

# MoDoBio Modélisation des Données en Biologie

## L3 Sciences de la Vie, parcours Biologie-Ecologie

Elodie Brunel MA et Céline Devaux BE

Université de Montpellier – Faculté des Sciences

2024-2025



# Objectifs

- ▶ construction de protocoles
- ▶ description numérique et graphique des données
- ▶ construction de modèles appropriés
- ▶ codage des modèles sous R
- ▶ interprétation de résultats

# Organisation

- ▶ Introduction - Rappels (cours 1)
  - pourquoi et comment faire des stats
  - deux exemples de projets de M1
  - mise en garde
  - retour sur les protocoles
- ▶ Partie I. Modèles Linéaires (cours 2, 3, 4 )
- ▶ Partie II. Analyse de la variance à 1 facteur (cours 5 )
- ▶ Partie III. Analyse de la covariance (cours 6, 7)
- ▶ Partie IV. Vers plus de complexité biologique (cours 8)

# Organisation

- ▶ 8 CM de 1h30, tous ensemble
  - ▶ E. Brunel
  - ▶ C. Devaux
- ▶ 4 TD de 3h, en groupes de 40
  - ▶ E. Brunel
  - ▶ C. Devaux/L. Etienne
- ▶ 3 TP sous R, en groupes de 20
  - ▶ Lucas Buffan
  - ▶ Timothée Chenin
  - ▶ Lucas Etienne
  - ▶ Léo Streith
  - ▶ Julien Offesson

# Modalités de contrôle des connaissances

- ▶ Quizz en présentiel 1H, le **mercredi 13 novembre 2024 à 18h30** = 20%
- ▶ Devoir Maison (DM) donné le 22 novembre, et à rendre pour le **lundi 2 décembre 2024 avant 12H** = 15%
- ▶ Devoir sur table (CT), en janvier = 65%  
**Attention la règle du MAX disparaît.** Le calcul de la note de session 1 est :

$$0.20 \times \text{QUIZZ} + 0.15 \times \text{DM} + 0.65 \times \text{CT}$$

- ▶ Si absence justifiée au QUIZZ dans les 8 jours, la note du QUIZZ sera "neutralisée".  
La remise du DM est obligatoire.

# Déroulement

- ▶ 4 ECTS = 80h de travail
  - ▶ en cours
  - ▶ en dehors du cours en groupe
  - ▶ en dehors du cours seul.e
- ▶ Aide
  - ▶ docs de cours
  - ▶ littérature conseillée
  - ▶ quizz de révision
  - ▶ annales des sessions précédentes

# Objectifs

- ▶ des questions biologiques réelles
- ▶ des données biologiques réelles
- ▶ des outils mathématiques utiles aux biologistes
- ▶ des logiciels utiles aux biologistes
- ▶ des méthodes (très) utiles pour le projet tuteuré (8 ECTS !)

# Pourquoi faire des stats ?



# Pourquoi faire des stats ?

- ▶ pour répondre à des questions biologiques
- ▶ parce qu'on ne peut pas mesurer toute une population
- ▶ parce qu'on doit considérer la variance d'échantillonnage
- ▶ parce qu'on veut quantifier le risque de se tromper

# Terminologie

- population biologique vs population statistique
- une variable aléatoire mesurée sur des individus \*
- un échantillon aléatoire, composé d'individus (statistiques) indépendants
- une hypothèse nulle  $H_0$  à tester \*
- un risque  $\alpha$  de se tromper en rejetant  $H_0$  \*
- un risque  $\beta$  de se tromper en conservant  $H_0$  \*

# Pourquoi construire des modèles ?

- ▶ pour répondre à des questions biologiques
- ▶ pour décrire (e.g. moyenne, variance, quantile, intervalle de confiance) \*
- ▶ pour tester une hypothèse biologique (c'est souvent  $H_1$  qu'on espère)
- ▶ pour prédire

# Exemple 1

Comment expliquer la latéralité chez les fourmis attines ?

- population biologique vs population statistique
- une variable aléatoire mesurée sur des individus
- un échantillon aléatoire, composé d'individus (statistiques) indépendants
- une hypothèse nulle  $H_0$  à tester
- un risque  $\alpha$  de se tromper en rejetant  $H_0$
- un risque  $\beta$  de se tromper en conservant  $H_0$

# Exemple 1

Comment expliquer la latéralité chez les fourmis attines ?

## **Hypothèses faites**

- ▶ leur choix est indépendant des autres fourmis
- ▶ efficacité de la découpe de feuilles
- ▶ usure de la mandibule

# Exemple 1

Quel est le lien entre la latéralité des fourmis et l'usure de leur mandibule ?

- ▶ observation = mesurer la variabilité des variables explicatives
- ▶ expérience = contrôler la variabilité des variables explicatives

# Exemple 1

Données : des variables mesurées sur plusieurs fourmis d'une même colonie

Etudier :

1. si les variables sont liées
2. quelle est la force du lien
3. si la variable d'intérêt peut-être prédite seulement en mesurant les autres

# Exemple 1

Données : des variables mesurées sur plusieurs fourmis d'une même colonie

- ▶ le nombre de découpes réalisées avec la mandibule droite
- ▶ le nombre de découpes réalisées avec la mandibule gauche
- ▶ l'usure de la mandibule droite
- ▶ l'usure de la mandibule gauche

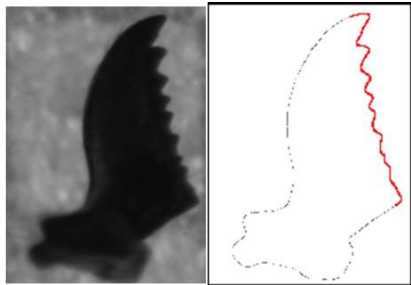


# Exemple 1

$n = 20$  fourmis, pour plein de découpes

$p_i$  proportion des découpes faites "à droite" pour la fourmi  $i$

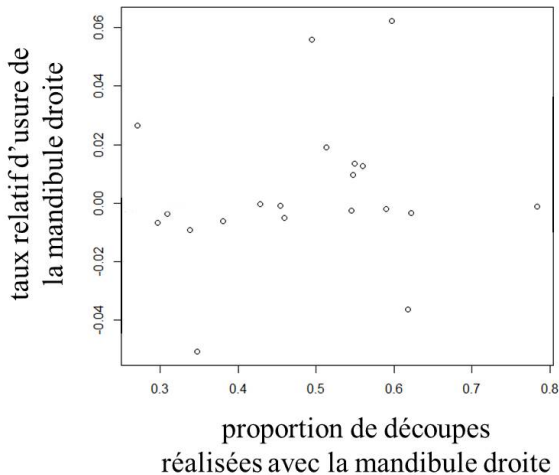
$t_i$  taux d'usure la mandibule droite relatif à celui de la gauche



stage de M1 de Yohan Vien, BIOSTATS - UM

# Exemple 1

Représentation synthétique des résultats



## Exemple 1

Comment tester que plus une mandibule est utilisée plus elle est usée ?

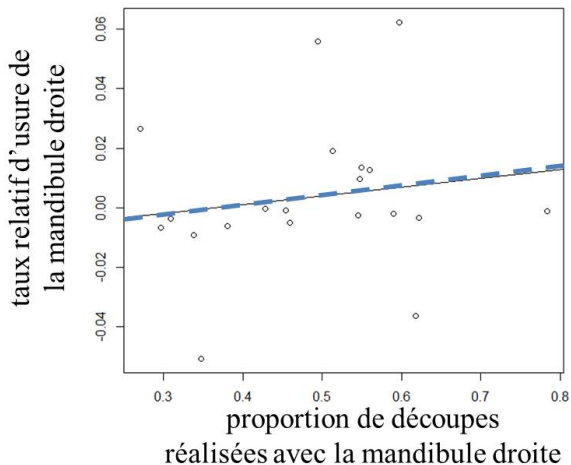
- ▶ en calculant un coefficient de corrélation linéaire  $\rho$  (noté aussi  $r$ ) entre l'usure relative  $t_i$  et la proportion de découpes  $p_i$  :

$$\rho = \frac{\sum_{i=1}^n (p_i - \bar{p})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2 \times \sum_{i=1}^n (t_i - \bar{t})^2}}$$

- ▶ en régressant l'usure relative sur la proportion de découpes avec la mandibules droite

## Exemple 1

Régression linéaire de l'usure relative sur l'usage relatif



## Exemple 1

Régression linéaire de l'usure relative sur l'usage relatif

1. si les variables sont liées ( $\hat{r} = 0.14$ )
2. quelle est la force du lien ( $\hat{\beta}_1 = 0.03$  ou  $R^2 = 0.02$ )
3. si la variable d'intérêt peut-être prédite seulement en mesurant les autres variables

mais

- des fourmis très différentes des autres
- une pente faible

## Exemple 2

Est-ce que les pollinisateurs sont bons pour "faire des graines" ?

- ▶ conditionnellement à une visite
- ▶ conditionnellement à l'efficacité du transfert de pollen
- ▶ qualité du pollen déposé

## Exemple 2

Quel est le lien entre le mode de pollinisation (vent vs insectes) et le succès reproducteur ?

- ▶ observation = mesurer la variabilité des variables explicatives
- ▶ expérience = contrôler la variabilité des variables explicatives

## Exemple 2

Données : des variables mesurées sur plusieurs plantes

Etudier :

1. si les variables sont liées
2. quelle est la force du lien
3. si la variable d'intérêt peut être prédite seulement en mesurant les autres



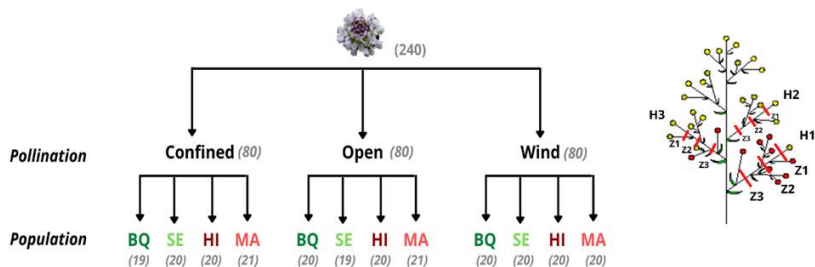
## Exemple 2

Données : des variables mesurées sur plusieurs plantes

- ▶ la moyenne du nombre de graines produites dans un fruit
- ▶ le mode de pollinisation (contrôlé)
- ▶ la population d'origine

## Exemple 2

Protocole expérimental équilibré pour les variables contrôlées

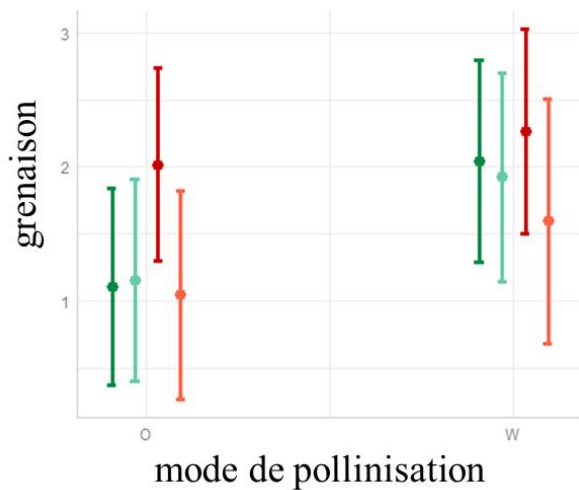


stage de M1 de Bushra Shahid, MEME - UM

► pourquoi le même nb de plantes dans chaque groupe ?

## Exemple 2

Représentation synthétique des résultats



## Exemple 2

Comment tester que

- ▶ les moyennes sont égales entre mode de pollinisation ?
- ▶ les moyennes sont égales entre populations ?
- ▶ le mode de pollinisation affecte identiquement les populations ?

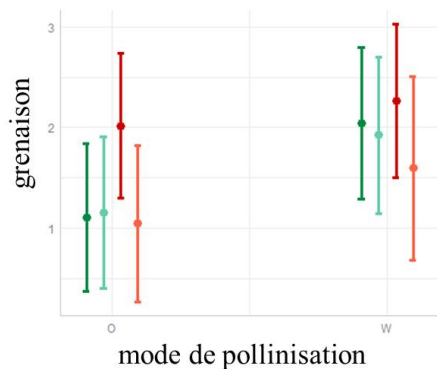
## Exemple 2

En faisant une ANOVA (ANalysis Of VAriance) qui teste les effets

- ▶ mode de pollinisation (p-value= 0.00026)
- ▶ populations (p-value= 0.0028)
- ▶ l'interaction pollinisation/population (p-value= 0.34)

## Exemple 2

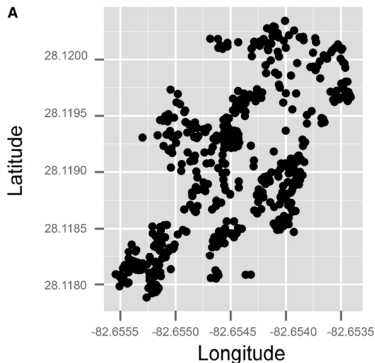
### Interprétation des résultats



la population "rouge" différente des autres ?  
le vent fait mieux que les insectes !

# Mise en garde

- ▶ question : niche écologique
- ▶ données
  - présence / absence de *Lilium catsebai* en Floride
  - variables environnementales



- ▶ expliquer la présence par les v. environnementales
- ▶ comparer les v. environnementales entre présence et absence

Sommers et al. 2011 tiré de Ecological Statistics, Eds. Fow, Negrete-Yankelevich et Sosa (2015)

# Mise en garde

- ▶ expliquer la présence par les v. environnementales
  - v. d'intérêt = présence, en fonction de l'environnement
  
- ▶ comparer les v. environnementales, entre présence et absence
  - v. d'intérêt = v. environnementales, en fonction de présence/absence

**présence/absence = traitement expérimental ! FAUX !!!**



# Planification d'expériences

Penser à :

1. où se trouve la variabilité, comment elle est structurée "en nature"
2. faire une figure des résultats attendus
3. bien poser la ou les hypothèses à tester
4. construire le modèle statistique
5. construire le protocole expérimental

# Planification d'expériences - cas d'une variable qualitative

Penser à :

1. des groupes témoins
2. randomiser les individus entre les groupes
3. réduire la variance des erreurs
4. avoir le même nombre d'individus par traitement