

Régression de Poisson

10 novembre 2021

1. Communautés d'insectes de cours d'eau

Le tableau de données `stream_composition.csv` indique le nombre d'espèces de cinq ordres d'insectes présentes dans 20 cours d'eau, en fonction de la température et du pH.

```
stream <- read.csv("stream_composition.csv")
str(stream)
```

```
## 'data.frame': 20 obs. of 8 variables:
## $ stream : int 1 2 3 4 5 6 7 8 9 10 ...
## $ pH : num 6.8 5.5 6.3 7.3 7.2 7 7 6.1 6.2 7.5 ...
## $ temperature: num 17.4 17.1 17 16.8 18.9 18.1 16.3 15 15.8 16.8 ...
## $ mayfly : int 26 17 7 17 27 28 19 6 9 19 ...
## $ stonefly : int 4 1 2 6 3 6 4 4 5 3 ...
## $ caddisfly : int 9 23 25 9 16 19 21 21 37 12 ...
## $ diptera : int 30 16 10 25 25 30 19 30 26 12 ...
## $ beetle : int 3 17 1 1 2 21 13 12 5 3 ...
```

- Estimez l'effet de la température et du pH sur le nombre d'espèces de plécoptères (`stonefly`), avec une régression de Poisson suivant la formule `stonefly ~ temperature + pH`. Vérifiez si les données sont surdispersées et corrigez vos estimés s'il y a lieu.
- Quelle portion de la variance du nombre d'espèces est expliquée par le modèle?
- Si l'une des deux variables a un effet significatif, interprétez la valeur du coefficient.
- Affichez le nombre d'espèces observées et les courbes de prédiction du modèle pour des valeurs du pH allant de 5.5 à 7.5 et pour trois valeurs de la température: 15, 17 et 19 degrés C.

Truc: Avec `ggplot`, pour que les courbes de prédiction relient les points ayant la même valeur d'une variable numérique (ex.: température), vous devez définir un groupe (ex.: `group = temperature`) dans la fonction `aes`. Vous pouvez aussi convertir la température en facteur.

- Répétez les étapes (a)-(d) pour un modèle du nombre d'espèces d'éphémères (`mayfly`).
- Quel est le nombre moyen d'espèces d'éphémères prédit par le modèle en (e) pour un cours d'eau avec une température de 17 degrés et un pH de 8.5? Cette prédiction est-elle fiable?

2. Salamandres dans différents paysages forestiers



Photo: Bill Bouton

Le fichier `salamander.csv` contient des données de Welsh et Lind (1995) sur le nombre de salamandres (`salaman`) de l'espèce *Plethodon elongatus* dans 47 placettes (`site`), en fonction du pourcentage de couvert forestier (`pct_cover`) et de l'âge du peuplement (`forest_age`).

```
sal <- read.csv("salamander.csv")
str(sal)
```

```
## 'data.frame': 47 obs. of 4 variables:
## $ site : int 1 2 3 4 5 6 7 8 9 10 ...
## $ salaman : int 13 11 11 9 8 7 6 6 5 5 ...
## $ pct_cover : int 85 86 90 88 89 83 83 91 88 90 ...
## $ forest_age: int 316 88 548 64 43 368 200 71 42 551 ...
```

- À partir d'une régression de Poisson, estimez l'effet du couvert forestier (`pct_cover`) sur le nombre de salamandres par placette.
- Est-ce que l'ajout du prédicteur `forest_age` améliore le pouvoir prédictif de ce modèle?
- Produisez un graphique du nombre de salamandres en fonction de l'âge du peuplement et superposez des points représentant les valeurs attendues (`fitted`) du modèle basé uniquement sur le couvert forestier. Qu'observez-vous?
- D'après ces résultats, croyez-vous que l'âge du peuplement influence directement la population de salamandres? Est-ce qu'il influence indirectement cette population?