## Étape 7. Choix de l'analyse

**Pourquoi notre problème requiert des analyses multivariées et non univariées?**

Les analyses multivariées portent sur l'analyse de données complexes en termes de dimensionalité, donc plusieurs variables avec plusieurs dimensions et de manière simultanée. Dans notre exemple :

* conditions physico-chimiques, présence-absence d’espèces, topologie des étangs sont de dimensions différentes et mesurées dans chacun de nos étangs
* une analyse simultanée de plus de 2 variables est une analyse multivariée.
* On veut analyser la réponse de plusieurs espèces à différents facteurs 🡪 analyse multiV

**Quelle analyse multivariée est appropriée et pourquoi?**

Dans notre cas : RDA partielle

Nous avons décidé d'utiliser une analyse de redondance (RDA) partielle pour analyser comment la présence de clôture, l'utlisation des étangs par les vaches et les variables physico-chimiques en plus des facteurs biotiques font varier la communauté d'amphibiens, en prennant en compte l'effet de l'aire et de la profondeur des étangs comme co-variables.

Pourquoi ?

Approprié pour mettre en lien des variables liées aux espèces (matrice Y) et des variables environnementales (matrice X) comme dans notre cas.

C’est un test non paramétrique donc pas besoin de supposer une distribution normale multivariée 🡪 la majorité de nos variables dans la RDA partielle ne sont pas distribués de manière normale. Les données d'abondances sont loins d'être \*Gaussienne\*. Les RDAs sont des tests non-paramétriques, donc elles conviennent très bien aux données d'abondance que nous avons.

RDA partielle parce qu’on a des covariables (profondeur et aire) qui ne nous intéressent pas mais qui vont influencer les relations entre les autres variables.

Combine une régression linéaire multiple à une ordination classique (PCA).

Pourquoi pas une MANOVA?

Test paramétrique donc requiert une distribution normale multivariée dans les abondances d’espèces, ce qui est rarement le cas en pratique. Elle utilise les distances euclidiennes qui ne sont pas appropriées pour les abondances d’espèces et aussi il y a souvent plus d’espèces que de degrés de libertés dans le système.

Pourquoi pas une PCA/analyse non contrainte?

La RDA est un test statistique basé sur des permutations, ce qui nous permet d'étudier le lien entre une matrice de variables réponse et une matrice de variables explicative. Les analyses non-contraintes (p. ex. une PCA) sont seulement descriptives: on ne teste pas de liens de cause à effet. L’analyse non-contrainbte est purement descriptive.

## Étape 7. Choix de transformation de la matrice d'abondance:

Lorsque plusieurs sites partagent la présence d'une espèce, nous savons que ces sites se ressemblent entre eux, (ils partagent les conditions nécessaires à la présence de cette espèce). Par contre, lorsque que plusieurs sites partagent l'absence d'une espèce, nous ne pouvons pas savoir si ces zéros sont dus aux même facteurs (absence réelle ou simplement pas été échantillonnée. Par conséquent, nous ne pouvons pas conclure qu'ils sont aussi ressemblants que s'ils partageaient la présence d'une espèce.

Il est donc nécessaire de transformer les données à l'aide d'une mesure de \*\*dissimilarité asymétrique\*\* (pour donner davantage de poids à la présence qu'à l'absence de l'espèce).

En standardisant les données par abondance totale d'espèce, les investigateurs ont donné davantage de poids aux espèces rares, ce qui est questionnable lorsqu'on veut établir les similarités dans une communauté. En effet, les espèces dominantes et bien-représentées pourraient être plus informatives quant aux ressemblances entre sites que les espèces rares qui tendent à être moins bien représentées dans les échantillons (Legendre et Legendre, 2012).

Nous avons 2 options :

1- 1- Utiliser les données d'abondance standardisées et les transformer

Les transformations les plus recommandées pour un usage général sont les transformations de "Chord" et "Hellinger" (Legendre et Legendre, 2012). Nous choisissons la transformation de Hellinger, recommandée par de nombreux auteurs pour l'ordination des données d'abondance d'espèces (Legendre et Legendre, 2012), et qui obtient le meilleur R^2 pour représenter la distance géographique réelle (Legendre et Gallagher, 2001). La transformation de Hellinger peut être utilisé sur les données standardisées par abondance totale d'espèce (Legendre et Legendre, 2012), ce qui a pour effet de donner une mesure de distance réduite entre les sites qui partagent une espèce rare par rapport aux sites partageant une espèce commune (Noy-Meir et al., 1975). Dans une communauté mature, il peut être judicieux de donner davantage de poids aux espèces rares (Legendre et Legendre, 2012).

2- Utiliser les données de présence-absence (binaire) et les transformer

En transformant les données en présence-absence, l’analyse de la communauté amphibienne se ferait en fonction de la co-occurrence des espèces, plutôt que de la co-fluctuations de leurs abondances (Legendre et Lengendre, 2012). Il est recommandé d'utiliser les données de présence-absence lorsque c'est la seule chose disponible ou lorsque nos variables quantitatives sont de qualité inégale ou incertaine (Borcard et al. 2011). Une seconde transformation à l'aide d'un coefficient binaire assymétrique est nécessaire afin de régler le problème des doubles zéros et donner plus de poids à la présence qu’à l’absence des espèces.

Nous choisissons la première option, qui permet de conserver davantage d’information sur la co-fluctuations des abondances et qui respecte l’idée des investigateurs de donner davantage de poids aux espèces rares dans l’analyse de la communauté amphibienne de ces étangs.