



Ingénieur, chercheur ou technicien supérieur au sein d'un service de R&D ou d'une unité de recherche dans les domaines de l'agronomie ou de l'agroalimentaire, vous êtes confronté à l'exploitation de données expérimentales, voire à la mise en place de dispositifs expérimentaux.

Modulaire et progressif, combinant formation à distance et présentiel, ce stage vous apportera à la fois les outils et les éléments méthodologiques permettant de manipuler, explorer, présenter et exploiter statistiquement vos jeux de données.

Objectifs pédagogiques

- Découvrir le **langage R et l'environnement RStudio**. Manipuler des données avec R. Rédiger des scripts.
- Connaître les principales **approches univariées et bivariées** de statistique descriptive ; savoir les mettre en œuvre à l'aide de R et RStudio.
- Savoir choisir et mettre en œuvre les principales **méthodes de statistique inférentielle en expérimentation agronomique** ; savoir choisir et mettre en œuvre les dispositifs expérimentaux classiques en agronomie ; savoir en traiter les résultats et en présenter les résultats.
- Savoir choisir et mettre en œuvre **les principales méthodes multivariées** pour l'exploration de données statistiques.

Déroulé pédagogique

Les parties A, B et C ci-dessous sont dissociables, mais les compétences acquises à la partie A (2j) sont un prérequis aux parties B (2j) et C (1j).

Partie A – Auto-formation à R et aux statistiques descriptives (2 jours, à distance¹)

Jour 1 – Découverte de R et RStudio

- I. **Prendre en main le langage R et l'environnement de développement RStudio**
 - Télécharger, installer, découvrir R et RStudio
 - Editer, manipuler un script R
 - Manipuler des variables (numériques, chaînes de caractères, vecteurs...)
 - Manipuler des fonctions, utiliser l'aide en ligne
 - Installer des Packages
- II. **Manipuler un jeu de données avec R**
 - Créer et importer un fichier CSV sur R
 - Visualiser un jeu de données individus/variables
 - Extraire un sous ensemble de variables ou d'individus
 - Modifier, sauvegarder un jeu de données.

Jour 2 – Statistique descriptive uni et bivariée

- III. **Mettre en œuvre des outils de descriptive univariée avec R**
 - Connaître, identifier, modifier les types de variables
 - Présenter une variable qualitative sous forme de tableau ou de graphique (diagrammes en barres, en secteurs)
 - Présenter une variable quantitative sous forme de résumés numériques (paramètres de tendance centrale et de dispersion, statistiques de rang, mise en classe) ou graphiques (histogrammes, box-plots)

¹ La durée de 2 jours est donnée ici à titre indicatif. C'est une durée de formation estimée. En pratique, cette première partie pourra débuter par une séance introductive courte puis s'étaler sur une période plus longue selon votre rythme d'apprentissage.

IV. Mettre en œuvre des outils de descriptive bivariable avec R

- Croiser deux variables qualitatives : produire et normaliser un tableau de contingence, le représenter graphiquement à l'aide de diagrammes en barres.
- Croiser une variable qualitative et une variable quantitative à l'aide de résumés numériques et graphiques (e.g. box-plots)
- Croiser deux variables quantitatives à l'aide de résumés numériques (covariance et coefficients de corrélation, paramétriques et non paramétriques) et graphiques (nuages de points)

Partie B – Statistique et expérimentation (2 jours, en présentiel)

Jour 3 – Principes d'échantillonnage, d'estimation et de tests hypothèses

V. Echantillonnage et estimation

- Notions de moyenne, d'écart-type, de variance ; lois usuelles (normale, binomiale...)
- Estimation ponctuelle d'une moyenne, d'une proportion, d'une variance
- Estimation par intervalle de confiance d'une moyenne, d'une proportion, d'une variance

VI. Tests d'hypothèses

- Comprendre les objectifs et principes des tests d'hypothèses
- Liens entre puissance d'un test, taille d'échantillon et variabilité
- Connaître, mettre en œuvre et interpréter les tests d'hypothèses usuels : test binomial, test de Student, test du khi2
- Vérification des conditions d'application et alternatives non paramétriques

Jour 4 – ANOVA et dispositifs en expérimentation agricole

VII. Analyse de Variance (ANOVA)

- Mettre en œuvre et analyser une ANOVA à 1 facteur et ses tests post-hoc
- Savoir mettre en œuvre et interpréter une ANOVA à 2 facteurs avec et sans interaction
- Savoir vérifier les conditions d'applications, appliquer une transformation ou choisir une alternative non paramétrique

VIII. Dispositifs expérimentaux

- Connaître les dispositifs expérimentaux usuels, savoir les mettre en œuvre et en traiter les résultats

Partie C – Analyse de données multivariées (1 jour, en présentiel)

Jour 5 – Méthodes multivariées pour l'analyse de variables quantitatives

IX. Analyse en composantes principales (ACP)

- Connaître les objectifs et le principe d'une ACP, savoir la mettre en œuvre et en interpréter les résultats

X. Classification non supervisée (K-means et classification ascendante hiérarchique)

- Connaître les principales méthodes de classification non supervisée. Savoir choisir, mettre en œuvre ces méthodes et en interpréter les résultats

OPTION dans le cadre d'une formation intra-entreprise, sur mesure

Partie D – Support à distance post formation

A l'issue de la formation, une assistance est apportée aux stagiaires afin de faciliter l'appropriation et la mise en pratique des outils et méthodes abordées dans leur contexte professionnel. L'équipe de formateurs se rendra disponible pour répondre aux sollicitations des stagiaires. L'assistance sera apportée, de façon asynchrone, sur un forum dédié mis en place sur la plate-forme pédagogique de l'établissement. Les supports de formations et autres supports pédagogiques seront mis à disposition des stagiaires sur cette même plate-forme.

- **Durée** : 5 jours (35h) dissociables en 2 j (14h à son rythme) + 2 j (14h) + 1 j (7h)
- **Stage proposé en** : inter-entreprises ou intra-entreprise
- **Dates** : voir le « **Calendrier des formations courtes** » (rubrique Formation Continue du site web)
- **Mode hybride** : 2 j distanciel (A)+ 2 j présentiel (B)+ 1 j présentiel (C)
- **Niveau** : découverte (A) à approfondissement (B et C)

- **Publics cibles** : techniciens ou ingénieurs en agriculture et environnement
- **Pré-requis ; conditions d'accès** : notions élémentaires en statistiques et traitement de données
- **Modalités d'évaluation** : test de positionnement (niveau, attente), évaluation QCM de l'atteinte des objectifs (pratique de R, statistiques descriptives, statistiques inférentielles, statistiques multivariées.), questionnaire de satisfaction
- **Equipe pédagogique** :
 - Jean-Pierre Da Costa, Professeur, Bordeaux Sciences Agro
 - Florent Arnal, Professeur agrégé, Université de Bordeaux
 - Lionel Bombrun, Maître de Conférences, Bordeaux Sciences Agro
- **Modalités pédagogiques**

A : Partie distancielle (asynchrone) :

 - utilisation de la plateforme pédagogique Moodle,
 - supports interactifs pour l'autoformation aux logiciels R et RStudio,
 - assistance asynchrone de l'équipe pédagogique sur un forum

B+C : Parties présentiels (synchrone) :

 - alternance d'apports théoriques et de mise en pratique sur logiciel
 - utilisation des logiciels R et RStudio

Dans les deux cas, auto-évaluations sous formes de QCM (sur Moodle).
- **Nombre de personnes mini-max par session** : de 6 à 15 personnes

Tarif : Partie A (distanciel) et/ou Partie B et C (présentiel)

- ✓ *inscription individuelle* : voir dates et tarifs sur le « **Calendrier des formations courtes** »
- ✓ *sur mesure – intra-entreprise* : nous contacter en écrivant à partenariats@agro-bordeaux.fr

Modalités de financement : plan de développement des compétences, OPCO, Pôle Emploi, financement personnel, ...

Responsable pédagogique : Da Costa Jean-Pierre , jean-pierre.dacosta@agro-bordeaux.fr

Contact inscriptions et financement: formco@agro-bordeaux.fr / 05 57 35 07 67