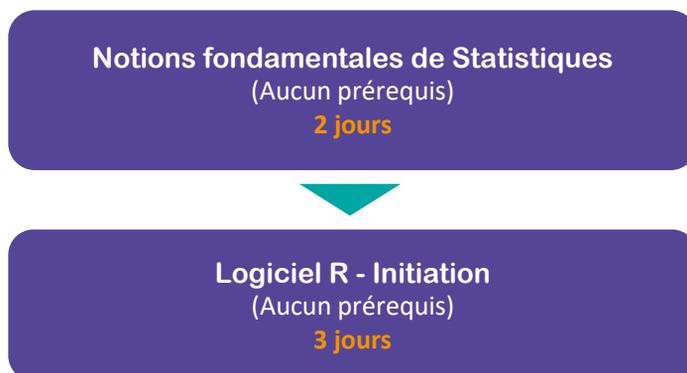


Chemin pédagogique - Formations Statistiques

Notions fondamentales

1



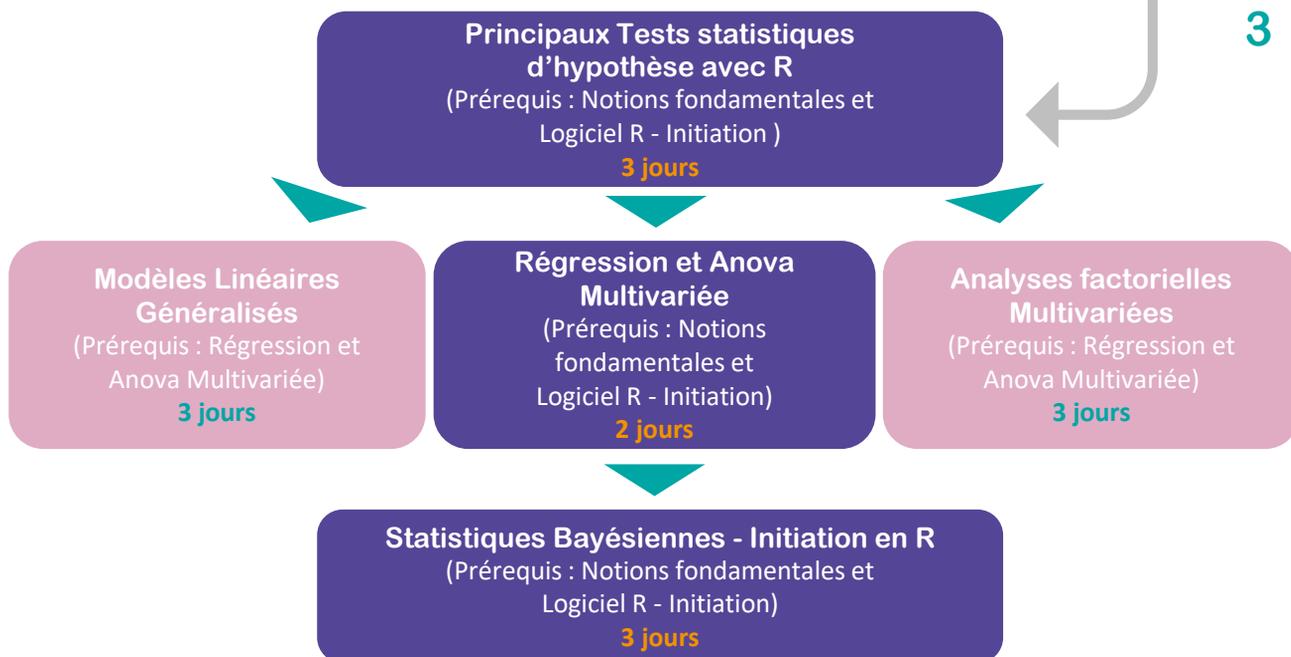
Approfondissement R

2



Statistiques avec R

3



NOTIONS FONDAMENTALES DE STATISTIQUES

Référence : 2021- STAT-N1-NF2J

2 jours (14h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Personne souhaitant s’initier aux fondamentaux de la statistique.

Prérequis

Aucun prérequis statistique particulier n’est nécessaire. L’usage d’un tableur type Excel est suffisant.

Objectifs pédagogiques

A l’issue de cette formation, l’apprenant sera capable de :

-  Connaître les différents types de données
-  Acquérir les notions élémentaires en statistique (moyenne, médiane, écart-type...)
-  Décrire synthétiquement et graphiquement une série de mesures quantitatives
-  Mesurer et interpréter le lien entre deux caractères
-  Produire et interpréter des graphiques descriptifs adaptés aux données

La formation en quelques mots

Généralités

Cette formation a essentiellement pour objectif de sensibiliser des utilisateurs à la pratique des techniques statistiques de base. Elle peut constituer la première approche de la pratique des statistiques et pourra être suivie de modules complémentaires de perfectionnement.

La réussite de cette formation est donc très importante puisqu’elle conditionne la qualité des premiers pas des apprenants en statistique.

Plus que d’objectifs opérationnels, il s’agit avant tout d’une formation dans laquelle l’enjeu est important. En effet, pour beaucoup de participants, cette formation est souvent la première approche des outils statistiques.

De plus, lorsque des apprenants ont des préjugés sur la statistique, ceux-ci ne sont pas toujours favorables envers cette dernière :

-  Outil trop mathématique,
-  Outil trop théorique,
-  On fait dire ce qu’on veut avec la stat, ...

Il est donc important que cette formation soit l’occasion de bien démarrer avec la statistique et qu’il en ressorte avant tout l’envie d’aller plus loin.

Cet objectif sera atteint si les apprenants :

-  Comprennent que la statistique est avant tout un outil complémentaire et support à leur métier plutôt qu’une finalité en soi.
-  Comprennent que les outils statistiques peuvent être mis en œuvre sans forcément maîtriser les fondamentaux mathématiques sous-jacents.
-  Perçoivent que l’utilisation des techniques statistiques peut offrir une réelle plus-value lorsqu’elle est panachée avec des compétences métiers.

Au-delà des compétences statistiques pures, l'important dans ce type de formation est de comprendre les mécanismes fondamentaux présents dans la plupart des manipulations des outils statistiques :

D'une façon générale, ces mécanismes sont les suivants :

-  Savoir à quoi sert un outil statistique précis,
-  Savoir dans quelles conditions on peut l'utiliser,
-  Interpréter les résultats
-  Attribuer un sens physique à l'outil statistique

Par exemple, dans le calcul d'un écart-type, élément clé présent dans un grand nombre de méthodes statistiques, il nous paraît plus important d'appréhender le sens physique de l'écart-type plus que la formule mathématique en traitant les points suivants :

-  Que mesure un écart-type ?
-  Dans quelle unité de mesures s'exprime un écart-type ?
-  A partir de combien de valeurs, je peux raisonnablement mesurer un écart-type ?
-  De quel écart-type je parle (s , σ , ...) ?

Les séquences et les compétences acquises :

Deux grandes séquences seront présentes dans cette formation.

- Une première séquence où on cherchera à acquérir de la compétence Méthodologique ET opérationnelle (Statistiques de base via l'usage d'un tableur)
- Une deuxième séquence où on cherchera à acquérir essentiellement de la compétence Méthodologique sans viser le côté « opérationnel ». Ceci concernera entre autres la thématique des tests d'hypothèses.

Ainsi, les chapitres 2 et 3 détaillés ci-dessous « TRAITEMENT STATISTIQUE BIVARIÉ : EXTENSION DES NOTIONS DE BASE À 2 VARIABLES » et « INTRODUCTION AUX TESTS D'HYPOTHÈSES : PRÉSENTATION DES BASES DE LA THÉORIE DES TESTS D'HYPOTHÈSE » seront animés sous la forme d'une ouverture.

Les autres modules de formation proposés permettront alors d'acquérir la compétence intégrant le savoir-faire notamment avec l'usage du logiciel R.

Programme

1 - Traitement statistique univarié : notions de base

-  Les objectifs de la statistique
-  Le vocabulaire de base
-  Présentation des différents types de variables (qualitatives et quantitatives)
-  Caractéristiques numériques : définition, calcul et interprétation ex : moyenne, médiane, écart-type, variance, étendue, quantiles
-  Caractéristiques des variables qualitatives : définitions, calcul, interprétation et usages
-  Représentations graphiques : bonnes pratiques et interprétation

2 - Traitement statistique bivarié : extension des notions de base à 2 variables

-  Tableaux de contingence
-  Représentations graphiques : bonnes pratiques et interprétation
-  Fréquences conditionnelles
-  Corrélation et causalité
-  Nuages de points, régression linéaire simple

3 - Introduction aux tests d'hypothèses : présentation des bases de la théorie des tests d'hypothèses

-  Hypothèses, niveau, puissance, région critique, p-value
-  Test sur la moyenne
-  Test sur la variance
-  Test sur proportion
-  Test de χ^2

Méthodes et moyens

-  Explications théoriques suivies de pratiques guidées puis de mises en autonomie
-  Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
-  Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
-  Votre formation se déroule sur notre plate-forme de formation avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.

LOGICIEL R - INITIATION

Référence : 2021- STAT-N1-INITR3J

3 jours (21h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant prendre en main le logiciel R, s'initier aux bases de l'utilisation et à la mise en œuvre d'analyses statistiques de bases.

Prérequis

Avoir suivi la formation « notions fondamentales de statistique » ou avoir un niveau équivalent.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

-  Exécuter et exploiter des scripts simples
-  Créer et manipuler des objets sous R (vecteurs, facteurs, dataframe)
-  Importer des données
-  Créer et modifier un graphique simple
-  Mettre en œuvre des analyses statistiques de base sous R

La formation en quelques mots

Cette formation s'adresse à des personnes souhaitant prendre en main le logiciel R.

Précisons que l'animation s'appuiera sur l'usage de R Studio. Les compétences acquises concerneront :

-  L'interface de l'outil
-  La configuration
-  La syntaxe
-  La gestion des données via les dataframes
-  Les bases de la programmation

Il ne s'agit pas d'une formation sur les méthodes statistiques proprement dites mais bien d'une formation sur le logiciel R. En outre, il ne s'agit pas d'une formation sur les techniques de programmation avancée sous R type création de packages.

A l'issue de la formation, le stagiaire sera capable de mettre en œuvre des scripts simples sur la manipulation et l'analyse de données peu complexes.

Deux volets principaux constitueront l'ossature de cette formation :

-  **Volet 1 :**
Un volet orienté sur la compétence purement R permettant la maîtrise de l'interface et surtout de la manipulation des données
-  **Volet 2 :**
Un volet orienté sur la mise en place d'analyses qts que l'on peut qualifier de niveau 1 (statistiques descriptives)

Le volet 2 s'inscrira parfaitement en complément du module n° 1 nommé « NOTIONS FONDAMENTALES DE STATISTIQUES. Il peut être vu comme le volet pratique des éléments méthodologiques du module 1.

L'ensemble de ces deux compétences sera fédéré par un exercice pratique de synthèse organisé sous forme d'atelier mixant les compétences de manipulations de données et de traitements statistiques.

Les éléments de bonnes pratiques de R

R est un logiciel basé sur une approche de lignes de commandes de programmation.

Quelles sont les qualités que l'on attendues d'un script, c'est-à-dire d'un programme au sens informatique :

- Bien faire ce qu'on lui demande (ce qui est déjà pas mal !)
- Faire rapidement ce que l'on lui demande
- Pouvoir être maintenu dans le temps et être compréhensible l'utilisateur concepteur mais aussi par d'autres utilisateurs

C'est pourquoi, au-delà de l'aspect technique proprement dit, il nous paraît important que l'accent soit mis sur l'acquisition d'une démarche visant à intégrer les **bonnes pratiques** de l'usage de R.

Les éléments de bonnes pratiques concerneront aussi bien :

-  L'usage de syntaxe explicite pour les objets
-  La clarté, l'aération des « scripts »
-  La mise en place de commentaires
-  Le paramétrage de l'interface (.Rprofile, ...)
-  L'usage de la vectorisation, clé de voute du logiciel R

Ces bonnes pratiques faciliteront l'usage de R et permettront l'appropriation de R dans le cadre d'un usage collectif en facilitant les échanges ultérieurs entre praticiens.

Outil logiciel

Par essence même, cette formation s'appuie sur le couple R / RStudio (principalement ce dernier).

En marge de la formation, il nous paraît également intéressant de montrer quelques outils libres s'appuyant sur R et permettant de disposer d'une interface conviviale pour faciliter l'accès à l'usage d'analyses statistiques.

Si cette suggestion est retenue, citons par exemple un outil tel que Jamovi.

Programme

1 - Introduction au logiciel R

-  Configuration du logiciel R/R Studio
-  Utiliser l'aide en ligne (recherche de commandes R)

2 - Principes généraux du langage R, principales structures de données, objets et fonction

-  Vecteurs
-  Types de données
-  Dataframe : structure de base des traitements statistiques
-  Listes
-  Packages

3 – Principaux éléments de R studio

- 🔄 Object browser, packages, fenêtre de scripts
- 🔄 Sauvegarder l'espace de travail, charger un fichier sauvegardé
- 🔄 Travailler en mode projet
- 🔄 Importer dans une dataframe. Exporter un fichier CSV/Excel

4 - Langage R : prise en main des objets R

- 🔄 Extraire des parties d'un vecteur
- 🔄 Concaténer, supprimer, insérer
- 🔄 Manipulation de différents types de données
- 🔄 Regrouper/ réordonner les niveaux d'un facteur
- 🔄 Extraire des parties d'une matrice
- 🔄 Extraire des parties d'une liste
- 🔄 Dataframe en tant que liste et en tant que matrice
- 🔄 Gérer les données manquantes

5 - Statistiques descriptives : principaux résumés statistiques sur des données numériques ou nominales

- 🔄 Médiane, moyenne, quantiles
- 🔄 Distribution statistique
- 🔄 Corrélations et covariances
- 🔄 Tableaux de contingences

6 - Graphiques : visualiser des données multi-variées sous R

- 🔄 Histogrammes
- 🔄 Box-plots, dot plots, scatter plot
- 🔄 Graphiques en barre
- 🔄 Graphiques en mosaïques
- 🔄 Fonction qplot de ggplot2

7 - Exercice pratique : réalisation d'un mini-projet (analyse de données et production du rapport)

- 🔄 Description de chaque variable
- 🔄 Corrélation et corrélations partielles
- 🔄 Représentation graphique multidimensionnelle

Méthodes et moyens

- 🔄 Explications théoriques suivies de pratiques guidées puis de mises en autonomie
- 🔄 Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
- 🔄 Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
- 🔄 Votre formation se déroule sur notre plate-forme de formation avec :

- 1 ordinateur
- 1 connexion Internet
- 1 adresse e-mail valide
- 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.

APPROFONDIR LE LANGAGE R

Référence : 2021- STAT-N2-PERFR2J

2 jours (14h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant consolider ses acquis en maîtrisant l'usage des fonctionnalités avancées du logiciel R.

Prérequis

Avoir suivi la formation « logiciel R initiation » ou avoir un niveau équivalent.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

-  Approfondir sa connaissance du langage R
-  Automatiser une chaîne de traitement complexe (tableaux, graphiques)

La formation en quelques mots

Même si la formation « N° 2 » en 3 jours « Logiciel R Initiation » permet d'acquérir une autonomie dans la manipulation des données ainsi que dans la mise en œuvre des analyses statistiques « simples », elle ne permet pas d'accéder à un usage avancé du logiciel R dans des thématiques telles que la manipulation de grands jeux de données, la mise en œuvre d'automatisations de traitements statistiques ou la mise en place de graphiques élaborés.

Ce module nommé « Approfondir le langage R » répond à ces besoins.

L'animation commencera par un rappel des fonctionnalités de base puis enchainera sur l'acquisition de nouvelles compétences proprement dite.

Plusieurs axes de compétences seront développés :

-  Les instructions de programmation proprement dite visant à l'automatisation de tâches
-  La manipulation avancée des tables de données en s'appuyant notamment sur le package `data.table`
-  La mise en place de graphiques avancés avec notamment l'usage avancé du package `ggplot2`

Précisions qu'il ne s'agit pas d'une formation sur les méthodes statistiques proprement dites mais bien d'une formation sur le logiciel R.

Les éléments de bonnes pratiques de la programmation sous R

Deux approches seront appréhendées lors de cette formation concernant le volet

« automatisation » :

-  Une approche classique visant à utiliser des outils tels que boucles, fonctions conditionnelles, ...
-  Une approche complémentaire visant à « essayer » d'éviter l'usage des outils classiques

On pourrait penser que ces deux approches sont contradictoires.

En fait paradoxalement, nous montrerons que ces approches sont complémentaires mais d'une manière générale, nous inviterons les apprenants à s'appuyer sur des instructions permettant d'optimiser leurs codes tant sur le plan du nombre de lignes que sur les temps d'exécution. C'est pourquoi l'usage des fonctions de type « apply », de la vectorisation ainsi que le développement de l'approche du package « magrittr » feront partie des points clés de la formation.

Parmi les autres bonnes pratiques sur lesquelles nous nous proposons d'insister, il y a aussi la règle consistant à stipuler qu'avant d'écrire soi-même une tâche avancée notamment de manipulations de données sous R par le biais de scripts de programmation, il est nécessaire de vérifier que des fonctions accessibles via des packages du Cran ne sont pas directement disponibles.

En d'autres termes, pourquoi écrire des lignes de codes pour réaliser une tâche que R sait très bien réaliser !

Enfin, nous insisterons également dans ce module sur l'importance de réaliser soi-même ses propres fonctions d'automatisation afin de les utiliser comme des fonctions R classiques avec passage d'arguments, ...

Outil logiciel

Par essence même, cette formation s'appuie sur le couple R / RStudio (principalement ce dernier).

Nous insistons sur le fait que dans cette formation des packages clés tels que magrittr, data.table ou ggplot2 seront très utilisés.

Programme

1 - Environnement R

-  Valeurs NA et NULL
-  Opérations vectorisées
-  Fonctions apply : appliquer une fonction à chaque élément d'une liste/data-frame/matrice
-  Facteurs et tables
-  Fonction grep

2 - Eléments de programmation avancée

-  Création de fonctions
-  Environnements local / with / within
-  Opérations vectorisées, boucles, instructions conditionnelles
-  Passage par valeur et par référence à l'aide des environnements
-  Package magrittr, style de programmation fonctionnel
-  Techniques de débogage : utilisation de log, fonction browser, assertions, tests unitaires

3 - Techniques avancées de traitement de données : utilisation du package data table

-  Opérations ensemblistes
-  Fusion de data frames et leur manipulation avec dplyr et data.table
-  Filtrage
-  Opérations par groupe

 Reformatier une dataframe : melt et cast

4 - Représentations graphiques complexes : package ggplot2 niveau avancé

-  Géométries, esthétiques, statistiques, échelles, coordonnées, données, annotations, thèmes
-  Graphiques en treillis (FACETING)
-  Génération dynamique d'un rapport au format word/pdf/html

Méthodes et moyens

-  Explications théoriques suivies de pratiques guidées puis de mises en autonomie
-  Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
-  Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
-  Votre formation se déroule sur notre plate-forme de formation avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.

PRINCIPAUX TESTS STATISTIQUES AVEC R

Référence : 2021- STAT-N2-PERFTESTS3J

3 jours (21h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant maîtriser la mise en œuvre des tests statistiques « classiques » avec une mise en application sous R.

Prérequis

- 🌀 Avoir suivi la formation « notions fondamentales en statistiques » ou avoir des connaissances équivalentes (moyenne, médiane, écart-type, variance, étendues, quantiles...)
- 🌀 Avoir suivi la formation « logiciel R initiation » ou avoir des connaissances équivalentes de base en R (vecteurs, types de données, dataframe, listes, packages, objets de R...)

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- 🌀 Savoir choisir un test adapté aux questions que l'on se pose et aux données dont on dispose
- 🌀 Savoir évaluer l'hétérogénéité d'un échantillon, comparer deux ou plusieurs échantillons, mesurer l'association entre plusieurs variables
- 🌀 Savoir mettre en œuvre les tests et interpréter les sorties logicielles

La formation en quelques mots

Le module de formation n°1 « Notions fondamentales de statistiques » abordait la notion des tests d'hypothèses via une ouverture sur cette thématique en fin de module.

Dans ce module, on cherche clairement à atteindre la maîtrise et l'autonomie sur la pratique des tests d'hypothèses avec une mise en application sous le logiciel R.

Deux axes de compétences seront développés :

- 🌀 Un axe consacré à la méthode des tests d'hypothèses proprement dite
- 🌀 Un axe consacré à la mise en pratique des tests d'hypothèses sous R

Il s'agit donc clairement d'une formation à dimension méthodologique importante.

Sur l'axe « Méthodologie Statistique », l'accent sera mis sur les éléments suivants :

- 🌀 Un test d'hypothèse correspond à un besoin spécifique doit être en phase avec la question posée (question posée au sens biologique, physique, du praticien, ...)
- 🌀 Un test d'hypothèse doit être en cohérence avec le protocole et la démarche expérimentale associée
- 🌀 Un test d'hypothèse s'utilise dans des conditions de validité données par la "théorie statistique"

Lien entre tests d'hypothèses et métier des praticiens

Dans la démarche des tests d'hypothèses, nous pensons important de ne pas réduire l'interprétation des résultats d'un test à la simple lecture de la p-value. Même si la p-value a un rôle bien sûr important, un comportement simplement mécanique de type $p < 0.05$, significatif, $p > 0.05$ pas significatif nous paraît dommageable.

Nous insisterons donc sur le sens des mots : que veut dire significatif en expliquant que significativité statistique n'est pas forcément preuve de significativité physique.

Au-delà de la panoplie des tests statistiques (choix maîtrisé de l'utilisateur), une vigilance particulière sera apportée sur les éléments portant sur la pertinence des conclusions statistiques notamment dans le cas classique des tests d'hypothèses ou H_0 n'a pas été rejetée.

La question fondamentale que l'on se pose est :

Nous n'avons pas rejeté H_0 , c'est à dire que nous n'avons vu de différence entre des groupes par exemple ... Mais avons-nous les moyens de voir une différence compte tenu de la taille de nos échantillons, de la variabilité de la mesure, de la qualité des données dont nous disposons, de la nature du protocole expérimental, ... ?

Nous soulevons alors le problème de la pertinence de l'outil statistique mis en œuvre. Cette problématique est particulièrement vraie lorsque les apprenants se situent dans un des cas de figure ci-dessous :

- Personnes non maîtres du protocole expérimental mais plutôt participants,
- Héritiers de travaux réalisés par des tiers
- Lecteurs de conclusions de travaux statistiques.

Le contexte particulier des petits échantillons

Nous insisterons également sur le cas particulier de la mise en place des tests statistiques dans le contexte de petits échantillons.

Cette spécificité nous amène ainsi à aborder certains thèmes tels que :

-  Problème de puissance
-  Hypothèses fondamentales délicates à vérifier
-  Fragilité des jeux de données

Outil logiciel

Par essence même, cette formation s'appuie sur le couple R / RStudio (principalement ce dernier).

Programme

1. Introduction aux tests d'hypothèses : se familiariser avec le vocabulaire et les définitions des tests d'hypothèse

-  Vocabulaire
-  Typologie des types de tests
-  Risque d'erreur
-  Puissance du test
-  Bonnes pratiques

2. Tests pour variables assimilables à une normale : revue des principaux tests applicables à des données approximativement normales

-  Tests de moyennes (t-test, z-test, T 2 Hotelling)
-  Tests de variance
-  Tests de comparaisons de moyennes (z-test, t-test, T 2, Westlake Schuirmann)
-  Tests de comparaisons de variances (F-test, t-test, Levene, Bartlett)

3. Tests de distribution

- Variables qualitatives (CHI 2, Fisher)
- Variable quantitatives (Kolmogorov Smirnov)

4. Test de normalité et transformation

5. Tests non paramétriques (wilcoxon, mannwhitney)

6. Tests multiples : contrôler le risque associé à une famille de tests

Méthodes et moyens

- Explications théoriques suivies de pratiques guidées puis de mises en autonomie
- Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
- Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
- Votre formation se déroule sur notre plate-forme de formation avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.

REGRESSION ET ANALYSE DE VARIANCE MULTIVARIEE

Référence : 2021- STAT-N2-PERFREGANOVA2J

2 jours (14h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant maîtriser la mise en œuvre des modèles de régression et d'Anova avec une mise en application sous R.

Prérequis

Avoir suivi les formations « notions fondamentales en statistiques » et « principaux tests statistiques avec R », ou avoir un niveau équivalent.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- Comprendre, maîtriser et mettre en œuvre les techniques usuelles de régression, y compris logistique et l'analyse de variance au-delà du cas à une variable.

La formation en quelques mots

Ce module de formation propose d'aborder la problématique de la modélisation via deux outils très importants en statistique : l'analyse de la variance et la régression

Dans les modules de formation proposés en amont, des notions tels que corrélation ou les comparaisons de moyennes ont déjà permis de s'approcher de ces méthodes.

Dans ce module, clairement c'est la modélisation qui est au cœur de la mise en place de ces outils.

Même si Anova et régression sont des outils frères ou cousins ! que ce soit dans la mise en place sous R ou dans la présentation des résultats restitués, une attention particulière sera apportée à la différence entre Anova et régression :

- Différence en termes de formalisme
- Différences en termes d'objectifs
- Différences en termes de mise en place sous R (problèmes liés à une syntaxe commune, voire ambivalente)

Nous insisterons également sur le fait qu'un outil de modélisation se doit d'être en phase avec la question ou problématique métier posée et avec le protocole expérimental réalisé.

En d'autres termes, l'anova et la régression ne sont qu'une façon de traduire les données collectées lors de l'expérimentation.

Deux axes de compétence seront traités lors de cette formation.

- Un premier axe visant clairement à acquérir l'autonomie (régression et Anova classiques multivariées)
- Un deuxième axe s'orientant vers une ouverture des modèles plus avancés :
 - Modèles mixtes en compléments de l'Anova
 - Régression logistique en complément de la régression

Outil logiciel

Par essence même, cette formation s'appuie sur le couple R / RStudio (principalement ce dernier). En marge de la formation, il nous paraît également intéressant de montrer quelques outils libres s'appuyant sur R et permettant de disposer d'une interface conviviale pour faciliter l'accès à l'usage d'analyses statistiques.

Si cette suggestion est retenue, citons par exemple un outil tel que Jamovi.

Programme

1. Approche multivariée non structurée : une version multivariée du t-test

- Lois multidimensionnelles (normale, Wishart, etc...)
- Tests d'hypothèses (T 2 de Hotelling...)
- Modèle à 1 et 2 échantillons

2. Analyse de variance multivariée

- Estimation de paramètres
- Tests d'hypothèses
- Analyse de profil
- Analyse de courbes de croissance

3. Introduction aux modèles mixtes : le modèle mixte comme moyen de prise en compte de mesures répétées

- Modèle linéaire général
- Modèle linéaire mixte
- Varying intercept, varying slope
- Régression logistique

Méthodes et moyens

- Explications théoriques suivies de pratiques guidées puis de mises en autonomie
- Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
- Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
- Votre formation se déroule sur notre plate-forme de formation avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.

STATISTIQUES BAYESIENNES INITIATION EN R

Référence : 2021- STAT-N2-PERFBAYES3J

3 jours (21h) – Présentiel ou À distance

Population visée : Ouvrier – Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant maîtriser la mise en œuvre des tests statistiques « classiques » avec une mise en application sous R.

Prérequis

Avoir suivi la formation « logiciel R initiation » et la formation « régression et analyse de variance » ou équivalent

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

-  Comprendre le principe de la méthode bayésienne et ce qui la distingue des tests statistiques classiques
-  Savoir utiliser les packages R (RJAGS, Coda,..)
-  Savoir tester des hypothèses pour un équivalent bayésien des modèles usuels (régression, t-test, anova, régression logistique, tableau de contingence, ...)

La formation en quelques mots

Les modules de statistique précédents (Notamment Module 4 & 5) traitant des tests statistiques et des outils de modélisation s'appuient sur une approche classique dite inférentielle.

Dans ce module de formation, l'orientation technique est différente puisque que l'on va traiter des problématiques de modélisations ou de comparaisons via l'approche dite Bayésienne.

Cette formation s'adresse à des personnes déjà spécialisées dans le domaine statistique à travers les méthodes classiques inférentielles.

Deux orientations seront prises lors de cette formation :

-  L'apprentissage des concepts de base de l'approche bayésienne
-  La comparaison entre les méthodes classiques « fréquentistes » et méthodes bayésiennes, illustrant que l'approche bayésienne peut être utilisée comme une approche inférentielle complète
-  La distinction entre ce qui est vraiment utile en statistique Bayésienne, et ce qui l'est moins
-  L'apprentissage de l'utilisation des réseaux bayésiens, qui sont un outil de modélisation très puissant, plus souple en termes de possibilités de modélisation que les approches fréquentistes classiques.

Ainsi, les avantages et les inconvénients des approches « Bayésienne » et « classique » seront présentés et discutés au fil de la formation.

Outil logiciel

Par essence même, cette formation s'appuie sur le couple R / RStudio (principalement ce dernier). Les packages principalement utilisés seront RJAGS, Coda, WINBUGS, ...)

Programme

1. Introduction à l'approche bayésienne

- 🌀 L'approche fréquentielle (classique)
- 🌀 L'approche bayésienne
- 🌀 Le théorème de Bayes
- 🌀 Trouver la distribution a posteriori : facile en théorie, difficile en pratique
- 🌀 Inférence a posteriori
- 🌀 Utilisation séquentielle du théorème de Bayes
- 🌀 Normalité asymptotique
- 🌀 Méthodes numériques en bayésien
- 🌀 Classique ou Bayésien : une synthèse

2. Application pratique avec RJAGS et CODA : estimation d'une proportion

- 🌀 Par la solution analytique
- 🌀 Par MCMC (Markov chain Monte Carlo)
- 🌀 Modélisation hiérarchique d'une proportion
- 🌀 Comparaison de modèles
- 🌀 Test d'hypothèses
- 🌀 Taille d'échantillon

3. Équivalent bayésien de méthodes classiques : application de la méthode bayésienne à des analyses courantes

- 🌀 tests
- 🌀 régressions
- 🌀 Travailler en mode projet
- 🌀 analyse de variance

Méthodes et moyens

- 🌀 Explications théoriques suivies de pratiques guidées puis de mises en autonomie
- 🌀 Votre formation a lieu en présentiel :
 - 1 vidéoprojecteur par salle
 - 1 ordinateur
- 🌀 Votre formation se déroule à distance avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide
 - 1 équipement audio (micro et enceintes ou casque)
 - 1 Webcam (facultatif – dans l'idéal)
 - 1 deuxième écran (facultatif – dans l'idéal)
- 🌀 Votre formation se déroule sur notre plate-forme de formation avec :
 - 1 ordinateur
 - 1 connexion Internet
 - 1 adresse e-mail valide

- 1 équipement audio (micro et enceintes ou casque - facultatif)

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.