

Fiche bilan SAE

Mohamed-Amir BELARBI

Nom de la SAE	Relation âge-performance		semestre / Période	S3-S4
volume horaire consacré par l'étudiant	avec enseignant	1h30	en autonomie	3h
coéquipiers :	Franck Tankapanya		Samuel Darmalingon	
	Enzo Leriche			

Sujet spécifique	Analyse de la relation âge-performance
Objectifs	Démontrer si oui ou non l'âge a un impact sur la performance
Livrables	Rendu d'un rapport

Bilan de la SAE

(reproduire le tableau autant de fois que de compétences mobilisées dans la SAE)

Compétence	Analyser Statistiquement des données
Apprentissages critiques sollicités	Comprendre l'intérêt des analyses multivariées pour synthétiser et résumer l'information portée par plusieurs variables
	Prendre conscience de la différence entre modélisation statistique et analyse exploratoire
	Apprécier les limites de validité et les conditions d'application d'une analyse
Composantes essentielles à respecter	En tenant compte du contexte de l'étude (économique, socio-démographique, commerciale, clinique...)
	En mettant en évidence les grandes tendances et les informations principales
	En identifiant et en mettant en œuvre les techniques adaptées aux données complexes (données massives, données mal structurées, flux de données...)

Compétence	Valoriser une production dans un contexte professionnel
------------	---

Fiche bilan SAE

Apprentissages critiques sollicités	Saisir la nécessité de choisir des indicateurs pertinents pour communiquer sur les résultats
	Prendre conscience de la rigueur requise dans ses productions et dans la communication à leur propos
	Savoir défendre ses choix d'analyses
Composantes essentielles à respecter	En interprétant et contextualisant les résultats (citations, vérification des sources, esprit critique)
	En utilisant la forme de restitution adaptée
	En s'exprimant correctement, aussi bien en français que dans une langue étrangère, à l'oral comme à l'écrit

Compétence	Modéliser les données dans un cadre statistiques
Apprentissages critiques sollicités	Appréhender les difficultés et les limites rencontrées dans la mise en œuvre d'un terrain de collecte
	Comprendre l'impact du type de données sur le choix de la modélisation à mettre en œuvre
	Comprendre l'intérêt de planifier le recueil des données
Composantes essentielles à respecter	En s'adaptant aux spécificités (données, enjeux, méthodes) d'un domaine d'application particulier (santé, marketing, assurance, qualité, socio-démographie...)
	En s'adaptant à la complexité des données (données massives, données mal structurées, flux de données...)
	En choisissant le modèle adapté à la situation

(reproduire le tableau autant de fois que de compétences mobilisées dans la SAÉ)

Ma démarche

Savoirs / connaissances	Savoir-faire	Savoir-être
Adaptation aux types des données Sélection de données	Analyser Interpréter Démontrer Conclure	Travail de groupe Communication

Evaluation du résultat

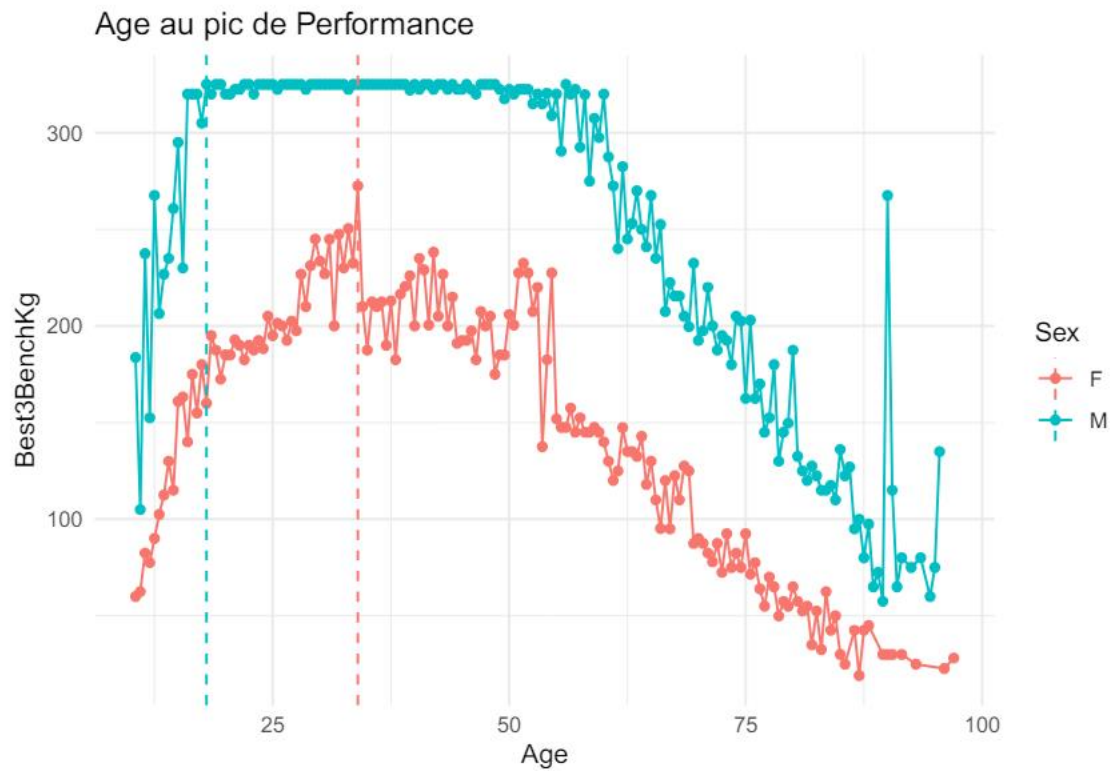
- Ce que je trouve bien réalisé, pourquoi ?
Je trouve que notre répartition du travail a été très bien géré ce qui nous a facilité la tâche, l'organisation du rapport a été très vite puisque chacun s'était immédiatement ce qu'il avait à faire et où le faire. On s'est très bien approprié le contexte des données car on avait déjà des connaissances dans ce domaine ce qui nous a facilité les tâches.
- Ce que je n'ai pas bien compris ; ce qui serait à améliorer pour une prochaine fois : pourquoi ? comment ?

Lors de ce projet, l'objectif était d'étudier la relation entre l'âge et la performance dans le domaine du powerlifting, mais aussi dans notre cas, l'influence du poids et de l'équipement sur la performance. Pour ce faire, nous avons accès à des données de la base Open Powerlifting. La base de données étant très vaste, j'ai dû la nettoyer en ne gardant uniquement les variables qui nous intéressaient comme l'âge, le poids des athlètes, leur meilleure performance en squat et leur meilleure performance en développé couché. J'ai ensuite réalisé des graphiques démontrant le lien entre l'âge et la performance selon le sexe, un graphique pour chacune des trois épreuves. Ensuite, nous avons utilisé le modèle de l'ANOVA pour déterminer si le poids avait une influence sur la performance des athlètes et nous avons déduit que oui. De même pour l'utilisation d'équipement, nous avons utilisé le modèle du Khi2 pour voir s'il y avait une association significative entre la performance et l'utilisation d'équipement et la p-valeur obtenue étant extrêmement faible, cela indique une probabilité extrêmement faible que l'association observée soit due au hasard.

Eléments de preuve, ce que je peux montrer

(Choisir des éléments précis à mettre annexe)

1)



2)

Fiche bilan SAE

Influence du poids sur la performance

Notre analyse se porte désormais sur le rôle que joue le poids des athlètes dans leurs performances en powerlifting. Pour approfondir cette question, nous avons ciblé le sous-échantillon des compétiteurs affiliés à la FFForce, qui constitue notre variable de féderation, pour évaluer l'impact de la masse corporelle sur leur capacité à soulever des poids.

```
data_ffforce <- data %>%
  filter(Federation == "FFForce")

mod2 <- lm(TotalKg ~ WeightClassKg, data = data_ffforce)
mod1 <- lm(TotalKg ~ 1, data=data_ffforce)
anova(mod1,mod2)
```

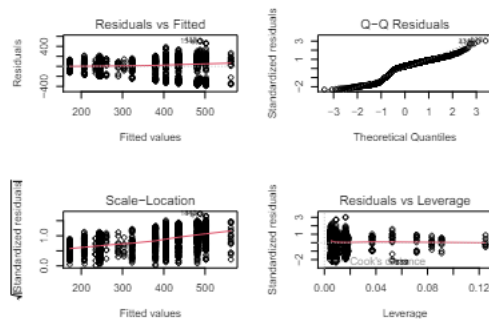
19

```
## Analysis of Variance Table
##
## Model 1: TotalKg ~ 1
## Model 2: TotalKg ~ WeightClassKg
## Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1    1412 58343099
## 2    1394 40429948 18 17913151 34.313 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nous avons effectué une ANOVA pour évaluer l'influence du poids sur la performance des athlètes. Le modèle linéaire utilisé révèle une association significative entre le poids des compétiteurs et leurs performances. L'analyse comparée à un modèle de base, qui ne considère que la performance isolée, montre une p-valeur extrêmement faible, quasiment nulle. Cette preuve de significativité statistique indique clairement que le poids a un impact considérable sur la performance en powerlifting.

Nous avons également procédé à une vérification de la normalité des résidus.
par(mfrow=c(2,2))
plot(mod2)

## Warning: les observations ayant un 'leverage' = 0 ne sont pas représentées sur le graphique :
## 776
```



Equipements

Le powerlifting permet l'utilisation de divers équipements pouvant potentiellement influencer la performance des athlètes. Ces équipements varient du « raw », qui signifie sans équipement, à des aides plus techniques comme le « single-ply » (combinaison à une couche) et le « multi-ply » (combinaison à plusieurs couches), ainsi que les « straps » (bandes élastiques pour fixer les poignets à la barre) et les « wraps » (bandes élastiques autour des articulations).

Pour étudier l'effet de ces équipements, nous avons d'abord transformé la variable quantitative de 'performance totale' (la somme des meilleures performances dans les trois mouvements) en une variable qualitative ordonnée à l'aide de la fonction `cut` du package `questionr`. Ensuite, nous avons appliqué un test du χ^2 en utilisant cette nouvelle variable de performance catégorisée à la variable 'Equipment' pour déterminer si l'utilisation de ces équipements a une influence significative sur la performance en powerlifting.

```
data$performance_cat <- cut(data$TotalKg, breaks = quantile(data$TotalKg, probs = 0:5/5, na.rm = TRUE),
                             labels = c("Q1", "Q2", "Q3", "Q4", "Q5"), ordered = TRUE)

table_equipment_performance <- table(data$Equipment, data$performance_cat)

chi2_results <- chisq.test(table_equipment_performance)

## Warning in chisq.test(table_equipment_performance): L'approximation du Chi-2
## est peut-être incorrecte
print(chi2_results)

##
## Pearson's Chi-squared test
##
## data:  table_equipment_performance
## X-squared = 114827, df = 16, p-value < 2.2e-16
```

Le test du χ^2 de Pearson est un outil statistique conçu pour évaluer l'existence d'une association significative entre deux variables catégorielles. Pour notre analyse, ce test a été appliqué pour explorer la relation entre le type d'équipement (variable `Equipment`) et les catégories de performance (variable `performance_cat`), qui ont été créées à partir des totaux soulevés (variable `TotalKg`) en powerlifting. La p-value obtenue, inférieure à $2.2e-16$, indique une probabilité extrêmement faible que l'association observée soit due au hasard. Par conséquent, il est clairement établi que le type d'équipement influence significativement la performance en powerlifting, confirmant l'existence d'une relation statistiquement significative entre les équipements utilisés et les performances des athlètes.

Nous approfondissons maintenant notre analyse pour déterminer quel type d'équipement favorise le plus la performance en powerlifting. Pour répondre à cette question, nous avons extrait et analysé les résidus standardisés du test du χ^2 réalisé précédemment. L'examen de ces résidus nous permet d'identifier l'équipement qui, au-delà des attentes statistiques, améliore le plus significativement les performances des athlètes.

```
chi2_results <- chisq.test(table_equipment_performance)
```