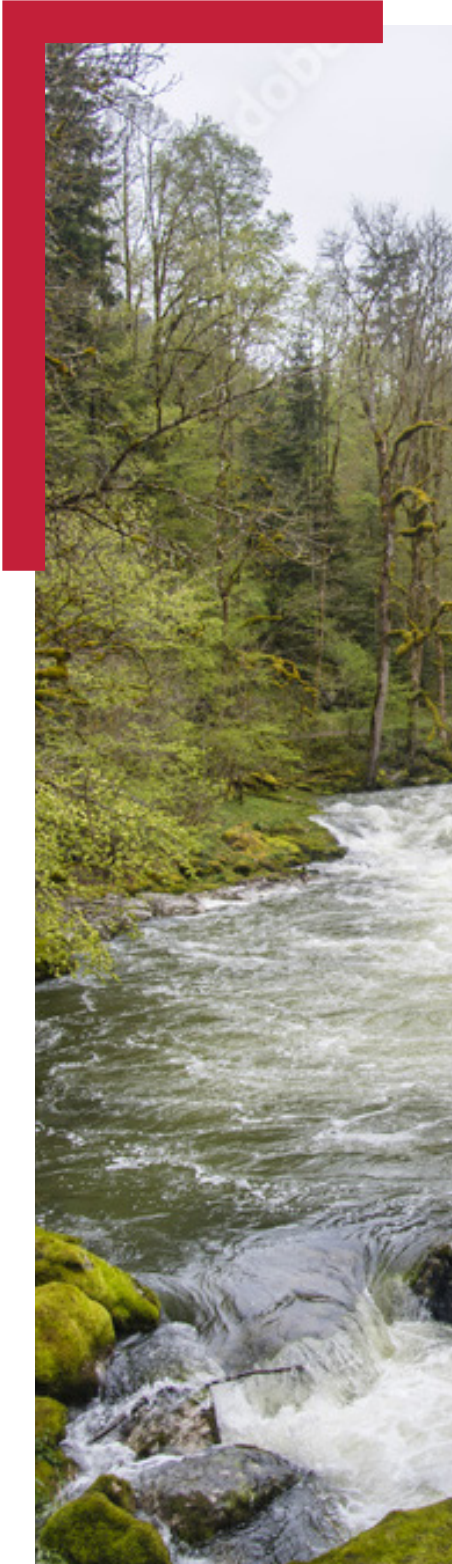




CERTIFICAT EN ANALYSE DE DONNÉES POUR L'ÉCOLOGIE, LE SUIVI DE LA BIODIVERSITÉ ET LA GESTION DES ESPACES NATURELS



L'École Pratique des Hautes Études vous propose une formation en trois niveaux aux méthodes généralistes d'analyse de données appliquées à l'écologie. Ce parcours vous donnera les compétences fondamentales nécessaires à comprendre des résultats d'analyse statistique, puis à construire, analyser et communiquer vous-mêmes vos propres jeux de données écologiques. Vous aborderez les notions fondamentales et les mettrez en application dans un esprit pratique, à partir de véritables cas d'étude.

Vous deviendrez progressivement autonome, dans la mesure de vos besoins : comprendre l'essentiel, mettre en pratique, puis devenir force de proposition. Vous vous préparerez à aborder des méthodes plus spécialisées : modélisation d'abondances ou de dynamiques de populations, modèles de distributions d'espèces, analyses génétiques ou acoustiques...

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 uniquement

Dates en 2025 : 17 - 21 mars

Horaires : 9h30 - 12h30, 13h30 - 17h (sauf lundi : début à 10h, vendredi : fin à 15h30)

Matériel à prévoir :

- Un ordinateur portable 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 bibliothèques R complémentaires vous seront communiquées par le formateur une semaine à dix jours avant le début de la formation (droits administrateur nécessaires pour l'installation).



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 de données** dans le contexte de l'écologie. Nous aborderons leur raison d'être, leurs enjeux et limites, la manière de construire un échantillonnage, d'explorer et de communiquer des données. En fin de formation, 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...

La formation est résolument axée vers **une approche pragmatique** : nous irons peu vers la théorie statistique, car nous pensons qu'en tant qu'utilisateur, votre but premier doit être avant tout de savoir manipuler correctement les méthodes et les outils, plutôt que de les décortiquer dans toute leur complexité. Nous accorderons les après-midi à une initiation de l'exploration de données sous le logiciel R.

Cette formation est principalement destinée à **des participants qui n'ont aucune expérience préalable en analyse de données** ou qui souhaitent recycler des connaissances et compétences fondamentales. Il n'y a donc aucun prérequis. L'objectif pour vous cette semaine n'est pas de devenir autonome sur vos analyses, mais d'appréhender les principales notions qui vous permettront de comprendre les statistiques présentées dans des rapports ou des articles, ou d'interagir plus facilement avec des chercheurs ou chargés d'études scientifiques qui travaillent au quotidien sur des données.

Nous vous incitons à communiquer dès le début de la formation sur les objectifs que vous vous êtes fixé sur cette semaine, afin de personnaliser au mieux les interactions avec le formateur. 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 statistiques appliquées aux données écologiques sont un champ vaste et changeant, d'autant plus que les questions relatives à l'étude et la gestion de la biodiversité sont elles-mêmes 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 encore plus loin, le niveau avancé 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	10h 12h	Pourquoi des analyses statistiques en écologie ?	La démarche statistique, son champ d'application, ce qu'on peut en attendre ou non, grandes notions centrales	<ul style="list-style-type: none"> Comprendre l'intérêt d'une démarche fondée sur l'analyse de données pour l'écologie Connaître les grandes étapes d'une analyse statistique et les principaux résultats à en attendre Maîtriser les éléments clé du vocabulaire Structurer un espace de travail en vue d'une analyse de données
	13h30 17h	Structurer un questionnaire et un jeu de données en vue d'une analyse	Préparation et vérification de données en vue de leur analyse, mise en place de l'environnement informatique, découverte de R	
Mardi	9h30 12h30	Qu'est-ce que la variabilité de données écologique et pourquoi s'en préoccupe-t-on ?	Indicateurs quantitatifs de séries de données, idée de descripteurs et d'estimateurs, notion d'échantillonnage aléatoire et de taille d'échantillon, utilité des lois de distributions, estimation d'un intervalles de confiance	<ul style="list-style-type: none"> Synthétiser et représenter des données au regard d'une question Interpréter la variabilité d'un jeu de données ; confronter des données à des attendus théoriques Mesurer l'incertitude sur la moyenne d'une série de données et comprendre son interprétation Communiquer des données par voie graphique Développer un regard critique sur des représentations de données
	13h30 17h	Représentations graphiques de données	Fonctions graphiques essentielles avec la librairie ggplot2 sous R, réalisation, interprétation et présentation de graphiques	
Mercredi	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 Maîtriser la notion d'aléa d'échantillonnage et sa traduction sous forme d'intervalle de confiance Comprendre et exploiter la relation incertitude - taille d'échantillon
	13h30 17h	L'échantillonnage de données écologiques : planifier et dimensionner	Construction d'un plan d'échantillonnage en réponse à une question, dimensionnement de l'effort en fonction du résultat attendu, idée de test de puissance	
Jeudi	9h30 12h30	Qu'est-ce qu'un modèle statistique ?	Modèle de régression linéaire à une variable, estimation et interprétation des paramètres	<ul style="list-style-type: none"> Répondre à une question à partir de données Comprendre la démarche de modélisation et son intérêt en écologie 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	Implémenter, interpréter et représenter un modèle linéaire 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		