

ECOLOGIE STATISTIQUE ET PROGRAMMATION SOUS R



Dr. Ir. Biaou S. S. Honoré
Maître de Conférences
Faculté d'Agronomie – Université de Parakou

PLAN

1. Echantillonnage et collecte des données sur la diversité
2. Tableaux de données en écologie
3. Analyses portant sur la richesse spécifique (TD1)
4. Analyses portant sur la diversité (TD2, TD3 et TD4)
5. Analyses portant sur des données issues de comptage (TD5 et TD6)
6. Analyses des données de présence-absence des espèces (TD7)
7. Analyses portant sur les différences de composition spécifique (TD8, TD9 et TD10)

Objectifs de la formation

- Développer chez les apprenants la capacité à organiser, analyser et interpréter les données de leurs propres recherches en écologie avec le logiciel R
- En insistant sur les questions liées à l'analyse de la Biodiversité: analyses portant sur **richesse spécifique**, la **diversité**, les **données issues de comptage**, les données de **présence-absence** des espèces et les différences de **composition spécifique**.

Méthodologie de la formation

- Exposés Powerpoint / Rappels :
 - ✓ Echantillonnage et collecte des données sur la diversité
 - ✓ Tableaux de données en écologie
- Travaux dirigés pour tous les autres modules (analyses avec le logiciel R) :
 - ✓ Comment préparer un fichier de script
 - ✓ Exécution du script dans R
 - ✓ Interprétation des résultats

Bibliographie

- Ce cours repose sur l'ouvrage de Kindt et al. (2005):
 - Kindt, R., & Coe, R. (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi, 196.
- Il est disponible gratuitement en téléchargement sur:
<https://www.worldagroforestry.org/downloads/Publications/PDFS/b13695.pdf>

1^{ère} Partie

ECHANTILLONNAGE ET
COLLECTE DES DONNEES SUR LA
DIVERSITE

Plan d'échantillonnage

- Le plan d'échantillonnage doit être fondé sur des **objectifs** de recherche et les **hypothèses** que vous voulez tester:
 - Quelles sont les variables dépendantes mesurées (variable(s) réponse) ?
 - Quelles sont les variables explicatives mesurées (variables que vous supposez influencer la réponse) ?

Plan d'échantillonnage

- Exemple d'hypothèse : « la richesse en espèces d'arbres dans les petites exploitations agricoles est influencée par le niveau d'intégration du marché de l'entreprise agricole, parce que l'intégration du marché détermine les arbres plantés et conservés ».
- Dans cet exemple pouvez vous identifier:
 - la variable de réponse ?
 - la/les variable(s) explicative(s) ?
 - l'unité de recherche ?

Plan d'échantillonnage

- Réponses :
 - la **richesse en espèces** est la variable de réponse
 - le **niveau d'intégration** du marché est une variable explicative.
 - l'hypothèse se réfère aux **petites exploitations agricole**, c'est donc l'unité de recherche.
 - la partie « parce que... » de l'hypothèse ajoute beaucoup de valeur à la recherche et son étude nécessite des informations supplémentaires sur le fait de savoir si des espèces ont été plantées ou conservées et pourquoi ?

Plan d'échantillonnage

- Exemple d'hypothèse : « **la diversité des arbres dans les exploitations agricoles d'Afrique de l'Est diminue avec l'éloignement des forêts car plus de graines sont dispersées par les arbres forestiers que par les arbres des espaces agricoles** ».
- Dans cet exemple pouvez vous identifier:
 - la variable de réponse ?
 - la/les variable(s) explicative(s) ?
 - l'unité de recherche ?

Plan d'échantillonnage

Taille des parcelles/unités d'observation

La taille des relevés influencera généralement les résultats :

- Vous trouverez normalement plus d'espèces dans des quadrats de 100 m² que dans des quadrats de 1 m².
- Mais 100 quadrats dispersés de 1 m² contiendront probablement plus d'espèces qu'une parcelle unique de 100 m².

Si le but n'est pas de trouver des espèces mais de comprendre un phénomène écologique, la taille de chaque parcelle doit être adaptée à l'ampleur des processus étudiés.

Plan d'échantillonnage

Forme des parcelles/unités d'observation

La forme du quadrat influencera souvent aussi les résultats:

- **Plus d'espèces d'arbres sont observées dans les quadrats rectangulaires que dans les quadrats carrés** de la même zone. Ce phénomène s'explique par le fait que les espèces d'arbres se présentent souvent en grappes, de sorte que davantage d'arbres de la même espèce seront observés dans les quadrats carrés.
- Lorsque les quadrats sont rectangulaires, leur orientation peut également poser problème (en général perpendiculaire au gradient par soucis d'homogénéité floristique).
- le marquage d'une parcelle circulaire peut être beaucoup plus difficile que le marquage d'une parcelle rectangulaire ou carrée dans les milieux très boisés.

Plan d'échantillonnage

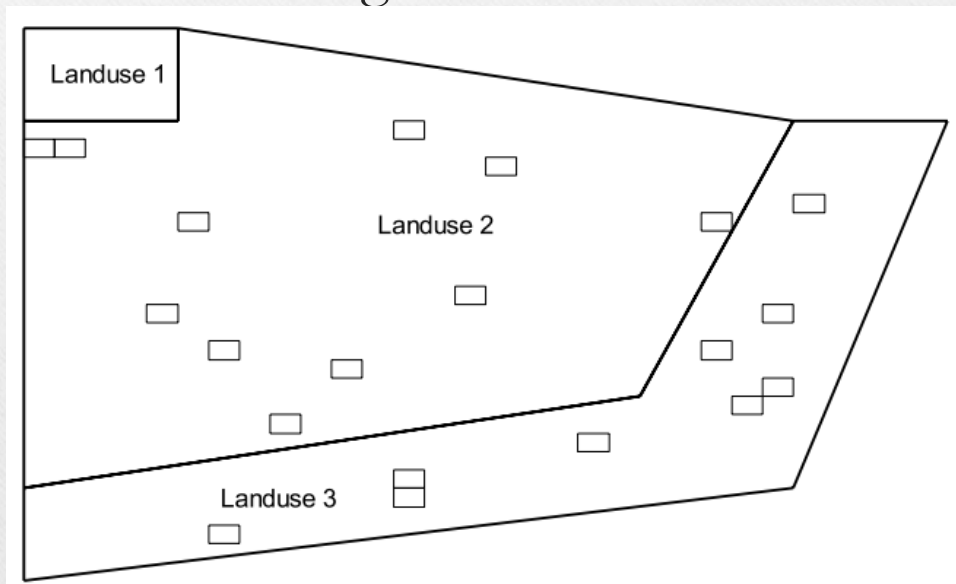
Étant donné que la taille et la forme de l'unité d'échantillonnage peuvent influencer sur les résultats, il est préférable de **ne conserver qu'une taille et une forme pour les quadrats au sein d'une étude.**

Si vous souhaitez comparer les résultats avec d'autres enquêtes, il sera alors plus facile d'utiliser les mêmes tailles et formes de quadrats.

Plan d'échantillonnage

Méthodes d'échantillonnage :

- Échantillonnage aléatoire

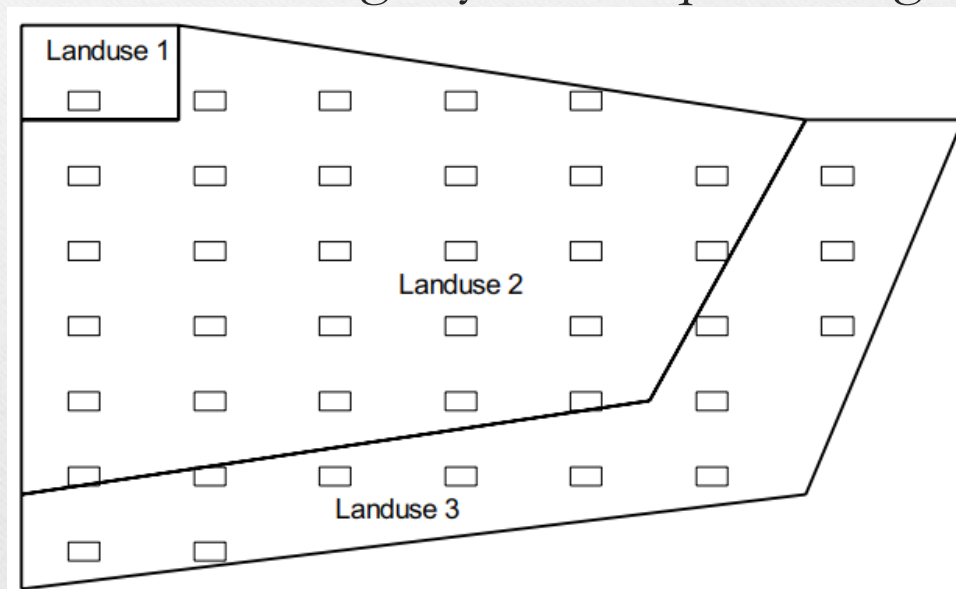


Il y a un risque que les strates faiblement représentées (telle que Land Use 1) ne soient pas échantillonnées

Plan d'échantillonnage

Méthodes d'échantillonnage :

- Échantillonnage systématique ou régulier

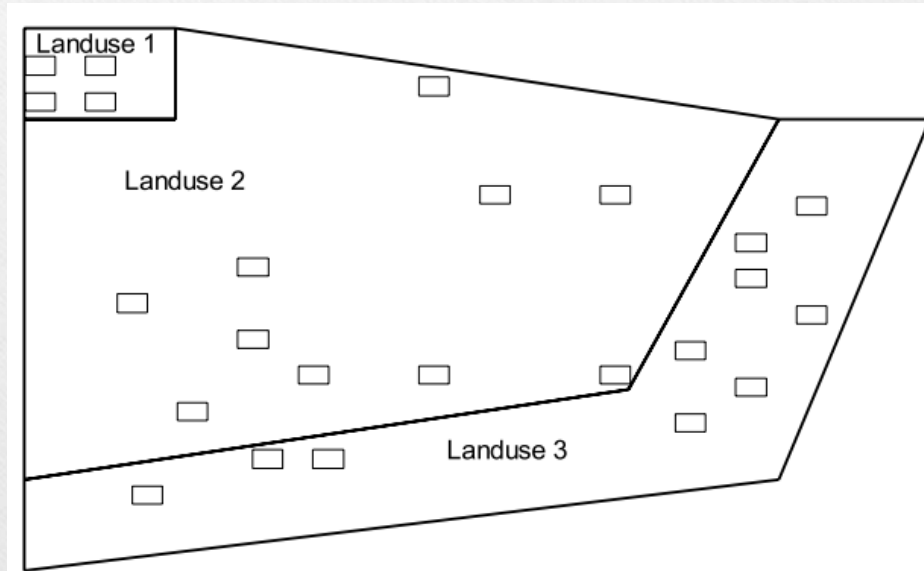


Garanti que les données soient collectées sur l'ensemble de la zone d'étude

Plan d'échantillonnage

Méthodes d'échantillonnage :

- Échantillonnage stratifié



Garanti que des observations soient effectuées dans chaque strate. Les échantillons de parcelles sont choisis au hasard pour chaque mode d'utilisation du sol

Plan d'échantillonnage

Taille de l'échantillon

La taille de l'échantillon nécessaire dépendra de la complexité des objectifs :

- **il faut plus de données pour démêler les relations complexes** entre plusieurs réponses et variables explicatives que pour simplement comparer la moyenne de deux groupes.
- cela **dépendra de la variabilité de la réponse étudiée**
- si chaque unité d'échantillonnage était la même, il suffit d'en mesurer une pour avoir toutes les informations!
- cela **dépendra également de la précision** avec laquelle vous aurez besoin de connaître les réponses.

Plan d'échantillonnage

Taille de l'échantillon

Si l'étude doit comparer différentes strates ou conditions, il est clair que nous avons besoin d'observations dans chaque strate ou représentant chaque ensemble de conditions.

Nous devons ensuite planifier des observations répétées dans une strate ou un ensemble de conditions.

Plan d'échantillonnage

Taille de l'échantillon

Selon Bellefontaine et al (2007):

Un échantillon préliminaire à très faible taux est à préconiser afin de caractériser l'hétérogénéité, c-à-d la valeur des coefficients de variation (CV) des variables étudiées, qui permettront d'estimer le nombre minimum d'unités nécessaires pour atteindre une précision donnée.

Plan d'échantillonnage

Taille de l'échantillon

La taille de l'échantillon pour obtenir une précision p_0 donnée peut être estimée par (Bellefontaine et al 2007):

$$n \geq \frac{4N(CV)^2}{(N-1)p_0^2 + 4(CV)^2}, \text{ et si } N \text{ est grand, } n \geq \frac{4(CV)^2}{p_0^2}$$

n = taille de l'échantillon;

N = nombre de parcelles potentielles sur la surface inventoriée;

CV = coefficient de variation de la variable réponse (volume, surface terrière, ou hauteur,..., attaché à chaque parcelle) obtenu avec l'échantillon préliminaire;

P_0 = précision souhaitée.

Plan d'échantillonnage

Taille de l'échantillon

Le pré échantillonnage est cependant lourd à mettre en œuvre et augmente les coûts.

- On peut aussi **calculer le coefficient de variation au fur et à mesure de l'avancement de l'inventaire** ou se reporter à des valeurs obtenues dans des situations analogues dans la littérature.

Plan d'échantillonnage

Nombre d'observations dans chaque strate

- Un échantillonnage avec le **même nombre d'observations par strate** sera optimal **si l'objectif est de comparer les différentes strates** et si vous ne disposez pas d'informations supplémentaires ou d'hypothèses sur d'autres sources de variation.

Plan d'échantillonnage

Nombre d'observations dans chaque strate

- Si l'objectif est d'estimer la moyenne de l'ensemble de la zone d'étude, une alternative consiste à **rendre le nombre d'observations par strate proportionnel à la taille de la strate.**
- L'un des avantages est que la moyenne de l'ensemble de la zone d'étude sera la moyenne de toutes les parcelles d'échantillonnage.

2^{ème} Partie

TABLEAUX DE DONNEES EN
ECOLOGIE

Tableaux de données

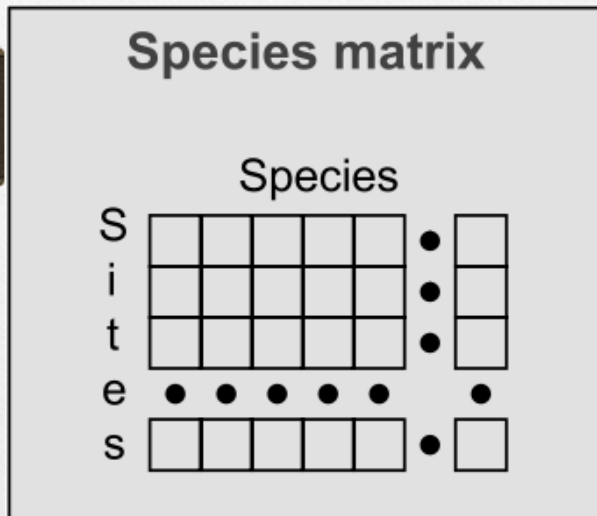
- En général, pour une analyse de la diversité, le chercheur dispose d'1 ou 2 tableaux de données, suivant le but de sa recherche:
 - 1 **tableau de données sur les espèces** (abondance, biomasse, présence-absence, coefficient d'abondance, etc.)
 - 1 **tableau de données sur les variables environnementales** (sol, climat, perturbations, modes de gestion, traitements spécifiques, etc.)

Tableaux de données (suite)

- Prévoir une colonne (généralement la 1^{ère} colonne) servant à **identifier les unités d'observation (sites, placettes, relevés, station, etc.)**
- Chacune des autres colonnes correspond à une **variable mesurée** sur chaque unité d'observation
 - Espèces pour le tableau des espèces
 - Variables environnementales

Tableaux de données (suite)

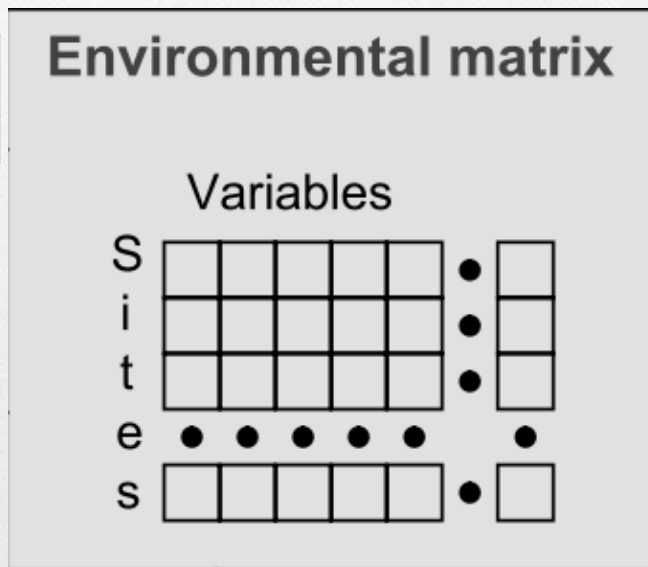
Especies :



Site	Species S1 (count)	Species S2 (count)	Species S3 (count)
A	1	1	1
B	4	0	1
C	2	2	0
D	0	1	2

Tableaux de données (suite)

Variables environnementales:



Site	Soil depth (m)
A	1.0
B	2.0
C	0.5
D	1.5

Lorsqu'on dispose de deux tableaux de données, les lignes des deux matrices doivent avoir strictement les mêmes noms et suivre le même ordre.

RAPPEL: quelques règles de saisie

- L'utilisation des espaces doit être évitée dans les données et/ou noms de variables. Les espaces doivent être remplacés par des « . » ou des « _ » pour éviter les espaces qui compliqueraient la lecture des données par le logiciel
- Les données manquantes, elles doivent toutes être codées **NA** par convention
- Utiliser de préférence une codification alphanumérique (texte) plutôt que numérique pour les variables qualitatives,

PREPARATION DES TRAVAUX DIRIGES

LOGICIELS

Installer la dernière version de R et les packages utiles pour les analyses: <https://cran.r-project.org/>

Packages à installer dans R:

- ✓ BiodiversityR
- ✓ Car
- ✓ gam
- ✓ vegan

REFERENCES UTILES

- ✓ Kindt R. and Coe R. 2005. Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre, 2005. Téléchargeable sur: www.worldagroforestry.org/downloads/Publications/PDFS/b13695.pdf
- ✓ Oksanen J., 2016. Vegan: ecological diversity. Téléchargeable sur: <https://cran.r-project.org/web/packages/vegan/vignettes/diversity-vegan.pdf>

FORMAT DES PRESENTATIONS

Modalités de présentation

- ✓ 2h par groupe d'étudiants :
 - Introduction aux conditions d'utilisation de la méthode
 - Présentation du script
 - Exécution de l'analyse dans R
 - Interprétation des résultats
 - Exemple de publication scientifique utilisant la méthode.

3^{ème} Partie

ANALYSES PORTANT SUR LA
RICHESSE SPECIFIQUE

TRAVAUX DIRIGES SUR PC (LOGICIEL R)

- Richesse spécifique = nombre d'espèces
- Référence : Analyse de la richesse spécifique - (Kindt et Coe, 2005; chap 4)

4^{ème} Partie

ANALYSES PORTANT SUR LA
DIVERSITE

TRAVAUX DIRIGES SUR PC (LOGICIEL R)

La diversité implique la richesse (ou le nombre d'espèces et l'équité (ou l'égalité dans le nombre d'individus pour chaque espèce)).

- Indices de diversité - (Kindt et Coe, 2005; chap 5)
- Diversité taxonomique - (Oksanen, 2016; page 3)
- Diversité fonctionnelle - (Oksanen, 2016; page 3)

Diversité

- Smith (2002). Ecological statistics. In El-Shaarawi and Piegorsch (eds), Encyclopedia of Environmetrics. John Wiley & Sons, Ltd, Chichester, Vol 2, pp 589–602

Table 2 Some commonly used **diversity measures**. S = observed number of species or taxa, p_i = the proportion of taxa i in the sample, n_i = the number of taxa i and N = the total abundance

Diversity measure	Formula
Species richness	S
Margalef index	$(S - 1) \ln\{N\}$
Menhinick	$\frac{S}{\sqrt{N}}$
Shannon–Wiener–Weaver	$-\sum_{i=1}^S p_i \ln\{p_i\}$
Simpson	$\sum_{i=1}^S p_i^2$
Berger–Parker	$\frac{\max\{n_i\}}{N}$
Log series α	Requires iterative solution

5^{ème} Partie

ANALYSES DES DONNEES ISSUES
DE COMPTAGES

TRAVAUX DIRIGES SUR PC (LOGICIEL R)

- Diagrammes rangs-fréquences et modèles d'abondance - (Oksanen, 2016; page 4)
- Régression linéaire simple - (Kindt et Coe, 2005; chap 6)
- GLM - (Kindt et Coe, 2005; chap 6)
- Generalized additive regression model – GAM - (Kindt et Coe, 2005; chap 6)
- Régression multiple - (Kindt et Coe, 2005; chap 6)

6^{ème} Partie

ANALYSES DES DONNEES DE PRESENCE-ABSENCE DES ESPECES

TRAVAUX DIRIGES SUR PC (LOGICIEL R)

- GLM - (Kindt et Coe, 2005; chap 7)
- GAM - (Kindt et Coe, 2005; chap 7)

7^{ème} Partie

ANALYSES PORTANT SUR LES DIFFERENCES DE COMPOSITION SPECIFIQUE

TRAVAUX DIRIGES SUR PC (LOGICIEL R)

- Test de Mantel - (Kindt et Coe, 2005; chap 8)
- Test ANOSIM - (Kindt et Coe, 2005; chap 8)
- Classification et Ordination - (Kindt et Coe, 2005; chap 9 & 10)