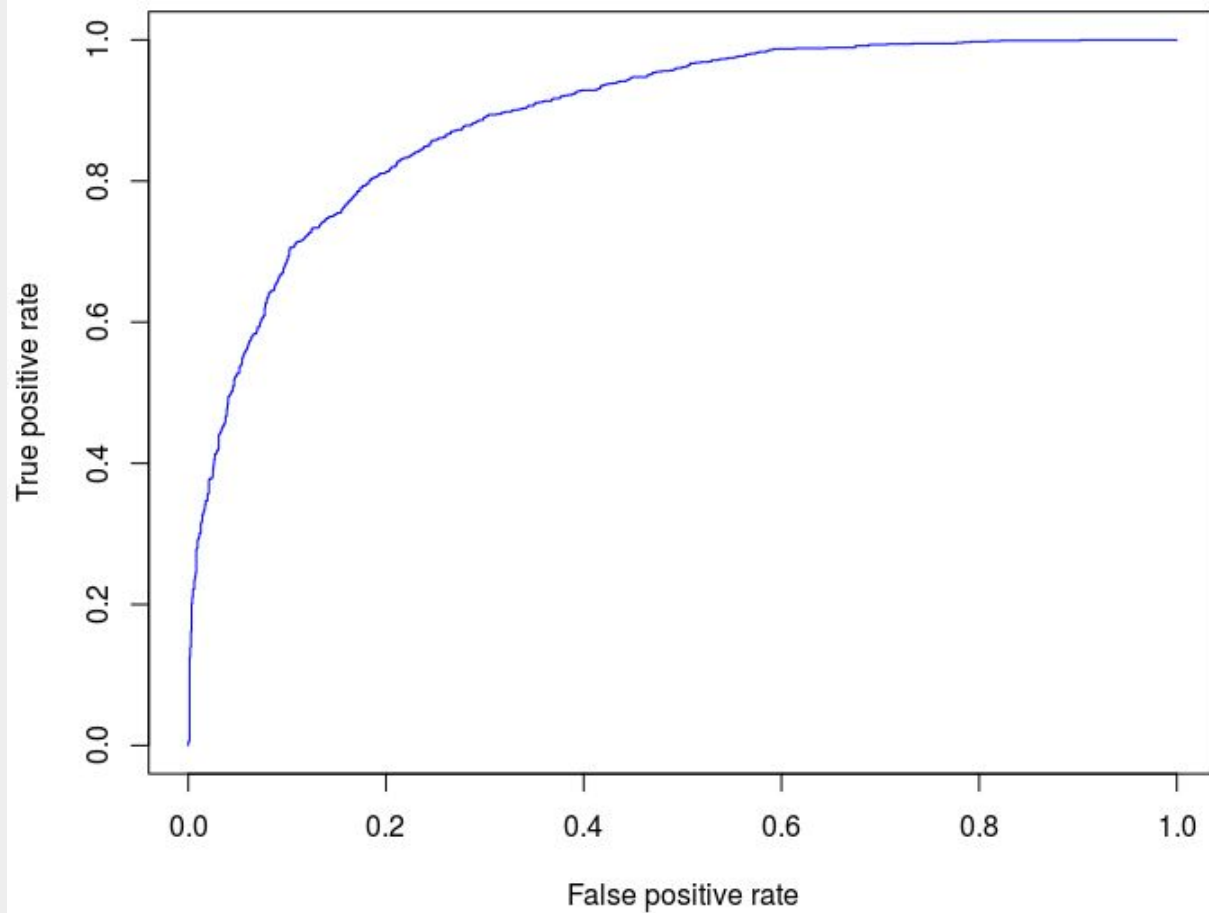
Bilan

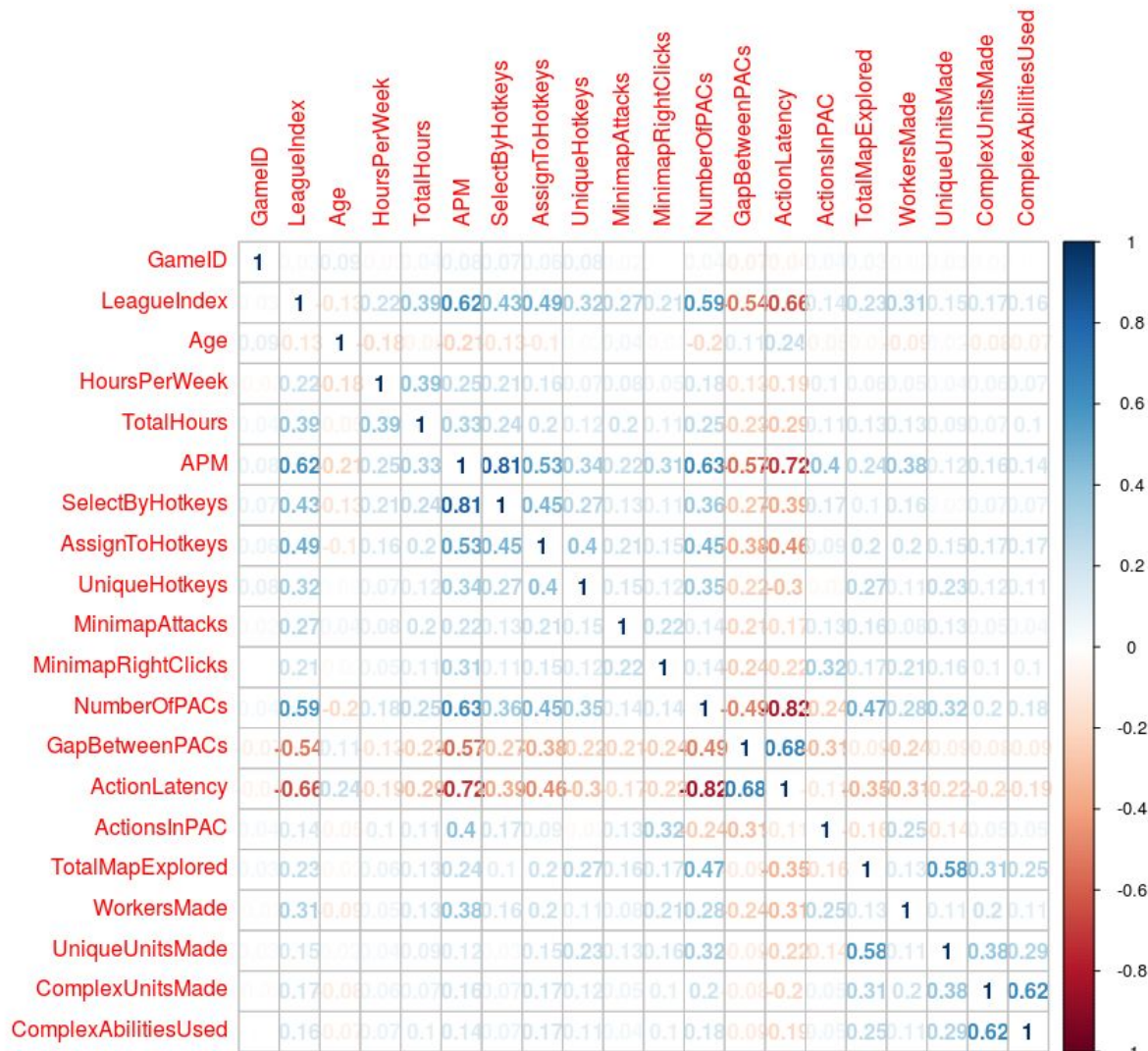
- Nous n'avons pas réussi à trouver de modèle capable d'estimer précisément le rang d'un joueur.
- La régression logistique nous a permis de correctement scinder les joueurs en deux groupes.
- En s'accordant un rang d'écart, la régression logistique multinomiale est satisfaisante.

Bilan

- Nous n'avons pas réussi à trouver de modèle capable d'estimer précisément le rang d'un joueur.
- La régression logistique nous a permis de correctement scinder les joueurs en deux groupes.
- En s'accordant un rang d'écart, la régression logistique multinomiale est assez satisfaisante.
- Nous avons pu observer que l'impact des variables sur le rang varie fortement avec le rang d'un joueur.

Courbe ROC





Im

