

Régression linéaire multiple

16 octobre 2020

Ce laboratoire doit être remis le **28 octobre à 17h sur Moodle**. Dans votre réponse pour chaque question, veuillez inclure le code R utilisé (s'il y a lieu) et les résultats obtenus.

1. Prises de charbonnier en Alaska

Le fichier `sablefish.csv` contient des données de Kimura (1988) sur le nombre de prises de charbonnier par unité d'effort (*catch*) dans quatre localités d'Alaska (*location*) pour chacune des six années entre 1978 et 1983.

```
sable <- read.csv("sablefish.csv")
str(sable)
```

```
## 'data.frame':    24 obs. of  3 variables:
## $ year      : int  1978 1978 1978 1978 1979 1979 1979 1979 1980 1980 ...
## $ location: chr  "Shumagin" "Chirikof" "Kodiak" "Yakutat" ...
## $ catch     : num  0.236 0.204 0.241 0.232 0.14 0.202 0.228 0.268 0.286 0.275 ...
```

- a) Transformez l'année et la localité en facteurs dans R, puis effectuez une ANOVA à deux facteurs pour déterminer si l'abondance varie significativement entre les années et entre les localités ($\alpha = 0.05$). À partir des graphiques de diagnostic, vérifiez que les suppositions du modèle d'ANOVA soient respectées.

Note: Nous supposons ici que les effets sont additifs. Aussi, puisqu'il y a une seule mesure pour chaque combinaison d'une année et d'une localité, il n'est pas possible d'estimer une interaction.

- b) Ré-analysez le modèle en (a) avec la fonction de régression linéaire `lm`. Utilisez les contrastes appropriés pour déterminer la moyenne générale des prises (*catch*) et les déviations de cette moyenne pour chacune des six années.
- c) Selon les résultats du modèle en (b), dans quelle localité les prises moyennes sont-elles les plus élevées?
- d) À l'aide du package *emmeans*, illustrez les prises moyennes estimées pour chacune des années avec leur intervalle de confiance. En utilisant un test de comparaisons multiples, indiquez entre quelles années les prises varient de façon significative.

2. Métabolisme d'un poisson selon la salinité

Le tableau de données `sardinella.csv` provient d'une étude de Wohlschlag (1957), "Differences in metabolic rates of migratory and resident freshwater forms of an Arctic Whitefish". Il contient des mesures du poids (*log_weight*) et de la consommation en oxygène (*log_O2*) pour des individus du corégone *Coregonus sardinella* capturés dans un environnement d'eau douce (*freshwater*) ou d'eau salée (*marine*).

```
sardinella <- read.csv("sardinella.csv")
str(sardinella)
```

```
## 'data.frame':    22 obs. of  3 variables:
## $ environment: chr  "marine" "marine" "marine" "marine" ...
## $ log_O2      : num  1.59 1.4 1.47 1.66 1.55 ...
## $ log_weight  : num  2.5 2.04 2.15 2.35 2.24 ...
```

- a) Estimez les effets additifs de l'environnement et du poids sur la consommation en oxygène de ce poisson. Comment interprétez-vous chacun des paramètres du modèle?
- b) Reprenez le modèle en (a) avec une version normalisée du prédicteur *log_weight* (*norm_weight*). Quelle est maintenant l'interprétation des coefficients?
- c) Reprenez le modèle en (b) en ajoutant l'interaction entre le poids (normalisé) et l'environnement. Quelle est l'interprétation des coefficients? Est-ce que l'interaction est significative?