



Centre d'Analyse et de Traitement des données Statistiques

MANUEL DE FORMATION

Formation pratique en analyse statistique avec R

I. Présentation de CAT-Stat

Le Centre d'Analyse et de Traitement des données Statistiques (CAT-Stat) est une structure spécialisée dans la formation, l'accompagnement méthodologique et l'analyse statistique appliquée aux projets académiques, professionnels et institutionnels. Il intervient aussi bien auprès des étudiants et chercheurs que des professionnels, des ONG et des institutions publiques ou privées.

La création de CAT-Stat en 2020 est née d'un constat largement partagé dans les milieux académique et professionnel : bien que de nombreuses données soient produites à travers les enquêtes, les systèmes d'information et les projets de recherche, leur analyse, leur interprétation et leur valorisation scientifique restent souvent limitées. Ces difficultés freinent la production de connaissances fiables et l'utilisation effective des données dans la prise de décision.

CAT-Stat a ainsi été conçu pour répondre à ce besoin à travers des formations rigoureuses, accessibles et résolument orientées vers la pratique, tout en tenant compte des réalités locales et des standards internationaux. La structure s'appuie sur un réseau de consultants internationaux, basés dans plusieurs pays, disposant d'une expertise reconnue en biostatistique, épidémiologie, santé publique, suivi-évaluation et analyse de données.

La mission de CAT-Stat est de renforcer durablement les capacités en analyse statistique et en traitement des données, en favorisant une utilisation autonome, rigoureuse et pertinente des

outils analytiques. Elle intervient également dans la conception, la planification, le suivi et l'évaluation des projets, notamment dans le domaine de la santé.

La vision de CAT-Stat est de devenir une référence en formation et en accompagnement en statistique appliquée, au service de la recherche scientifique, de la santé publique et de la prise de décision fondée sur des données probantes.

II. Contexte et justification de la formation

L'analyse statistique occupe aujourd'hui une place centrale dans la recherche scientifique, l'évaluation des programmes et l'élaboration des politiques publiques. Le développement des systèmes d'information, la multiplication des enquêtes et l'essor de la recherche génèrent des volumes de données de plus en plus importants, nécessitant des outils puissants, flexibles et reproductibles.

Dans ce contexte, le logiciel **R** s'est imposé comme une référence internationale en analyse statistique et en data science. Sa gratuité, sa robustesse, sa capacité à produire des analyses reproductibles et la richesse de son écosystème de packages en font un outil privilégié aussi bien dans le monde académique que professionnel. Cependant, malgré ses nombreux avantages, l'apprentissage de R est encore perçu par beaucoup comme complexe, technique et peu accessible, en particulier pour les débutants.

La présente formation a été conçue pour démystifier l'utilisation de R, en proposant une approche progressive, pédagogique et fortement orientée vers la pratique. Elle vise à permettre aux participants de dépasser la simple exécution de commandes pour comprendre la logique de l'analyse des données et être capables d'appliquer R à leurs propres projets.

Par ailleurs, une enquête préalable menée auprès de participants potentiels a mis en évidence une forte préférence pour le format en ligne, jugé plus accessible et mieux compatible avec les contraintes académiques, professionnelles et géographiques. Ce constat a guidé le choix d'un dispositif de formation principalement à distance, tout en maintenant une forte interaction entre formateurs et apprenants.

III. Présentation générale de la formation

La formation intitulée « **Passer à la vitesse supérieure avec R : formation pratique en analyse statistique** » s'étend sur une durée de deux mois, du **10 janvier au 15 mars 2026**. Elle se déroule principalement en ligne, sous forme de sessions interactives en direct, permettant aux participants de suivre les cours en temps réel, de poser des questions et de travailler sur des exercices pratiques.

Cette formation s'adresse aux étudiants de niveau licence, master ou doctorat, aux chercheurs ainsi qu'aux professionnels amenés à utiliser des données dans le cadre de leurs activités. Elle ne se limite pas à un public de statisticiens, mais vise toute personne souhaitant acquérir des compétences solides et opérationnelles en analyse de données avec R.

IV. Objectifs pédagogiques et compétences visées

L'objectif général de la formation est de rendre les participants autonomes dans l'utilisation du logiciel R pour l'analyse statistique de données académiques ou professionnelles.

À l'issue de la formation, les participants seront capables d'utiliser efficacement R et l'environnement RStudio, d'importer et de structurer des bases de données issues de différentes sources, de nettoyer et transformer les données, y compris la gestion des données manquantes, et de produire des visualisations claires, informatives et scientifiquement pertinentes. Ils seront également en mesure de réaliser les principaux tests statistiques, d'en choisir les méthodes appropriées selon le contexte d'étude et d'interpréter correctement les résultats obtenus.

V. Contenu et organisation pédagogique de la formation

La formation est structurée en quatre modules progressifs, répartis sur l'ensemble du calendrier, alternant des séances de cours interactives et des séances spécifiquement dédiées aux travaux dirigés, exercices pratiques et études de cas.

Le premier module, consacré à la prise en main de R et RStudio, se déroule lors des premières séances de la formation. Il est centré sur l'installation et la configuration des outils, la découverte de l'environnement de travail et les bases du langage R. Les participants apprennent à se familiariser avec la syntaxe, les objets et les premières manipulations simples.

Le deuxième module porte sur la manipulation et la gestion des données. Il occupe une place centrale dans la formation et s'étend sur plusieurs séances. Les participants apprennent à importer des données à partir de différents formats, à nettoyer et transformer les bases de

données, à recoder des variables et à gérer les données manquantes. Une part importante de ce module est consacrée à la pratique, avec des jeux de données réels et des séances de travaux dirigés.

Le troisième module est dédié à la visualisation des données. Il permet aux participants de comprendre les principes fondamentaux d'une bonne visualisation statistique et de produire des graphiques de qualité professionnelle à l'aide du package ggplot2. Les séances alternent entre démonstrations et exercices pratiques, afin de permettre une appropriation progressive des outils.

Le quatrième module est consacré à l'analyse statistique. Il aborde les principaux tests statistiques usuels, les critères de choix des méthodes selon les types de données et les objectifs d'analyse, ainsi que l'interprétation correcte des résultats. Des séances spécifiques sont réservées à l'application des tests sur des données réelles et à la discussion des résultats obtenus.

Tout au long de la formation, certaines séances sont volontairement réservées exclusivement aux travaux dirigés, exercices pratiques et études de cas, afin de consolider les acquis et d'accompagner les participants vers une réelle autonomie.

VI. Calendrier de la formation

Modules et contenus	Date
Module 1 – Introduction générale à la formation et prise en main de R et RStudio : présentation de la formation, installation et configuration de R et RStudio, découverte de l'environnement de travail, premières commandes et objets de base	11/04//2026
Module 1 – Bases du langage R : types d'objets, fonctions, scripts, logique du langage R et bonnes pratiques de travail	12/04/2026
TD pratique (Module 1) : exercices guidés sur la prise en main de RStudio, manipulation des objets, résolution d'exercices simples et correction collective	18/04/2026

Module 2 – Importation et structuration des données : importation de données depuis différents formats (CSV, Excel), exploration des bases de données et structuration des variables avec les commandes (« <i>head</i> », « <i>str</i> », « <i>summary</i> », <i>etc...</i>)	19/04/2026
Module 2 – Nettoyage et transformation des données : recodage des variables, création de nouvelles variables, filtrage et sélection des données	25/04/2026
Module 2 – Gestion des données manquantes : identification des valeurs manquantes, stratégies de gestion et implications sur l'analyse	26/04/2026
TD pratique (Module 2) : travaux dirigés sur des bases de données réelles, nettoyage complet d'une base, exercices appliqués et accompagnement personnalisé	02/05/2026
TD pratique (Module 2) : approfondissement, correction détaillée des exercices, application sur les données des participants	03/05/2026
Module 3 – Principes de visualisation des données : règles de base d'une bonne visualisation statistique, choix des graphiques selon les données	09/05/2026
Module 3 – Visualisation avec ggplot2 : construction de graphiques, personnalisation, amélioration de la lisibilité et interprétation	10/05/2026
TD pratique (Module 3) : réalisation de graphiques à partir de données réelles, exercices guidés et études de cas	16/05/2026
TD pratique (Module 3) : application avancée de ggplot2, correction collective et optimisation des visualisations	17/05/2026

Module 4 – Introduction à l'analyse statistique : rappels conceptuels, types de tests statistiques et conditions d'application	23/05/2026
Module 4 – Tests statistiques usuels avec R : tests t de student, test de Wilcoxon, test du Chi2, etc..., et interprétation des résultats	24/05/2026
TD pratique (Module 4) : application des tests statistiques sur des jeux de données réels, interprétation guidée des résultats	30/05/2026
TD pratique (Module 4) : exercices approfondis, discussion des résultats et erreurs fréquentes	31/05/2026
Séance intégrative : analyse complète d'un cas pratique de A à Z (importation, nettoyage, visualisation et analyse statistique)	06/06/2026
TD encadré : révision générale, clarification des difficultés rencontrées et application sur les données personnelles ou professionnelles des participants	07/06/2026
Séance d'examen : EVALUATION FINALE (TEST ECRIT)	13/06/2026
Séance de clôture : restitution des corrections, échanges, évaluation de la formation ; clôture officielle et remise des attestations de formation	14/06/2026

VII. Méthodologie pédagogique

La formation adopte une approche **résolument pratique**, fondée sur des démonstrations en temps réel, **des exercices appliqués à chaque séance et des études de cas issues de situations réelles**. Les formateurs accompagnent progressivement les participants afin de leur permettre de gagner en autonomie, tout en restant disponibles pour répondre aux questions et apporter un appui méthodologique.

VIII. Prérequis et moyens matériels

La formation ne requiert aucun niveau avancé en programmation. Des connaissances de base en statistique sont toutefois recommandées afin de faciliter la compréhension des concepts abordés. Chaque participant doit disposer :

- d'un ordinateur portable,
- d'une connexion Internet stable,
- les logiciels R et RStudio, vous seront installés au début de la formation avec les packages nécessaires.

IX. Organisation, évaluation et attestation

Les séances sont organisées conformément au calendrier communiqué à l'avance. L'évaluation repose essentiellement sur la participation active aux séances et la réalisation des exercices pratiques. À l'issue de la formation, une **attestation de participation** est délivrée aux participants ayant suivi régulièrement les activités et ayant obtenu au moins une note **12/20 (note minimum)**.

X. Coût et modalités de paiement

Le coût de la formation est fixé à **100 000 FCFA**. Les modalités vous seront exposées lors de la 1^{ere} séance de la formation.

XI. Présentation des formateurs

La formation est animée par une équipe de formateurs expérimentés, composée de médecins, chercheurs et spécialistes en santé publique et analyse de données.



Dr Moulaye Berthe, MD, MPH

Dr Moulaye Berthe est médecin généraliste, titulaire d'un Master of Public Health (MPH), parcours Méthodes Quantitatives et Économétriques pour la Recherche en Santé (MQERS), obtenu à Aix-Marseille Université avec la distinction honorifique *Cum Laude*.

Lauréat du concours « Mon stage en 180 secondes » pour la meilleure présentation scientifique, il est certifié en modélisation macroéconomique pour le développement durable par l'Institut Africain de Développement Économique et de Planification (IDEP) à Dakar, ainsi qu'en suivi-évaluation de projets de développement par l'Agence Française de Développement.

Il a exercé comme épidémiologiste et responsable de recherche pour Médecins Sans Frontières France au Mali, puis comme Chargé MEAL sur des projets financés par l'USAID. Il est actuellement Coordinateur de Recherche au sein de l'ONG Muso Health, où il pilote des études opérationnelles et stratégiques visant à renforcer les systèmes de santé et l'accès équitable aux soins.



Dr Zoumana Cheick Bérété, MD, MPH, MSc

Dr Zoumana Cheick Bérété est médecin généraliste, titulaire d'un MPH en Informatique de Santé et Systèmes d'Information en Santé (ISSIS) obtenu à la FMOS, et d'un MSc en pharmacoépidémiologie de l'Université Laval (Canada).

Professionnel de recherche au CHU de Québec – Université Laval, il dispose d'une solide expérience en enseignement, recherche clinique et analyse de données.

Au sein de CAT-Stat, il intervient dans la formation en analyse de données et la planification des projets de recherche.



Dr Alahaye Mahamane Maïga, MD, MPH

Dr Alahaye Mahamane Maïga est médecin généraliste, titulaire d'un MPH (parcours MQERS) obtenu à Aix-Marseille Université en 2024. Il possède des compétences avancées en biostatistique, modélisation et analyse de données biomédicales.

Il travaille actuellement dans un laboratoire de recherche de niveau de confinement ACL2, spécialisé dans la biologie moléculaire et la surveillance des arthropodes d'importance médicale.



Dr Barth Lowe, MD, MPH

Dr Barth Lowe est médecin depuis 2023 et titulaire d'un MPH (parcours MQERS) d'Aix-Marseille Université. Il poursuit actuellement un Master spécialisé en santé globale au sein du programme Pasteur–CNAM à Paris.

Il a participé à plusieurs projets de recherche en Afrique de l'Ouest et en Afrique centrale, notamment en analyse spatiale du paludisme, ainsi que dans des programmes liés au VIH



Dr Sidy Coulibaly, MD, MPH student

Dr Sidy Coulibaly est médecin, diplômé de la Faculté de Médecine et d'Odonto-Stomatologie (FMOS) du Mali. Il est en M2-MPH (parcours MQERS) à Aix-Marseille Université.

Il est également certifié en modélisation macroéconomique pour le développement durable par l'IDEP et occupe actuellement le poste de Chargé de vérification du mécanisme Direct Health Facility Financing (DHFF) au sein de l'ONG Muso Health

Le PDG du CAT-Stat

A handwritten signature in blue ink, appearing to read 'Moulaye'.



Dr Moulaye BERTHE, MD, MPH