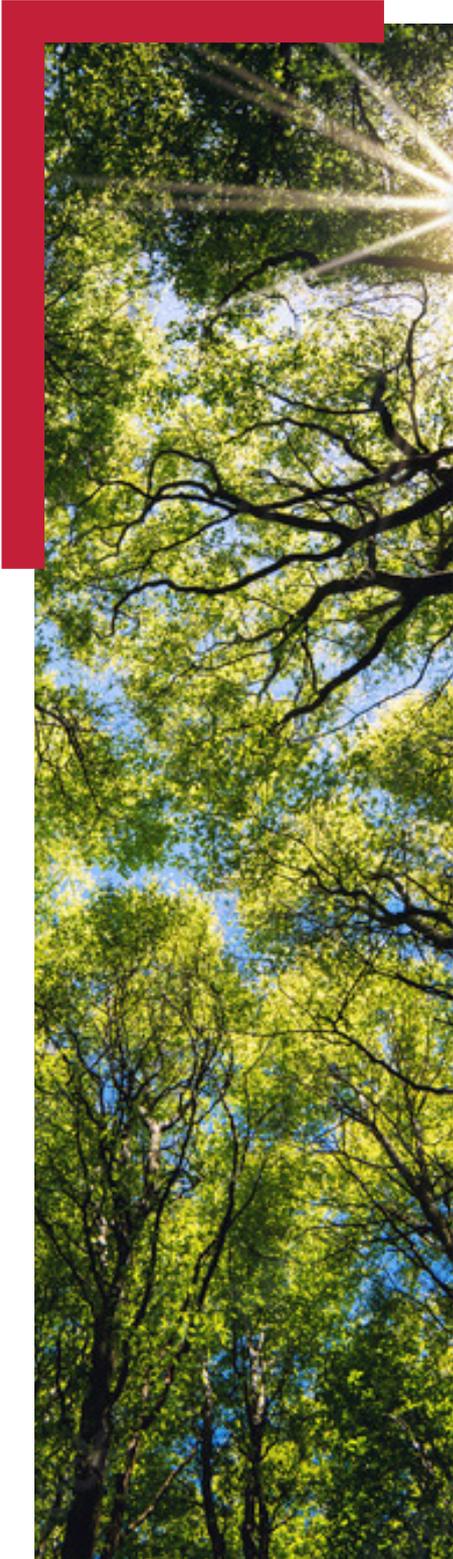




CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE ET LA GESTION DE LA BIODIVERSITÉ



La formation de l'École Pratique des Hautes Études, en analyse de données pour l'écologie et la gestion de la biodiversité ambitionne de vous donner des bases solides en statistiques généralistes afin de vous permettre de construire et d'analyser vos propres jeux de données. Elle vous fournira les compétences nécessaires afin de vous approprier des méthodes plus spécialisées, lorsque vous en aurez besoin : suivis de populations, modèles de distributions d'espèces, analyses génétiques...

La formation est structurée en trois niveaux. Le niveau de découverte est accessible à tous sans prérequis initial. Un test de positionnement vous permet d'évaluer vos compétences préalables si vous souhaitez démarrer aux niveaux suivants.

INFORMATIONS

Formateur : Jean-Yves Barnagaud (jean-yves.barnagaud@ephe.psl.eu)

Lieu : Maison des Sciences de l'Homme, 54 boulevard Raspail 75006 Paris. Métro Sèvres Babylone ou Rennes.

Format présentiel

Horaires : 9h30 - 12h, 13h30 - 17h (sauf le vendredi : fin à 15h30)

Matériel à prévoir :

- Un ordinateur portable (de préférence sous Windows, mais les Mac sont acceptés) avec MS Excel ou tout autre logiciel de tableur.
- Installer au préalable les logiciels gratuits R et R-studio, téléchargeables à ces adresses (contacter le formateur en cas de difficultés) :
<https://cran.r-project.org/>
<https://rstudio.com/products/rstudio/download/>
- Des directives pour l'installation de modules complémentaires vous seront communiquées par le formateur une semaine à dix jours avant le début de la formation.



NIVEAU 1 : DÉCOUVERTE

Une première approche pratique de l'analyse de jeux de données écologiques

Ce premier niveau est une découverte pratique de l'univers des analyses statistiques, telles qu'elles sont le plus souvent pratiquées par les écologues actuellement. Nous aborderons le pourquoi des analyses statistiques, leurs apports et leurs limites, la manière de construire un échantillonnage, d'explorer et de communiquer des données. Enfin, nous commencerons à construire des modèles statistiques en réponse à des questions simples. Nous utiliserons essentiellement des jeux de données réels, prêtés pour cette formation par une diversité de chercheurs et d'organismes que vous connaissez : CNRS, OFB, INRAE, ONF, associations, réserves naturelles...

Nous accorderons les après-midi à la pratique sous le logiciel R, qui est certes un peu difficile au premier regard, mais offre une flexibilité incomparable pour vos futures analyses. En revanche, nous irons peu vers la théorie statistique : en tant qu'utilisateur, votre but premier doit être avant tout de savoir manipuler correctement les concepts et les outils, plus que de les décortiquer dans toute leur complexité. Parce que cette formation est résolument orientée vers la pratique, soyez-en acteurs : questionnez, échangez, critiquez, afin que cette semaine soit la plus riche possible.

Et ensuite ?

Les analyses statistiques sont un domaine vaste et complexe - ceci d'autant plus que les données écologiques et les questions d'étude et de gestion de la biodiversité sont elles-mêmes souvent complexes. Ne vous attendez donc pas à être parfaitement autonome en une semaine de découverte ! A l'issue de ce premier niveau, vous disposerez d'une base solide pour comprendre à quoi sert l'analyse statistique, comment monter un protocole d'échantillonnage et comprendre un résultat statistique. Vous saurez synthétiser des jeux de données simples. Si vous recherchez une véritable autonomie, il vous faudra maîtriser des méthodes plus avancées permettant d'analyser simultanément plusieurs variables : c'est l'objectif du niveau de renforcement.

Pour aller plus loin : le niveau 3 vous procurera une véritable compétence pratique afin d'analyser en autonomie des jeux de données à long terme ou large échelle spatiale : suivis de populations, dynamiques de distributions ou communautés écologiques.

PLANNING

	Heures	Thèmes	Méthodes abordées	Compétences
Lundi	9h30 12h30	Que sont les statistiques et à quoi servent elles?	Intérêt des statistiques, définitions, questions et hypothèses, démarches scientifiques, notion d'erreur, description de données, représentations graphiques	<ul style="list-style-type: none"> • Mettre en place une démarche d'analyse • Connaître le vocabulaire • Interpréter des statistiques descriptives • Choisir une représentation graphique en fonction des données et de la question posée • Structurer un espace de travail ; anticiper le flux de travail
	13h30 17h	Travaux pratiques : formater son jeu de données et connaître les manipulations de base sous R	Préparation et vérification de données en vue de leur analyse, mise en place de l'environnement informatique, découverte de R	
Mardi	9h30 12h30	Pourquoi et comment échantillonner ?	Motivations de l'échantillonnage protocolé, échantillonnages probabilistes et non probabilistes, cas concrets	<ul style="list-style-type: none"> • Construire un plan d'échantillonnage en réponse à une question, adapter un plan d'échantillonnage aux compromis du terrain • Importer et explorer une table de données sous R ; manipuler des tables de données sous R • Identifier les problèmes et y remédier
	13h30 17h	Manipuler des données sous R	Import de données, synthèse de données, filtrage sur critère, formatage et jointures, formats d'export, messages d'erreur	
Mercredi	9h30 12h30	Quantifier la variabilité d'un jeu de données	Descripteurs quantitatifs, impact de la taille d'échantillon, lois de distributions, intervalles de confiance	<ul style="list-style-type: none"> • Interpréter la variabilité d'un jeu de données ; confronter des données à des attendus théoriques • Mesurer l'incertitude sur la moyenne d'une série quantitative • Communiquer des données par voie graphique ; améliorer l'esthétique de ses graphiques • Connaître les deux principaux langages graphiques de R
	13h30 17h	Travaux pratiques : représentations graphiques sous R	Histogrammes, nuages de points, boîtes à moustaches, diagrammes en barres, apparence des graphiques	
Jeudi	9h30 12h30	Relations entre variables : une introduction à la modélisation statistique	Modèle de régression linéaire à une variable, estimation et interprétation des paramètres, tests pour petits échantillons	<ul style="list-style-type: none"> • Répondre à une question à partir de données • Construire le modèle adéquat en fonction d'une hypothèse • Évaluer l'incertitude sur un résultat • Représenter un résultat d'analyse • Implémenter son analyse sous R • Visualiser et comprendre les résultats ; communiquer le résultat d'une analyse
	13h30 17h	Travaux pratiques : implémenter un modèle statistique sous R	Modèles linéaires sous R, vérification et inférence sur les résultats, interprétation et critique des résultats, restitution d'analyse	
Vendredi	9h30 12h30	Questions personnalisées	Tous les thèmes abordés dans la semaine	
	13h30 15h30	Évaluation du niveau 1		