

# Régression linéaire

16 octobre 2019

## 1. 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 corgéone *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: Factor w/ 2 levels "freshwater","marine": 2 2 2 2 2 2 2 2 1 1 ...
## $ 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 ...
```

- 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?
- Reprenez le modèle en (a) avec une version normalisée du prédicteur *log\_weight* (*norm\_weight*). Quelle est maintenant l'interprétation des coefficients?
- Reprenez le modèle en (b) en ajoutant l'interaction entre le poids (normalisé) et l'environnement. Cette interaction est-elle significative? Quelle est l'interprétation des coefficients?

## 2. Diversité des plantes sur des îles britanniques

Le tableau de données `britain_species.csv` provient de l'étude de Johnson et Simberloff (1974), "Environmental determinants of island species numbers in the British Isles". Ces données indiquent le nombre d'espèces de plantes vasculaires (*species*) pour 42 îles britanniques en fonction de différents prédicteurs: surface en km<sup>2</sup> (*area*), altitude en m (*elevation*), nombre de types de sol (*soil\_types*), latitude et distance de la Grande-Bretagne en km (*dist\_britain*).

```
iles <- read.csv("britain_species.csv")
str(iles)
```

```
## 'data.frame': 42 obs. of 7 variables:
## $ island : Factor w/ 42 levels "Ailsa","Anglesey",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ area : num 0.8 712.5 429.4 18.4 31.1 ...
## $ elevation : int 340 127 874 384 226 1343 210 103 143 393 ...
## $ soil_types : int 1 3 4 2 1 16 1 3 1 1 ...
## $ latitude : num 55.3 53.3 55.6 57 60.1 54.3 57.1 56.6 56.1 56.9 ...
## $ dist_britain: num 14 0.2 5.2 77.4 201.6 ...
## $ species : int 75 855 577 409 177 1666 300 443 482 453 ...
```

- Supposons qu'une théorie prédit que le nombre d'espèces (*S*) dépend de la surface d'une île (*A*) en fonction de l'équation suivante, où *c* et *z* sont des paramètres à déterminer:

$$S = cA^z$$

Utilisez un modèle linéaire pour tester l'hypothèse que le nombre d'espèces de plantes vasculaires suit cette équation avec une exposant  $z = 0.25$  (un quart).

*Indice:* Supposons que la valeur estimée de  $z$  suit une distribution normale. À partir de la valeur estimée de  $z$  et de son erreur-type, utilisez la formule vue au deuxième cours pour calculer l'intervalle de confiance.

$$(\hat{z} + t_{df, \alpha/2} SE, \hat{z} + t_{df, 1-\alpha/2} SE)$$

Dans cette formule,  $SE$  est l'erreur-type,  $\alpha$  est le seuil de significativité que vous choisissez (ex.: 0.05) et  $df$  est le nombre de degrés de liberté de la distribution  $t$ , que vous pouvez déterminer à partir du sommaire de la régression.

- b) Estimez maintenant le modèle suivant, où le nombre d'espèce dépend à la fois de la surface de l'île et de sa distance de la Grande-Bretagne, sur une échelle logarithmique. Vous devrez d'abord exclure l'île de Grande-Bretagne (**Britain**) du jeu de données.

$$\log(\text{species}) \sim \log(\text{area}) + \log(\text{dist\_britain})$$

- c) À partir du modèle en (b), donnez un intervalle de prédiction à 90% du nombre d'espèces pour (i) une île de 1 km<sup>2</sup> située à une distance de 5 km et (ii) une île de 40 km<sup>2</sup> située à une distance de 20 km.

*Notes:*

- Modifiez le % de l'intervalle de prédiction avec l'argument **level** de **predict**.
- Puisque la réponse du modèle est **log(species)**, le résultat de **predict** sera sur une échelle logarithmique.