université PARIS-SACLAY

ÉCOLE D'ÉTÉ DE SANTÉ PUBLIQUE ET D'ÉPIDÉMIOLOGIE DE BICÊTRE



Institut National de la Santé et de la Recherche Médicale
Faculté de Médecine Paris-Saclay
Graduate School de Santé Publique Paris-Saclay
Responsable : Marianne Canonico

Du 21 juin au 2 juillet 2027

Ecole d'été de Santé Publique et d'Epidémiologie de Bicêtre - Inserm CESP U1018
82, rue du Général Leclerc - 94276 LE KREMLIN-BICÊTRE cedex, France

Contact : Azadeh Rezaeï - E-mail : ecole-ete.medecine@universite-paris-saclay.fr

Plus d'informations : https://eespe.vjf.inserm.fr

L'école d'Eté de Santé Publique et d'Epidémiologie de Bicêtre comprend des enseignements théoriques dans les domaines de la biostatistique, de l'épidémiologie et de l'informatique et des enseignements pratiques concernant plus particulièrement la formulation, la résolution et la discussion de problèmes concrets de Santé Publique. L'accent est surtout mis sur l'utilisation des connaissances acquises au cours de cet enseignement. Les enseignants exercent une activité de recherche dans une Unité de recherche de l'INSERM, dans un laboratoire universitaire ou dans une Agence de Santé. La participation aux enseignements donne droit à la délivrance d'une attestation.
D'avance, merci pour votre participation !

Marianne Canonico, responsable de l'Ecole d'Eté de Santé Publique et d'Epidémiologie de Bicêtre

---

Personnes concernées

Le programme propose des enseignements en méthodologie et applications pour répondre aux besoins des professionnels de la Santé, décideurs, médecins, chercheurs, qui souhaitent découvrir ou approfondir les bases du raisonnement en sciences de la Santé et son utilisation dans le domaine de la Santé Publique. Tout dossier devra, pour être accepté, avoir été examiné par le Comité de Direction de l'Ecole d'Eté. Le nombre de participants est limité.

Modalités et droits d'inscription

Le formulaire de pré-inscription devra être rempli en ligne avant le 1 juin 2027 (https://eespe.vjf.inserm.fr).

Le règlement du montant total des frais d'inscription est obligatoire pour toute pré-inscription à titre individuel. Le montant total de ces frais figure ci-après :

Tarifs individuels étudiants :

- 750 euros pour un enseignement à temps complet (sur deux semaines)
- 600 euros pour un enseignement à 75% du temps complet
- 450 euros pour un enseignement à mi-temps sur deux semaines, ou à temps plein sur une semaine (modules 1, 2)
- 300 euros pour un enseignement à temps plein sur deux jours et demi ou mi- temps sur une semaine (modules 3, 4, 5)

Tarifs individuels non étudiants :

- 1750 euros pour un enseignement à temps complet (sur deux semaines)
- 1400 euros pour un enseignement à 75% du temps complet
- 1050 euros pour un enseignement à mi-temps sur deux semaines, ou à temps plein sur une semaine (modules 1, 2)
- 700 euros pour un enseignement à temps plein sur deux jours et demi ou mi- temps sur une semaine (modules 3, 4, 5)

Tarifs Formation Permanente :

- 2500 euros pour un enseignement à temps complet (sur deux semaines)
- 2000 euros pour un enseignement à 75% du temps complet
- 1500 euros pour un enseignement à mi-temps.
- 1000 euros pour un enseignement à temps plein sur deux jours et demi ou mi-temps sur une semaine (modules 3, 4, 5)

Pour les inscriptions au titre de la Formation Permanente, consulter le site internet ou contacter le secrétariat. Une Convention de Formation Professionnelle sera établie par la Faculté de Médecine Paris-Saclay.

Après la notification de l'acceptation de votre dossier, un courriel vous sera adressé comprenant les indications pratiques. Les documents pédagogiques seront transmis aux participants **directement par courriel.**

Liste des modules pour 2027

Module 1 : Régression multiple en épidémiologie : régression logistique et modèle de Cox
O. Lambert et A. Nambiema - 1 semaine du 21 au 25 juin, temps complet

Module 2 : Introduction à la modélisation des données longitudinales sous R : modèles marginaux et modèles mixtes
F. Artaud et H. Panjo - 1 semaine du 28 juin au 2 juillet, temps complet

Module 3 : Stratégie de recherche en épidémiologie
M. Canonico - 2 jours et demi du 21 au 23 juin midi, temps complet

Module 4 : Initiation à l'analyse de données avec R (tidyverse)
R. Bauer et R. Foare - 1 semaine du 21 au 25 juin, après-midi

Module 5 : R avancé, automatiser ses rapports d'analyses avec R
R. Bauer et R. Foare - 1 semaine du 28 juin au 2 juillet, après-midi

1. Régression multiple en épidémiologie : régression logistique et modèle de Cox

O. Lambert et A. Nambiema

Objectifs et contenu

L'objectif du cours est de présenter deux des principaux modèles de régression multivariée les plus fréquemment utilisés en épidémiologie et en recherche biomédicale : la régression logistique et le modèle de Cox.

Il s'agira de développer une démarche d'analyse, depuis la vérification des hypothèses inhérentes à chaque modèle, le choix et le codage des facteurs de confusion, jusqu'à l'interprétation des résultats. Certains aspects moins usuels seront abordés (variable dépendante du temps dans le modèle de Cox).

A côté des bases théoriques sous-tendant l'utilisation de ces modèles, de nombreux exemples et situations concrètes seront présentés, notamment sur la base de lectures critiques d'articles. A l'issue de cet enseignement, les étudiants auront acquis les connaissances leur permettant d'utiliser ces modèles de régression en fonction de leur problématique, et d'avoir un regard critique dans l'interprétation des résultats.

Pré-requis

Enseignement de base ou cours de statistique et d'épidémiologie de niveau CESAM, M1 de sciences biomédicales, de santé publique ou de mathématiques. Avoir des connaissances de base concernant le modèle de régression linéaire.

Plan

I. Régression logistique (5 demi-journées) :

- Rappels sur la régression linéaire (2h)
- Régression logistique binaire :
  - Présentation du modèle, estimation et tests des paramètres
  - Codage des variables et interprétation des coefficients
  - Evaluation du modèle
  - Sélection des variables selon l'objectif de la modélisation (étiologique ou prédictif)
  - Introduction à la méthode de sélection de variables par pénalisation LASSO

II. Données de survie (5 demi-journées) :

Ce module est consacré à l'analyse des données de survie, largement utilisées en épidémiologie et en santé publique pour l'étude du temps jusqu'à la survenue d'un événement (décès, maladie, rechute, etc.). Il aborde les méthodes d'estimation et de comparaison des courbes de survie, ainsi que l'utilisation du modèle de Cox pour l'analyse multivariée.

Contenu du module :

- Problématique des données de survie (censure, temps à l'événement)
- Construction des courbes de survie :
  - méthode de Kaplan-Meier
  - méthode actuarielle
- Comparaison des courbes de survie :
  - test du logrank
- Modèle de Cox :
  - hypothèses
  - analyse multivariée (procédures manuelles vs automatiques, étude des interactions)
  - variables dépendantes du temps
- Application des modèles de Cox à l'estimation du risque cardiovasculaire individuel, dans une perspective de médecine personnalisée

2. Introduction à la modélisation des données longitudinales sous R : modèles marginaux et modèles mixtes

F. Artaud et H. Panjo

Objectifs

Dans les études longitudinales où l'on mesure pour différents individus à des échéances successives une réponse biologique ou clinique, il est possible de modéliser l'évolution de cette réponse au cours du temps ainsi que l'influence des caractéristiques des sujets sur cette évolution. Cependant il existe une corrélation entre les mesures d'un même sujet qui devra être prise en compte dans l'analyse statistique.

L'objectif de ce module est de présenter, d'une façon pratique et appliquée, les méthodes et les modèles statistiques les plus utilisés pour analyser ce type de données : d'une part, les modèles mixtes, et d'autre part, les modèles GEE (Generalized Estimating Equations).

Pré-requis

Ce cours nécessite une bonne connaissance théorique et pratique niveau M2 santé publique des modèles de régression usuels (les modèles de régression linéaire et logistique).

Programme

- Caractéristiques des données longitudinales et implications
- Modélisation des évolutions moyennes (données continues et discrètes)
- Utilisations des modèles à effets mixtes
- Utilisations des modèles marginaux et GEE

Modalités pratiques

Les applications seront réalisées avec R. Les codes de programmation et les sorties seront fournis pour être décrits et interprétés pendant le cours.

3. Stratégie de recherche en épidémiologie

M. Canonico

Objectifs

Décrire et comparer les différentes stratégies de recherche qui peuvent être utilisées dans l'approche d'un problème de santé en population (étude cas-témoins, cohorte, essai d'intervention).

Cet enseignement interactif résumera les points clés d'une étude épidémiologique en prenant comme exemple l'évaluation des effets secondaires d'un traitement. L'apport de la pharmacoépidémiologie dans l'évolution des pratiques médicales sera également discuté.

Pré-requis

Connaissances de base en méthodologie statistique et épidémiologie.

Programme

Stratégies de recherche en population et élaboration d'un protocole.
Bonnes pratiques en recherche épidémiologique.
Régression logistique conditionnelle et modèle de Cox avec exposition dépendante du temps (applications pratiques à partir de sorties SAS).
Interprétation des résultats (biais, puissance, analyse quantitative des données de la littérature, ...) et écriture scientifique.
Impact de données nouvelles sur les pratiques médicales (agences sanitaires, médias,..).

Séance 1 : Stratégies de recherche en population
Elaboration d'un protocole I – Approche Cas-Témoins

Séance 2 : Analyse d'une étude cas-témoin
Interprétation d'une étude cas-témoins

Séance 3 : Elaboration d'un protocole II – Etude de cohorte (protocole et modèle de Cox avec exposition dépendante du temps)
Comparaison cohorte/cas-témoins

Séance 4 : Principe des méta-analyses et applications - interaction/ajustement avant les méta-analyses

Séance 5 : Elaboration d'un protocole III – Réaliser une étude de cohorte de A à Z

Enseignement sur 2 jours et demi : les cours auront lieu du lundi matin au mercredi midi.

4. Initiation à l'analyse de données avec R (introduction au tidyverse)

R. Bauer et R. Foare

Objectifs

Etre autonome avec le logiciel R pour l'importation d'un tableau de données, la génération de variables, la réalisation de statistiques descriptives, et des tests et modélisations les plus courants en épidémiologie et recherche clinique.

Le but de ce module n'est pas de discuter des résultats d'analyses statistiques mais d'apprendre à utiliser les commandes pour réaliser ces analyses.

Public concerné

Le cours est prévu pour les « vrais débutants » en programmation R. En revanche, des bases théoriques sont requises : théorie des tests, principaux tests univariés, régression linéaire

Pré-requis

Connaissances théoriques en Biostatistique et Epidémiologie :

- au minimum Licence pro en statistique ou M1 de Santé Publique (CESAM ou équivalent)
- idéalement M2 de Santé Publique - Epidémiologie ou Recherche clinique

Connaissances pratiques :

- environnement Windows : pratique courante indispensable.

Programme

- Prise en main du logiciel R et de l'environnement RStudio.
- Manipulation d'une base de données : Importation d'une table, tri sur une variable, sélection de lignes selon des conditions définies, fusion de plusieurs tables, génération de nouvelles variables, etc.
- Production de statistiques descriptives simples et de graphiques.
- La manipulation de base de données et la production de graphiques seront présentées avec l'extension *tidyverse*. L'extension *tidyverse* permet d'effectuer un grand nombre d'opérations courantes et vise à faciliter le travail de nettoyage et de préparation de bases de données, préalable à l'analyse statistique.

Modalités pédagogiques

Cours en salle d'informatique.

Un stagiaire par ordinateur.

Possibilité d'utiliser son ordinateur personnel, avec le logiciel R et l'environnement RStudio préalablement installés (installation gratuite).

5. R avancé, automatiser ses rapports d'analyses avec R

R. Bauer et R. Foare

Objectifs

Le but de ce module n'est pas de discuter des résultats des analyses statistiques mais d'apprendre à utiliser les commandes pour réaliser ces analyses.

Public concerné

Utilisateur de R souhaitant améliorer l'esthétique de leur rapport de résultats et gagner du temps en automatisant la réalisation de certaines tâches et la production de documents scientifiques.

Pré-requis

Connaissance et utilisation préalable de R.

Programme

- Production de tableaux « prêts à être publiés » avec l'extension *gtsummary.*
- Générer un document word, pdf ou html avec Rmarkdown.
- Initiation à la programmation (boucles, fonction apply, lapply, sapply, création de fonctions).

Modalités pédagogiques

Cours en salle d'informatique.
Un stagiaire par ordinateur.
Possibilité d'utiliser son ordinateur personnel, avec le logiciel R et l'environnement RStudio préalablement installés (installation gratuite).

Enseignants et membres du comité de direction

ARTAUD F.	INSERM CESP U1018 – Exposome et hérédité
BAUER R.	INSERM – Centre Régional en Antibiothérapie Ile de France
CANONICO M.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie
CLARES VERA S.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie
FOARE R.	INSERM SC10 – US19
LAMBERT O.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie
MEYER L.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie
NAMBIEMA A.	INSERM U970 – Paris – Centre de recherche Cardiovasculaire (PARCC)
PANJO H.	INSERM CESP U1018 – Soins primaires et prévention et santé des femmes
REZAEÏ A.	EDSP
ROBERGEAU F.	INSERM CESP U1018
SERA B.	INSERM SC10 US19
TEGLAS J.P.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie
TRAN L.	INSERM CESP U1018 – Epidémiologie clinique et pharmacoépidémiologie