



Formations, études,
et conseil en statistiques

Catalogue 2025



Ne jouez pas aux dés
avec vos données

www.anastats.fr

SOMMAIRE

<u>Notre équipe</u>	<u>p. 3</u>
<u>Les Etudes de données et conseil</u>	<u>p. 5</u>
<u>18 formations dispensées par des formateurs agréés</u>	<u>p. 6</u>
Notions fondamentales en statistiques _____	<u>p. 7</u>
Introduction au logiciel R pour les statistiques _____	<u>p. 9</u>
Traitement statistique des petits échantillons _____	<u>p. 11</u>
Pratique et interprétation des analyses de variance _____	<u>p. 13</u>
Construction et analyse des plans d'expériences _____	<u>p. 15</u>
Décrire et analyser les données multifactorielles _____	<u>p. 17</u>
Machine learning _____	<u>p. 19</u>
Introduction aux modèles mixtes linéaires et non linéaires _____	<u>p. 21</u>
Modélisation paramétrique, niveau avancé _____	<u>p. 23</u>
Analyse statistique des données textuelles avec IRaMuTeQ _____	<u>p. 25</u>
Utilisation avancée du logiciel R pour les statistiques _____	<u>p. 27</u>
Introduction à l'analyse de données à haut débit avec le logiciel R _____	<u>p. 29</u>
Les statistiques bayésiennes : introduction et pratique _____	<u>p. 31</u>
Introduction à l'analyse et à la modélisation des séries temporelles _____	<u>p. 33</u>
Les études sensorielles : éléments de méthodologie et analyses statistiques _____	<u>p. 35</u>
Manipulation et visualisation des données avec le logiciel R _____	<u>p. 37</u>
Créer des applications Shiny avec le logiciel R _____	<u>p. 39</u>
Ateliers de traitement de données personnelles _____	<u>p. 41</u>

Notre équipe, notre expérience

Docteurs et chercheurs en biologie.

AnaStats est composée d'une équipe de quatre docteurs en biologie, dont les domaines de compétences sont différentes et complémentaires : psychologie, neurosciences, éthologie, biologie marine et écologie.

Chaque membre a reçu une initiation différente aux statistiques ce qui est une force pédagogique pour notre équipe. En effet nous avons appris les statistiques dans le cadre de formations initiales diplômantes jusqu'à un niveau BAC+5, dans le cadre de la formation professionnelle mais aussi en tant qu'utilisateurs des statistiques pour l'analyse de nos propres données lors de nos activités de recherche respectives. Toujours en quête de nous perfectionner, nous assurons une veille technologique qui nous permet de proposer des solutions innovantes à nos clients.



AnaStats collabore également au développement de logiciels d'analyse des données

Formations, études et conseil pour :

- l'INSERM, le CNRS, l'INRAE, l'Institut Pasteur, CIRAD, CEA, IRSN, l'ANSES, l'IFREMER ...
- Les Universités : Bordeaux, Lyon, Paris-Diderot, Poitiers, Tours...
- Les laboratoires L'Oréal, BioSol, Sanofi, CIBA, Aventis-Pasteur, Atlantic Bone Screen, IRIS PHARMA, ReproPharm, ATLANTHERA...
- SUEZ, VEOLIA...



Les formations à l'utilisation des statistiques que nous vous proposons ne sont pas ordinaires

Trois particularités essentielles :

1. Elles sont dispensées par des UTILISATEURS des statistiques

- ☑ Issus des sciences biologiques expérimentales, les formateurs sont forts de nombreuses années d'application des statistiques dans divers domaines de recherche. Ils participent régulièrement à la recherche de solutions statistiques aux problèmes des expérimentalistes. Ils ont assuré pendant plus de dix ans des formations en direction des chercheurs et techniciens de l'industrie et du service public, dont ils partagent le langage et les préoccupations.
- ☑ Ils ont perfectionné au fil des années un enseignement des statistiques rigoureux au plan théorique et néanmoins accessible sans compétence particulière en mathématiques.

2. Elles sont tournées vers L'EFFICACITE

- ☑ Le choix de la méthode statistique la plus adaptée à tel ou tel problème est une question essentielle à résoudre en amont de toute analyse. Pourtant, force est de constater que cette étape pose souvent bien des problèmes aux utilisateurs. Explications et exercices pratiques donneront ici une solide formation aux participants.
- ☑ La connaissance des méthodes statistiques ne suffit pas : les stagiaires doivent comprendre les sorties fournies par les logiciels qu'ils utilisent. Elles sont passées en revue, examinées, analysées. L'outil de calcul est mis à la portée de l'utilisateur.
- ☑ Des exercices pratiques sont réalisés par les stagiaires eux-mêmes sur ordinateur, permettant une première familiarisation avec les logiciels et une confrontation aux problèmes concrets rencontrés lors des analyses.
- ☑ Les cours sont faits de manière vivante, en interaction avec les stagiaires, à l'écoute de leur demande. Les exemples traités sont apportés par les participants eux-mêmes, à partir de leurs préoccupations quotidiennes.

3. Elles incluent les APPROCHES MODERNES

La puissance de calcul des micro-ordinateurs a permis le développement de nouvelles techniques statistiques, très bien adaptées aux situations concrètes rencontrées par les utilisateurs : validations des résultats par ré-échantillonnages, calculs adaptés aux petits échantillons par les méthodes exactes de permutations, exploration des données multifactorielles, échantillonnages Monte Carlo dans des chaînes de Markov. Ces progrès récents sont mis à la portée de tous.

Service Etudes de données et Conseil en statistiques

Très nombreux sont les chercheurs, les ingénieurs, les industriels, les gestionnaires confrontés à des solutions statistiques difficiles à mettre en œuvre ou disposant de tableaux de données complexes dont « *il y aurait certainement des informations intéressantes à tirer* »

COMPETENCES TECHNIQUES

Disposant des meilleurs outils pour l'analyse des données et de l'expérience nécessaire à leur utilisation, nous pouvons réaliser tout type de traitement.

- ✓ Recherche des variables les plus sensibles à un traitement
- ✓ Quelles sont, parmi toutes les mesures effectuées, celles qui sont corrélées, et de quelle manière ?
- ✓ Trouver ce qui caractérise au mieux tel groupe d'individus
- ✓ Modélisations
- ✓ Traitement des données à haut débit
- ✓ Statistiques bayésiennes
- ✓ Analyses de questionnaires
- ✓ Traitement des questions ouvertes

COMMUNICATION

Ecoute d'une demande qui n'est pas toujours formulée d'emblée en termes statistiques précis, rapports d'analyses facilement utilisables, lisibles par les non spécialistes, sont le résultat de notre grande expérience dans ce domaine, avec des clients issus de milieux très diversifiés (industrie, commerce, recherche scientifique, sociologie...)

CONSEIL

Les statistiques se décident avant la récolte des données

Evitez de vous retrouver dans une impasse au moment de l'analyse de vos résultats. Faites appel à nous avant la mise en place de votre protocole expérimental.

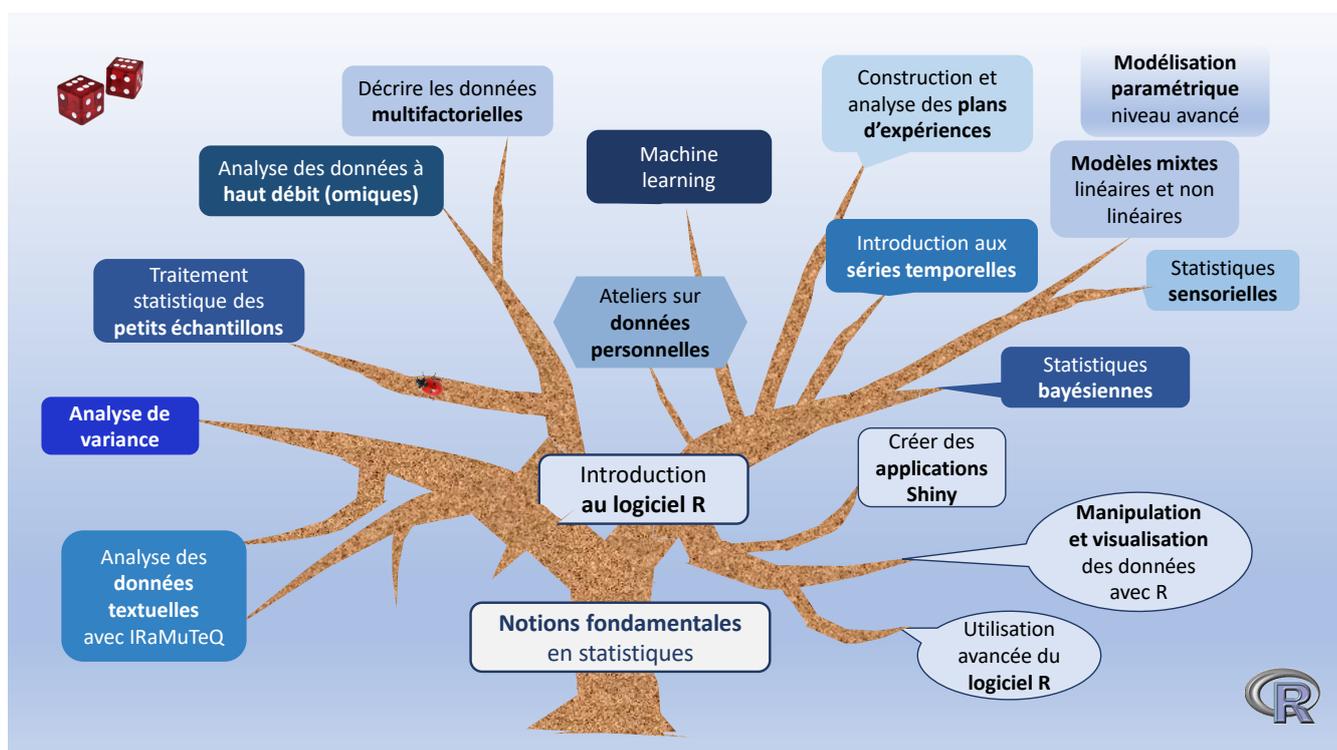
Nous pouvons également vous conseiller pour vos plans d'expériences ou analyses statistiques à tous les stades de votre étude.

**Vous avez des données,
ne laissez pas se perdre le résultat de vos investigations,
ne les abandonnez pas au fond d'un tiroir,
n'hésitez pas à nous en parler !**

18 formations dispensées par des formateurs agréés

Vous trouverez ci-dessous l'organisation de nos formations, et de l'atelier.

Ce schéma permet de mieux appréhender les liens entre ces enseignements et de mieux comprendre les pré-requis détaillés dans les pages suivantes.



Qualiopi
processus certifié

REPUBLICQUE FRANÇAISE

La certification qualité a été délivrée au titre de la catégorie d'action suivante : actions de formation.

Méthode générale

Les formations consistent en des parties théoriques, puis dirigées (démonstration par l'exemple), puis de mise en situation lors d'applications pratiques (exercices en autonomie avec correction de groupe).

Évaluation

L'évaluation se fait au travers de la réalisation d'exercices en autonomie et d'un questionnaire final de validation des acquis.

1. Notions fondamentales en statistiques

Public concerné

Toute personne ayant à traiter des données, caractériser ou comparer des échantillons, analyser des résultats d'expériences, mais ne disposant pas des connaissances de base en statistiques.

Prérequis

Savoir utiliser un tableur tel que Excel, cela sera évalué en préalable à la formation.

Objectifs

- Être capable de décrire une série de valeurs quantitative ou qualitative.
- Savoir identifier correctement les problèmes que les statistiques peuvent résoudre.
- Savoir choisir et mettre en pratique le test statistique adapté à chaque situation.
- Savoir mettre en pratique des tests de comparaisons de 1 ou 2 échantillons
- Savoir estimer et tester les relations entre variables
- Savoir interpréter les sorties logicielles correspondantes.

Méthode

Dans chaque rubrique on donne d'abord les principes de calcul, sans recours à des notions mathématiques complexes, en privilégiant une compréhension intuitive. On indique les limites des techniques et les conditions nécessaires pour les appliquer. On les met en pratique en utilisant les fonctions statistiques du logiciel Excel (un ordinateur par stagiaire). On décrit et on interprète les sorties logicielles. On donne des illustrations des usages que l'expérimentateur peut faire de chaque technique. Les situations concrètes rencontrées par les stagiaires sont examinées et discutées.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

Dates : nous consulter

Intra-entreprise

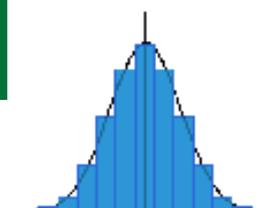
groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Un tour d'horizon des notions de base indispensables pour aborder, comprendre et utiliser différentes méthodes statistiques.



Programme

Fonctions des statistiques

Description (statistiques descriptives)

Aide à la décision (statistiques décisionnelles)

Ce que disent... et ne disent pas les statistiques

Caractérisation d'un échantillon

Caractérisation de la tendance centrale

Caractérisation de la dispersion

Représentations de la dispersion

Le cas des données qualitatives

Représentation, calcul de l'intervalle de confiance d'une proportion, mesure de la tendance centrale

Tous ces aspects seront étudiés dans le cas général et dans le cas des petits échantillons

Tests statistiques de comparaisons

Différentes échelles de mesure et tests correspondants

Formulation statistique de l'hypothèse du biologiste

Notion de hasard en statistiques, problématique des tests, réflexion sur la relation entre taille d'échantillon et risque dans les comparaisons de moyennes

Choix d'un seuil de risque

Notion d'« hypothèse nulle » (notée H_0)

Risques de première et de seconde espèce

Hypothèses bilatérales et unilatérales

Echantillons indépendants et séries de mesures appariées

Puissance d'un test

Choix des tests et exercices

Comparaison d'une moyenne observée à une moyenne théorique

Comparaison de deux séries paires

Comparaison de deux échantillons indépendants

Conditions de validité du test t de Student

Normalité des distributions

Homogénéité des variances des deux groupes

Indépendance entre les erreurs et les effets des traitements

Principes de l'analyse de variance et des comparaisons multiples (tests "post hoc")

Principe des tests non paramétriques pour le traitement statistique des petits échantillons et des données ordinales

Rappels sur les particularités des petits échantillons

Les tests non paramétriques: avantages et inconvénients

Mesure de la liaison entre variables quantitatives

Liaison entre 2 variables quantitatives: la corrélation

Introduire une causalité: la régression linéaire simple

Mesure de la relation entre variables qualitatives : test du Chi² et test binomial

2. Introduction au logiciel R pour les statistiques

Public concerné

Techniciens, ingénieurs ou chercheurs confrontés à l'exploitation statistique des données.

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; connaître les différences entre test paramétrique et test non-paramétrique ; savoir ce qu'est la puissance d'un test. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Savoir décrire des lignes de commandes simples
- Savoir exécuter des lignes de commandes simples
- Savoir utiliser l'interface graphique R Commander
- Savoir importer des fichiers externes à R (.txt, .xlsx) via R Commander
- Être capable de mettre en pratique des manipulations simples de données via R Commander
- Être capable de créer des graphiques simples via R Commander
- Être capable de décrire des données via R Commander
- Être capable de mettre en pratique des tests statistiques classiques (t de Student, Shapiro-Wilk, ...) via R Commander

Méthode

Après une présentation rapide de l'interface et de quelques notions élémentaires du langage S, l'enseignement s'appuie sur l'interface graphique "R commander". L'utilisation de cette interface permet aux stagiaires d'être rapidement opérationnels, tout en assimilant très progressivement les possibilités offertes par les commandes par scripts.

Les stagiaires réalisent tous les exemples eux-mêmes et des exercices pratiques à réaliser seuls leur sont ensuite proposés.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Le logiciel R, logiciel libre fonctionnant sur toute plateforme, est un outil très puissant, tant pour les tests statistiques, l'exploration des données que pour la réalisation de graphiques de qualité. Outil internationalement reconnu, il est maintenant très facile d'abord grâce



Programme

Présentation du « Projet R »

Concepts de base du fonctionnement de R

Utilisation de l'aide et de la documentation disponible sur la toile

Les bases du langage S. Calculs simples depuis la console de R

Gestion des objets en mémoire : manipuler les données et les enregistrer

L'interface "Rcmdr"

Importation de fichiers

Gestion des données

Editer et modifier les données, renommer des variables, éliminer des lignes ou des colonnes, empiler des variables, gérer les données manquantes, recoder des variables, découper une variable numérique en classes

Introduction aux graphiques avec R

Principes généraux, fonctions graphiques principales et secondaires, les paramètres graphiques et leur enregistrement, utilisation des symboles, les graphiques multiples conditionnés. L'interface GraphR.

Statistiques descriptives

Matrices de corrélations et tests de corrélation

Tests non paramétriques

Tests sur les rangs : Mann & Whitney, Wilcoxon, Kruskal & Wallis, Friedman, comparaisons multiples

Tests paramétriques

Test de normalité, tests de l'homogénéité des variances, puissance et effectifs nécessaires

Test t de Student et test de Welch

Analyses de variance à une dimension; à deux dimensions; à mesures répétées: construction du modèle, vérifications, comparaisons multiples

Régressions linéaires

Représentations graphiques

Construction et utilisation du modèle linéaire pour la régression linéaire simple

Vérification des résidus: normalité, hétéroscédasticité, autocorrélation

3. Le traitement statistique des petits échantillons

Public concerné

Techniciens, ingénieurs ou chercheurs confrontés au traitement statistique de petits échantillons.

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; connaître les différences entre test paramétrique et test non-paramétrique. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Être capable de décrire une petite série de données
- Savoir choisir un test adapté aux questions que l'on se pose et aux données dont on dispose
- Savoir comparer deux ou plusieurs échantillons
- Savoir estimer et tester l'association entre plusieurs variables.
- Savoir mettre en pratique ces tests et interpréter les sorties logicielles.

Méthode

Pour chaque question étudiée, l'enseignement est divisé en quatre parties :

1. Principe des tests disponibles. Choix du test correspondant à la situation.
2. Méthode de calcul (manuel ou informatisé).
3. Puissance et conditions de validité.
4. Solutions logicielles, mise en œuvre avec le logiciel StatXact, examen et interprétation des sorties.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Les développements de l'informatique modifient grandement le paysage des statistiques. Des techniques récentes permettent d'aborder facilement et sûrement l'analyse des petits échantillons.



Programme

Présentation théorique

Rappels sur les fonctions des statistiques

Premières réflexions sur la taille de l'échantillon

Représentation des paramètres d'un échantillon

Hypothèses et risques

Problématique d'un test statistique

Directionnalité de l'hypothèse

Risque et taille de l'échantillon

Risques de première et de deuxième espèce

Puissance d'un test

Trois familles de tests

Choix d'un test statistique

Les différentes échelles de mesure

Echantillons indépendants et séries de mesures appariées

Paramétrique ou non paramétrique ?

Grille de choix d'un test non paramétrique

Exercices de choix d'un test

Les tests « exacts » ou tests par permutations

Solutions logicielles pour les tests non paramétriques

Réalisations pratiques

Présentation des outils utilisés

Comparaison d'échantillons

Mesures en échelle d'intervalle : séries appariées et indépendantes, données

censurées, échantillons stratifiés, comparaisons multiples

Représentations graphiques

Résolution d'exercices en utilisant des tests sur les rangs et des tests exacts

Interprétation des résultats

Les mesures sont des catégories ordinales : échantillons indépendants et des séries appariées

Représentations graphiques

Résolution d'exercices en utilisant des tests sur les rangs et des tests exacts

Interprétation des résultats

Les mesures sont des catégories nominales

Réalisation et interprétation du test du χ^2 de bonne adéquation, du test du χ^2 de contin-

gence, du test du χ^2 exact, du test binomial, du test de McNemar, du test exact de Fisher, du

test de Cochran

Mesure de la relation entre variables

Les données sont en échelle au moins ordinale : corrélation et concordance

Réalisation du test de corrélation par rangs de Spearman ainsi que de son calcul exact

Calcul du coefficient de Kendall et du test de concordance de Kendall

Les données sont en échelle nominale : coefficients de contingence, coefficient Kappa de Cohen

4. Pratique et interprétation des analyses de variance

Public concerné

Toute personne ayant à comparer des échantillons, analyser des résultats d'expériences, mais ne possédant pas une bonne maîtrise des méthodes d'analyse de variance. Une connaissance des bases du logiciel R ou XLSTAT (selon le logiciel choisi pour réaliser la formation) est très souhaitable.

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; connaître les différences entre test paramétrique et test non-paramétrique ; savoir ce qu'est la puissance d'un test. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Savoir situer l'anova et calculer une variance
- Savoir réaliser une anova à une dimension dans R Commander
- Savoir vérifier que les conditions d'application sont remplies
- Savoir réaliser les comparaisons par paires à l'issue de l'analyse globale
- Savoir reconnaître lorsqu'une expérimentation comporte plus d'une dimension étudiée
- Être capable de distinguer un facteur indépendant d'un facteur apparié
- Comprendre l'intérêt de prendre en compte plus d'une dimension
- Être capable d'interpréter les probabilités associées aux facteurs et à l'interaction entre eux

Méthode

Les principes de calculs sont exposés sans recours à des notions mathématiques complexes. On alterne exposés théoriques et mise en pratique des notions abordées à l'aide du logiciel XLSTAT et des outils du tableur Excel. A la demande, la formation peut être réalisée avec le logiciel R. Les stagiaires réalisent eux-mêmes des applications nombreuses à partir de situations expérimentales très diverses (un ordinateur par stagiaire). Les sorties logicielles sont examinées et interprétées en s'aidant de représentations graphiques.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise

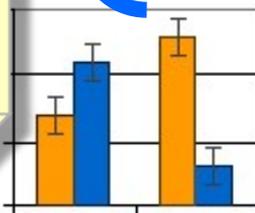
groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Outil incontournable pour l'analyse des données issues de plans d'expériences, l'analyse de variance fait partie des tests statistiques les plus utilisés mais elle n'en reste pas moins complexe si l'on s'intéresse à ses conditions d'utilisation et à ses diverses variantes



Programme

L'analyse de variance à une dimension

Echantillons indépendants

Objectif et hypothèses

Principe général

Calculs et interprétations

Conditions de validité, transformations d'échelles, puissance

Solutions non paramétriques

Comparaisons multiples

Intérêts de prendre en compte plus d'une dimension

Analyses de variance à plusieurs dimensions

Dans le cas de mesures indépendantes seules, de séries appariées et du croisement des deux types, de plans en blocs:

Objectifs

Principes et calculs

Conditions d'application

Comparaisons multiples

Solutions non paramétriques

Le modèle linéaire et sa généralisation

Modèle linéaire

La régression linéaire simple

Analyse de variance et régression linéaire

Analyses multivariées de la variance

L'analyse de covariance

Généralisation du modèle linéaire

5. Construction et analyse des plans d'expériences

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants confrontés à la mise en place de protocoles expérimentaux, à l'analyse et à l'interprétation des résultats qui en découlent. Une bonne familiarisation au logiciel R est souhaitable.

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU avoir suivi la formation Pratique et interprétation des analyses de variance OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; connaître les différences entre test paramétrique et test non-paramétrique ; savoir ce qu'est la puissance d'un test. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Comprendre le principe de la limitation du nombre d'essais et de contrôle des variations parasites dans le cadre d'un protocole expérimental et de ses contraintes
- Savoir identifier une source de variation résiduelle versus un facteur fixe
- Comprendre ce qu'est le screening et savoir le mettre en place
- Connaître le principe d'un plan en blocs complets
- Savoir avantage et limitation du plan en carré latin, et du plan en carrés gréco-latins
- Savoir faire un plan à surface de réponse
- Savoir trier de nouveaux produits
- Être capable de lire les sorties et graphiques
- Savoir choisir le type de plan adapté à une problématique, aux données

Méthode

L'enseignement repose sur une alternance constante entre des explications théoriques sans recours à des notions mathématiques complexes et de nombreuses applications pratiques. Chaque méthode est illustrée par un exemple en vraie grandeur, reposant sur le domaine de recherche des stagiaires.

Les stagiaires exécutent à l'ordinateur avec le logiciel R tous les exemples proposés et réalisent ensuite de nombreuses applications en autonomie.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise

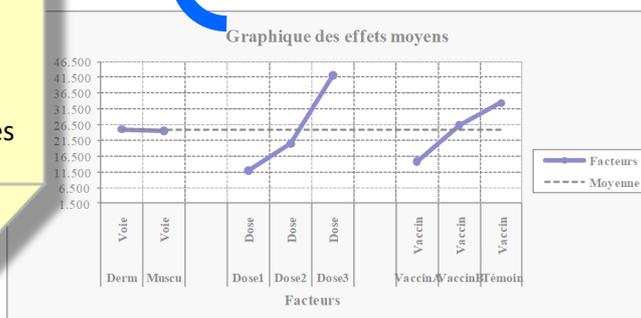
groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

La mise en place d'un plan d'expériences est un outil précieux voire indispensable pour l'optimisation d'un protocole expérimental.



Programme

Rappel des notions statistiques de base

Variation contrôlée et non contrôlée

Quelques définitions : notions de variable, population, échantillon, individu, échantillons indépendants et mesures appariées

Exigences méthodologiques : randomisation, biais expérimentaux

Plans factoriels complets et leur analyse

Définitions.

Ecriture conventionnelle.

Calcul du nombre d'expériences nécessaires dans un plan complet.

Analyse statistique et interprétation

Les interactions : définition, calcul de l'interaction dans un plan à deux facteurs, représentation graphique, calculs avec plus de deux facteurs, calculs et interprétation dans le cas général

Questions de méthode

Hypothèses et contraintes statistiques

Choix du plan adapté à une problématique

Recherche d'effets de facteurs

Le criblage (ou screening)

principe, création, analyse des résultats

Plans fractionnaires

principe, création, analyse des résultats

Recherche d'une réponse optimale

Plans d'expériences pour la réduction de la variation résiduelle

Principe

Plan en blocs complets : principe, création, analyse des résultats

Plan en carré latin : principe, création, analyse des résultats

Plan en carrés gréco-latins : principe, création, analyse des résultats

Plans en blocs incomplets : principe, création, analyse des résultats

Plans à surface de réponse

Principe

Les plans composites centrés : principe, création, analyse des résultats

Tri de nouveaux produits

Plans en blocs augmentés

Principe, création, analyse des résultats

6. Décrire et analyser les données multifactorielles

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants ayant à décrire des grands tableaux de données quantitatives et/ou qualitatives.

Prérequis

Aucun prérequis nécessaire.

Objectifs

- Savoir définir un facteur qualitatif versus quantitatif
- Savoir préparer son tableau (recodage, données manquantes, etc.)
- Connaître la taille/configuration des tableaux propres à chaque analyse
- Savoir comment lire une matrice de corrélations multiples, et calculer un Chi-Deux
- Savoir réaliser une ACP dans FactoMineR, et lire les plans
- Savoir vérifier la robustesse des analyses à partir des calculs associés
- Savoir quand utiliser une classification et comment elle se construit
- Être capable de comprendre l'intérêt d'une AFM versus ACM
- Savoir associer son analyse factorielle à une classification avec FactoMineR
- Être capable d'interpréter les sorties logicielles des analyses

Méthode

Pour chaque question étudiée, l'enseignement comporte une explication de la méthode sans recours à des notions mathématiques. Pour chaque technique on traite un exemple en vraie grandeur à l'aide du logiciel Spad. Dans tous les cas on examine les sorties logicielles et on insiste sur les règles et les difficultés d'interprétation des résultats. Des aides à l'interprétation sont systématiquement mises en pratique. Les stagiaires réalisent eux-mêmes des analyses en autonomie.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

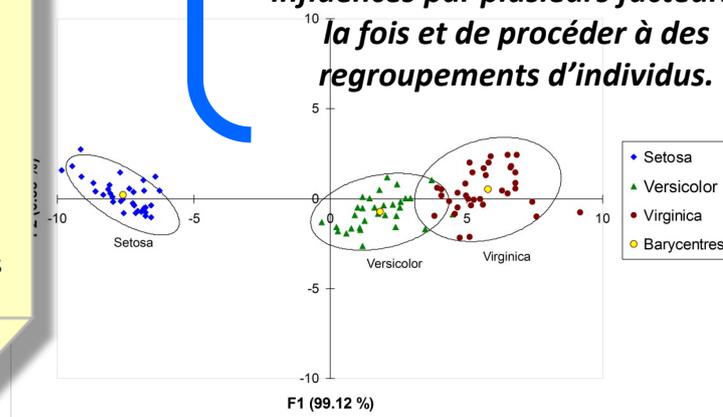
groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Les statistiques multifactorielles permettent d'explorer de grands tableaux, d'analyser les processus influencés par plusieurs facteurs à la fois et de procéder à des regroupements d'individus.



Programme

Introduction à l'analyse multifactorielle des données

Problématique générale des approches descriptives multifactorielles
Les différents types de mesures et les méthodes correspondantes

Données quantitatives : l'analyse en composantes principales

Exemples variés d'utilisations
Principes généraux à partir d'un exemple simple
Variables actives et illustratives
Réalisation d'une analyse en vraie grandeur
Aides à l'interprétation des résultats d'une ACP
Vérification des conditions de validité et transformations éventuelles des variables

Données qualitatives : l'analyse factorielle des correspondances

L'analyse des correspondances simples
Principe et réalisation
Lecture du graphe factoriel
Validation et aides à l'interprétation
Pratique sur des données en vraie grandeur
L'analyse des correspondances multiples
Objectifs et principe
Mise en œuvre et interprétation des sorties logicielles
Validation et aides à l'interprétation

Quelques autres méthodes

L'analyse factorielle multiple
L'analyse factorielle de données mixtes
L'analyse canonique des correspondances

Les méthodes de classification ou *clustering*

Méthode d'agrégation
Principe de la classification ascendante hiérarchique
Le critère de WARD
La lecture d'un dendrogramme
Relations entre les classes d'individus et les variables
Partition de l'arbre et description des classes formées
Validité d'une classification
Classification par la méthode des centres mobiles, principe et outils
La classification descendante hiérarchique, principe et outils

7. Machine learning

Public concerné

Techniciens, ingénieurs ou chercheurs souhaitant prendre des décisions à partir de grands tableaux de données. Des compétences en mathématiques ne sont pas nécessaires. Une utilisation régulière du logiciel R est très souhaitable.

Prérequis

Avoir suivi la formation Pratique et interprétation des analyses de variance et/ou Introduction aux modèles mixtes OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; savoir réaliser des modèles statistiques courants (régression linéaire, ANOVA) et connaître les conditions de validité. Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires et exécuter des lignes de codes. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Savoir situer les méthodes de régressions et d'analyses discriminantes dans le machine learning
- Savoir différencier les méthodes de régression des méthodes d'analyse discriminante
- Être capable de construire un modèle, de vérifier sa qualité et d'interpréter son contenu
- Être capable de choisir le meilleur modèle parmi les modèles construits
- Savoir mettre en pratique ces méthodes avec le logiciel R

Méthode

L'enseignement s'appuie sur l'utilisation du logiciel XLSTAT. Pour chaque question étudiée on donne une explication sans recours à des notions mathématiques complexes. Pour chaque technique, les stagiaires traitent un exemple en vraie grandeur, éventuellement sur leurs données. On examine les sorties logicielles en insistant sur les difficultés d'interprétation.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

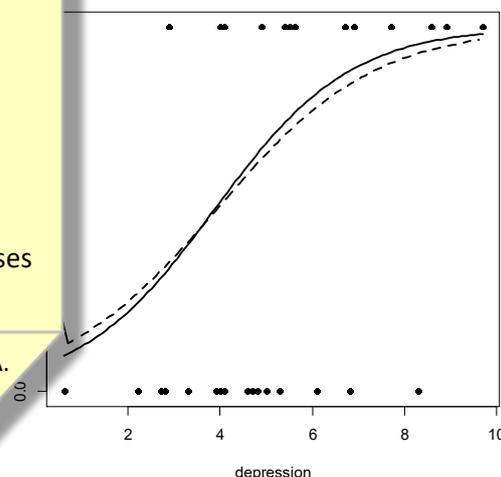
Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Modéliser des phénomènes et faire des prédictions à partir de tableaux de données complexes. Toute une panoplie d'outils à votre service.

régression logistique



Programme

Les outils de la fouille des données

Apprentissage statistique et Machine Learning.

Le surapprentissage et comment l'éviter.

La bibliothèque caret.

Les régressions et leurs conditions de validité

Présentation et description des données

La régression linéaire multiple

Conditions de validité. Mise en œuvre. Régression pas à pas.

La régression PLS

Intérêts ; Interprétation des sorties logicielles et des graphiques factoriels

Interprétation des résultats de la régression

Classement des variables

Estimation de la qualité du modèle.

Comparaisons de modèles et bilan

Exercices en autonomie

Les analyses discriminantes

Généralités, matrice de confusion et courbe ROC.

Importance des variables, prévision de nouveaux cas.

Présentation et description des données

L'analyse discriminante linéaire

La régression logistique

L'analyse discriminante sur facteurs, PLS-DA

Les forêts aléatoires (RandomForest)

Les réseaux de neurones

Les méthodes à vecteurs support (SVM)

Aspects théoriques. Conditions de validité. Mise en œuvre.

Comparaisons de modèles et bilan

Exercices en autonomie

8. Introduction aux modèles mixtes linéaires et non linéaires

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants confrontés à l'analyse approfondie de données complexes linéaires et non linéaires et pouvant comporter des emboîtements et/ou des données dépendantes. Une utilisation régulière du logiciel R est souhaitable.

Prérequis

Avoir suivi la formation Pratique et interprétation des analyses de variance OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; savoir réaliser des modèles statistiques courants (régression linéaire, ANOVA à effets fixes) et connaître les conditions de validité. Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires et exécuter des lignes de codes. Un questionnaire préalable permettra l'évaluation.

Objectifs

- Savoir choisir entre modèles linéaires et modèles non linéaires
- Savoir identifier un effet fixe
- Savoir identifier un effet aléatoire
- Savoir identifier l'agencement des effets (croisés, hiérarchisés, ...)
- Savoir choisir et mettre en oeuvre un modèle avec le logiciel R
- Savoir sélectionner le meilleur modèle
- Savoir vérifier la qualité du modèle choisi
- Savoir interpréter les sorties logicielles

Méthode

Des rappels seront d'abord faits sur les modèles linéaires fixes avant d'aborder les modèles linéaires mixtes puis les modèles non linéaires (effets fixes ou mixtes).

La théorie statistique est abordée si nécessaire et sans recours à des notions mathématiques complexes.

Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

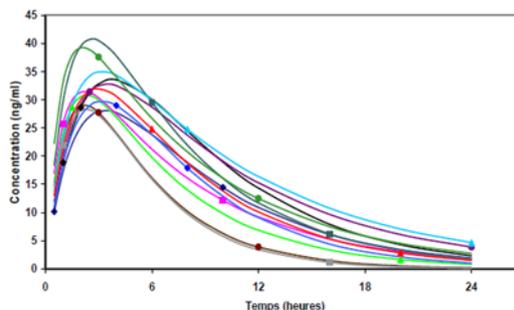
Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises

4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Savoir utiliser des outils permettant de construire des modèles performants à partir de données complexes.



Programme

Généralités

Variation contrôlée et non contrôlée
Définition et objectifs de la modélisation
Différentes familles de modèles
Stratégie en modélisation

Les modèles linéaires à effets fixes

Rappels sur le modèle linéaire simple : ANOVA et régression linéaire
Modèle linéaire généralisé

Les modèles linéaires mixtes

Présentation théorique
Analyse des plans en blocs complets
Modèles avec effets aléatoires croisés
Modèles avec effets hiérarchisés
Analyser les données longitudinales (mesures répétées)
Analyse paramétrique classique
Analyse non paramétrique
Introduction aux modèles linéaires généralisés (GLM) mixte avec l'exemple de la régression logistique mixte
Comparaisons de modèles

Modélisation non linéaire à effets fixes

Principe général et différents modes d'approche
Choix d'une fonction non linéaire adaptée aux données, commandes et interprétations
Comparaisons de modèles

Modélisation non linéaire à effets mixtes

Principe général
Choix d'une fonction, commandes et interprétations
Comparaisons de modèles

Annexes

Grille de choix d'un modèle
Fonctions pour l'analyse paramétrique des données longitudinales
Fonctions non linéaires courantes
Syntaxe des arguments 'random'

9. Modélisation paramétrique, niveau avancé

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants.

Prérequis

Bonnes bases en statistiques. Maîtrise des analyses de variances et des régressions linéaires. Avoir déjà utilisé des modèles mixtes. Bonne maîtrise du logiciel R. Un questionnaire préalable permettra l'évaluation.

Objectifs

- Savoir choisir entre modèles linéaires et modèles non linéaires
- Savoir identifier un effet fixe et un effet aléatoire
- Savoir identifier l'agencement des effets (croisés, hiérarchisés, ...)
- Savoir sélectionner le meilleur modèle
- Savoir vérifier la qualité du modèle choisi : identifier les écarts aux hypothèses
- Savoir choisir une structure de variance
- Savoir choisir une structure de variances-covariances
- Savoir gérer les données surdispersées, avec beaucoup de 0
- Savoir mettre en oeuvre un modèle avec le logiciel R et interpréter les sorties logicielles

Méthode

Des rappels seront d'abord faits sur la stratégie globale de modélisation. Les écarts aux hypothèses seront présentés et illustrés. L'utilisation d'effets aléatoires dans les modèles sera ensuite discutée. Différents cas de modèles mixtes présentant éventuellement des écarts aux hypothèses seront développés.



Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

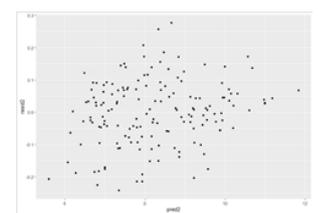
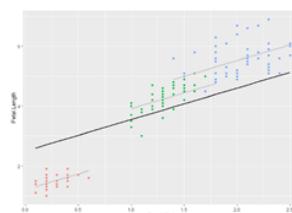
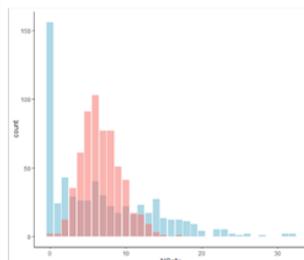
groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises

4926 € intra-entreprise

Nos formations sont exonérées de TVA.



Programme

Introduction

Modélisation statistique : définition et objectifs

Variation contrôlée et variation non contrôlée

Méthodes de modélisation

Démarche globale de modélisation

Représentations graphiques : exploration des données

Choix du meilleur modèle

Validation du modèle : vérifier les hypothèses

Homoscédasticité des résidus : utiliser une structure de variance

Le cas des variables possédant beaucoup de valeurs nulles

Les modèles mixtes

Prendre en compte des effets aléatoires dans un modèle

Pourquoi ? Dans quelles situations ?

Trois types de modèles à effets aléatoires

Les différentes structures de covariances

Les messages d'erreur courants : signification et réglages éventuels

Modèle linéaire à effets mixtes avec 2 effets aléatoires croisés

Exploration des données

Choix du meilleur modèle

Vérification du modèle choisi et analyse des résultats

Modèle linéaire avec mesures répétées

Observation des données

Choix du meilleur modèle

Vérification du modèle choisi et analyse des résultats

Régression de Poisson à effets mixtes

Exploration des données

Choix du meilleur modèle

Vérification du modèle choisi et analyse des résultats

Régression non linéaire à effets mixtes

Particularités des modèles non linéaires

Exploration des données

Choix du meilleur modèle

Vérification du modèle choisi et analyse des résultats

10. Analyse statistique des données textuelles avec IRaMuTeQ

Public concerné

Techniciens, ingénieurs ou chercheurs confrontés à l'exploitation de données textuelles. La connaissance des méthodes descriptives multifactorielles (AFC en particulier) est souhaitable.

Prérequis

Savoir interpréter le résultat d'un test Chi². Un questionnaire préalable permettra leur évaluation.

Objectifs

- Se familiariser avec le vocabulaire spécifique aux analyses lexicales
- Comprendre l'intérêt et l'utilisation de segments de texte
- Maîtriser la fonction de la lemmatisation et savoir la paramétrer
- Réaliser l'analyse des spécificités. Comprendre et en interpréter les sorties logicielles
- Comprendre le principe de la classification descendante et savoir interpréter ses résultats
- Comprendre le principe, réaliser et interpréter l'analyse des similitudes
- Savoir construire un nuage de mots et en comprendre la fonction, les limites et l'utilisation
- Évaluer l'intérêt, réaliser et interpréter une analyse prototypique

Méthode

Après une courte introduction les stagiaires apprennent à utiliser l'interface du logiciel libre Iramuteq (travaillant à partir des bibliothèques du logiciel R). La méthode Reinert de classification et une approche très fine des contextes de mots par les analyses de similitudes sont expliquées.

A partir d'un exemple d'enquête sociologique comportant des questions fermées et ouvertes, ils réalisent en pratique l'ensemble de l'analyse.

Les éléments théoriques sont présentés sans recours à des connaissances particulières en mathématiques.

Durée : Deux jours

Interentreprises (Tours)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1192 €/personne
interentreprises
3448 € intra-entreprise

Nos formations sont exonérées de TVA.

Des logiciels spécialisés permettent aujourd'hui l'analyse automatique des langues naturelles. Les méthodes multivariées permettent une exploitation poussée de ces données, questions ouvertes dans les enquêtes, entretiens, veille technologique .

 IRaMuTeQ



Programme

Généralités

Présentation du logiciel IRaMuTeQ

Présentation générale

Principes de fonctionnement

Paramétrage du texte et importation

Lemmatisation et statistiques élémentaires

Lemmatisation : Définitions, Mise en œuvre

Statistiques élémentaires

Spécificités et AFC

Rappels sur l'analyse des correspondances : Fonction de l'AFC, Rappel sur le χ^2 , Les calculs de l'AFC, Lecture du graphe factoriel d'une AFC

Classification méthode Reinert

Principe de la classification descendante hiérarchique sur segments de texte

Mise en œuvre de la classification : *Paramétrage de la classification, Lecture et interprétation*

Analyses de similitudes

Principe

Mise en œuvre des analyses de similitudes : *Paramétrage de l'analyse, Lecture et interprétation, Autres graphes de similitudes*

Nuages de mots

Généralités

Mise en œuvre : *Paramétrage, Lecture*

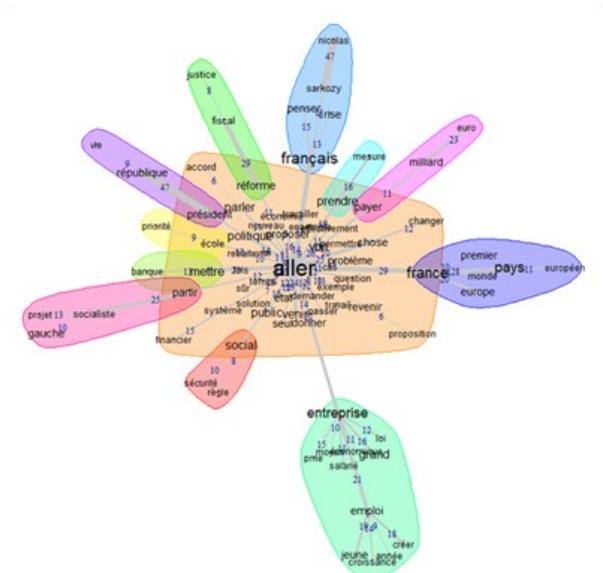
Analyse de matrices et analyse prototypique

Généralités

Mise en œuvre : *Fréquences, Fréquences multiples, χ^2 , Classification, Analyses de similitude, Analyse prototypique*

Limites de l'analyse automatique du texte

Exercices en autonomie



11. Utilisation avancée du logiciel R pour les statistiques

Public concerné

Techniciens, ingénieurs ou chercheurs désirant apprendre à créer des scripts complexes et à automatiser leur traitement statistique. Il ne s'agit pas ici de la création de packages.

Prérequis

Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires ; savoir modifier des lignes de codes ou savoir écrire ses propres lignes de codes. Un questionnaire permettra leur évaluation.

Objectifs

- Savoir créer des rapports automatisés
- Savoir manipuler efficacement des jeux de données
- Savoir mettre en pratique des conditions
- Savoir mettre en pratique des boucles
- Savoir mettre en pratique des fonctions personnelles
- Savoir utiliser les paramètres graphiques
- Être capable d'appliquer les bonnes pratiques de programmation

Méthode

Chaque sous-paragraphe est accompagné de plusieurs applications pratiques. Au moins une journée sera consacrée à des exercices en autonomie, axés principalement sur l'écriture de scripts contenant des conditions et des boucles plus ou moins complexes.

Ce module peut à la demande être entièrement réalisé sans pratique des tests statistiques

Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

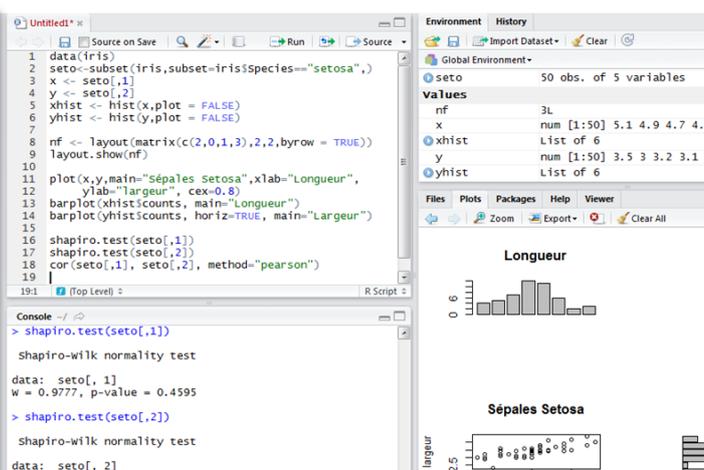
Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.



```
1 data(iris)
2 seto<-subset(iris,subset=iris$Species=="setosa",)
3 x <- seto[,1]
4 y <- seto[,2]
5 xhist <- hist(x,plot = FALSE)
6 yhist <- hist(y,plot = FALSE)
7
8 nf <- layout(matrix(c(2,0,1,3),2,2,byrow = TRUE))
9 layout.show(nf)
10
11 plot(x,y,main="Sépales Setosa",xlab="Longueur",
12      ylab="largeur", cex=0.6)
13 barplot(xhist$counts, main="Longueur")
14 barplot(yhist$counts, horiz=TRUE, main="Largeur")
15
16 shapiro.test(seto[,1])
17 shapiro.test(seto[,2])
18 cor(seto[,1], seto[,2], method="pearson")
19 |
```

Environment History

Global Environment

seto 50 obs. of 5 variables

Values

mf 3L

x num [1:50] 5.1 4.9 4.7 4.6

xhist List of 6

y num [1:50] 3.5 3.3 3.2 3.1 3

yhist List of 6

Files Plots Packages Help Viewer

Zoom Export Clear All

Longueur

Sépales Setosa

largeur

Console

```
> shapiro.test(seto[,1])
Shapiro-wilk normality test
data: seto[, 1]
W = 0.9777, p-value = 0.4595
> shapiro.test(seto[,2])
Shapiro-wilk normality test
data: seto[, 2]
```



Programme

Les éditeurs

Un choix... personnel

R-Studio : présentation et prise en mains

Produire des documents à partir des sorties R

Principe du R Markdown

Créer un document HTML, Word ou pdf avec Rstudio

La syntaxe de base de R Markdown

Introduction à la programmation sous R

Les fonctions prédéfinies

Les opérateurs

Les fonctions de manipulation

Les fonctions mathématiques et statistiques

Les boucles préprogrammées

Les structures classiques de programmation

Les commandes groupées

Les conditions

Les boucles

La structure des fonctions

Les fonctions prédéfinies

Créer ses propres fonctions

Débogage d'une fonction

Conseils pour l'écriture des scripts

Optimiser les graphiques sous R

Paramétrer et enrichir les graphiques simples

Fenêtres graphiques

Fonctions graphiques et paramètres

Fonctions graphiques courantes

Utiliser {ggplot2}

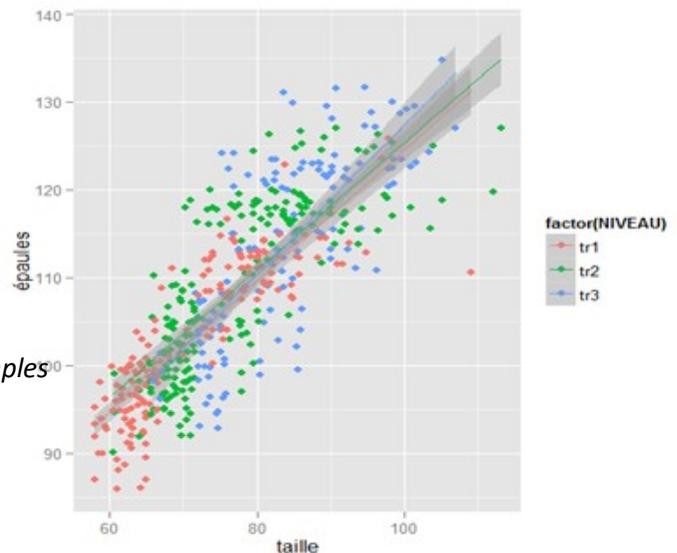
Principes généraux

Utiliser qplot()

Utiliser ggplot()

Améliorations

Utiliser {ggpubr}



Si le logiciel R est riche de milliers de fonctions, il permet aussi d'automatiser nombre de calculs en écrivant soi-même des commandes adaptées à ses données et aux représentations graphiques souhaitées. Une utilisation en routine de ce logiciel permet ainsi un gain de temps non négligeable dans le traitement statistique et dans la production de rapports clairs et de qualité.

12. Introduction à l'analyse de données à haut débit avec le logiciel R

Public concerné

Chercheurs, ingénieurs, techniciens, doctorants ayant à traiter des données issues de spectrométrie, de protéomique, de métabolomique ou génomique (les omiques).

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue ; connaître les différences entre test paramétrique et test non-paramétrique. Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires et exécuter des lignes de codes. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Savoir vérifier le jeu de données (données manquantes, distributions, qualité ...)
- Être capable mettre en pratique les méthodes de filtrage des données
- Savoir utiliser des méthodes de réduction de dimensions et en extraire de l'information
- Savoir mettre en pratique les tests statistiques inférentiels, adaptés à la taille du jeu de données

Méthode

Après une présentation générale des défis particuliers associés à ce type de données, on présente les méthodes descriptives puis décisionnelles, sans recours à des notions mathématiques complexes. Tout au long de la formation, les stagiaires mettent les acquis en pratique – en autonomie – sur données réelles. Une rapide introduction au logiciel R est proposée si nécessaire.

Durée : Trois jours

Interentreprises (Tours ou Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.



Les biologistes sont maintenant très nombreux à produire des données à haut débit. Des milliers de variables avec des petits effectifs posent des problèmes spécifiques mais les outils sont aujourd'hui disponibles pour répondre à ces défis.

Programme

Introduction

Données "OMIQUES", qu'est-ce que c'est ?
Données brutes et matrices d'expression
Problématiques et défis particuliers
Logiciels

Brève introduction au logiciel R

Pratique sur données réelles : manipulations préliminaires

Chargement et résumé
Quelques graphiques simples
Travailler sur les log de quantités
Transformation d'un tableau de données en matrice d'expression
Contrôle de qualité des données
Filtrage non spécifique, Filtrage spécifique

Statistiques descriptives exploratoires

L'analyse en composantes principales (ACP)
Classification ascendante hiérarchique (CAH)
Heatmap

Statistiques décisionnelles

Tests usuels paramétriques et non paramétriques
Tri des variables
Inflation du risque alpha et comparaisons multiples
Pratique : correction des p-values
Vérifications par rééchantillonnage

Pratique en autonomie sur des jeux de données réelles.

13. Les statistiques bayésiennes : introduction et pratique

Public concerné

Techniciens, ingénieurs ou chercheurs désirant

- 1) comprendre de manière intuitive le principe des statistiques bayésiennes et les intérêts qu'elles présentent par rapport aux statistiques conventionnelles ,
- 2) mettre en application les méthodes bayésiennes pour répondre à des problématiques plus ou moins complexes.

Prérequis

Avoir suivi la formation Notions fondamentales en statistiques OU savoir poser les hypothèses d'un test statistique et interpréter la p-value obtenue. Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires ; savoir modifier des lignes de codes ou savoir écrire ses propres lignes de codes. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Être capable d'expliquer, de manière intuitive, le raisonnement et le calcul mis en jeu en statistiques bayésiennes (sans recours à des formules mathématiques complexes)
- Être capable d'énoncer les différences avec les statistiques inférentielles
- Être capable de construire ses connaissances *a priori*
- Être capable de mettre en pratique des modèles classiques (régression, ANOVA, etc.) dans un univers bayésien
- Être capable d'interpréter les résultats avec l'inférence bayésienne

Méthode

Après une introduction théorique intuitive, ludique et presque dépourvue de formules mathématiques complexes, des modèles utilisés couramment en statistiques conventionnelles (régression, ANOVA, modèle mixte, etc.) sont mis en application dans un univers bayésien grâce aux logiciels R et JAGS. Des fichiers de lignes de code sont fournis avec explications détaillées, de manière à ce que les participants puissent les adapter à leurs propres problématiques à l'issue de la formation.

Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

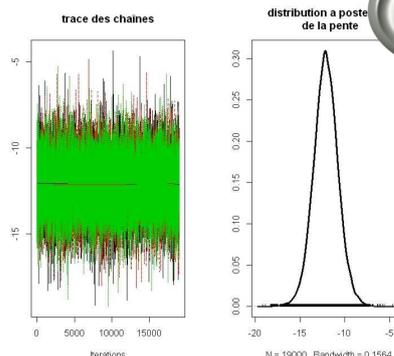
Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.



Programme

Théorie

Préambule et historique

Quelques problèmes rencontrés en statistiques classiques, apport des statistiques bayésiennes

Quand utiliser les statistiques classiques et quand utiliser les statistiques bayésiennes ?

Forme conceptuelle et intuitive du calcul bayésien

Forme mathématique du théorème de Bayes

Description d'une population : vision fréquentiste versus vision bayésienne

Limites du calcul analytique et utilité des algorithmes MCMC (Markov Chain Monte Carlo)

Illustration de la marche aléatoire de Metropolis

Construire sa connaissance a priori

Trois objectifs de l'inférence revus sous un angle bayésien : estimation, prédiction, comparaison

Applications JAGS (Just Another Gibbs Sampler)

Présentation/ installation

JAGS – quelques notes sur la syntaxe

Démarche systématique pour cette formation

Calcul de la densité d'espèces dans une pelouse calcaire

Régression linéaire simple

Analyse de Variance à un facteur

Analyse de Variance à deux facteurs

Modèle hiérarchique

Imputation de données manquantes

Régression logistique multiple

Les statistiques bayésiennes sont une manière différente de construire la connaissance par rapport aux statistiques conventionnelles : elles offrent la possibilité de mettre à jour ce que nous savons déjà à la lumière de ce que nous apportent nos expérimentations et observations, le tout dans un calcul très intuitif. Elles peuvent résoudre un grand nombre de problèmes rencontrés en statistiques classiques (hétérosedasticité, multicollinéarité, petits échantillons, données manquantes, modèles très complexes...). On s'attend à ce que ces méthodes remplacent les outils classiques d'inférence dans les années à venir.

14. Introduction à l'analyse et à la modélisation des séries temporelles avec le logiciel R

Public concerné

Toute personne, étudiant, doctorant, ingénieur, enseignant ou chercheur ayant recours à des séries chronologiques et désireuse de les analyser sous R.

Prérequis

Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires et exécuter des lignes de codes. Un questionnaire préalable permettra une évaluation.

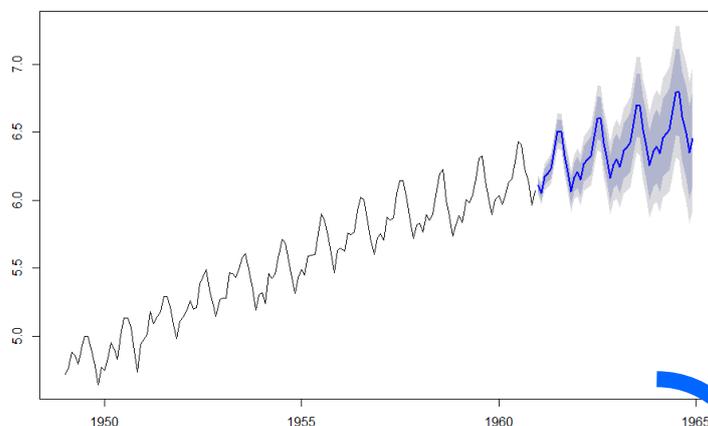
Objectifs

- Savoir identifier et construire des séries temporelles dans R
- Savoir construire les graphiques adaptés pour visualiser des séries temporelles
- Savoir décrire des séries temporelles
- Savoir choisir la modélisation adaptée en fonction de la série temporelle et de l'objectif visé
- Savoir comparer des séries temporelles entre elles
- Savoir utiliser les séries temporelles modélisées pour réaliser des prédictions

Méthode

Les différentes étapes de l'analyse des séries temporelles sont abordées en alternant théorie et application pratique. La théorie est abordée sans recours à des notions mathématiques complexes. Des exercices en autonomie sont réalisés régulièrement au cours de la formation.

Prévision pour AP



Durée : Trois jours

Interentreprises (Narbonne)

nombre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne
interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Les séries temporelles font parties du quotidien de tous : ce sont des mesures d'un même phénomène au cours du temps. Ce phénomène peut être écologique (la température mensuelle à Paris), économique (le cours de l'euro) ou encore sociologique (le taux de chômage). Ces phénomènes peuvent avoir des échelles de temps allant de quelques microsecondes à des centaines d'années. Les séries temporelles peuvent montrer plusieurs composantes (tendance, saisonnalité, bruit) et être modélisées. Avec un modèle satisfaisant, des prédictions peuvent être réalisées.

Programme

Introduction

Généralités

Caractéristiques des séries temporelles

Packages R spécifiques

Les dates et heures dans R

La fonction `ts()` et l'objet 'time series'

Principaux arguments de la fonction `ts()`

Exemples de paramétrages de séries temporelles

Conversion d'un tableau en série temporelle

Démarche globale et objectifs

Description des séries temporelles

Composantes d'une série temporelle

Le chronogramme

AirPassengers

Taux de change livres-dollars

Repérage de ruptures de tendance

Modifications de la série temporelle

Imputation de données manquantes ou aberrantes

Transformations des données

Décomposition : tendance, saisonnalité et bruit

Corrélations croisées

Modélisations et prévisions

Généralités

Filtrage et lissage exponentiel

Filtrage par moyenne mobile

Lissages exponentiels

Modèles ARIMA

Modèles (G)ARCH

Les modèles à seuil TAR

Quelques mots sur...

Les modèles multivariés

L'analyse spectrale

Les séries chronologiques à pas irrégulier

Exercices en autonomie sous RStudio

15. Les études sensorielles : éléments de méthodologie et analyses statistiques avec R

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants ayant à réaliser des études sensorielles et les analyser statistiquement. Les notions de base sur l'analyse de variance et les analyses multifactorielles (ACP, AFC, AFM, CAH) sont vivement recommandées.

Prérequis

Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires OU avoir suivi la formation Décrire et analyser les données multifactorielles. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Être capable de planifier une évaluation sensorielle en évitant les biais méthodologiques.
- Savoir choisir, réaliser et interpréter les analyses statistiques utiles aux analyses sensorielles à l'aide du logiciel R.
- Savoir choisir et mettre en œuvre l'épreuve sensorielle correspondant à une problématique de discrimination ou de similitude.
- Savoir analyser statistiquement et interpréter les résultats des études analytiques quantitatives.
- Être capable de mettre en œuvre et d'analyser les résultats d'une épreuve hédonique de classement réalisée sur un panel de consommateurs.
- Savoir construire et interpréter une carte de préférences.

Méthode

Pour chaque question, une explication de la méthode sans recours à des notions mathématiques. Un exemple en vraie grandeur est utilisé pour chaque technique. Les sorties logicielles et graphiques sont commentés. Les stagiaires réalisent eux-mêmes des analyses en auto-

mie.



Durée : Trois jours

Interentreprises (Narbonne)

nombre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

De nos jours des spécialistes testent, comparent, analysent, apprécient – ou pas – bien des objets de notre vie courante, objets alimentaires ou autres. Ces opérations, regroupées sous le nom "d'analyses sensorielles", donnent lieu à des tests de discrimination, à des évaluations par des juges entraînés, à des mises en relation des performances sensorielles avec des compositions physicochimiques, à des mesures de préférences par des publics variés de consommateurs potentiels.

Programme

Domaines d'application

Produits et consommateurs

Normalisation et approches sensorielles

Quelques éléments de méthodologie

Études orientées consommateur et études orientées experts.

Des biais nombreux et variés.

Traitements statistiques

Présentation rapide des outils logiciels : R et les bibliothèques utilisées

Rappels de statistiques : ACP, AFC, AFM, classification, Analyse de variance, concordance entre juges.

Épreuves analytiques

Épreuves discriminatives : test du triangle

Épreuves de similitudes : utilisation de notes, napping.

Épreuves quantitatives de notation : analyses de panel et de produit, relations avec les caractéristiques physico-chimiques.

Épreuves hédoniques

Épreuves de classement

Épreuves utilisant le discours

Cartographie des préférences



16. Manipulation et visualisation des données avec le logiciel R

Public concerné

Techniciens, ingénieurs, doctorants ou chercheurs désirant apprendre à manipuler les données et à les visualiser avec le logiciel R. Les notions de statistiques ne seront pas abordées dans cette formation.

Prérequis

Avoir suivi la formation Introduction au logiciel R OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires ; savoir modifier des lignes de codes ou savoir écrire ses propres lignes de codes. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Être capable d'utiliser les fonctionnalités de base de RStudio en autonomie
- Être capable de créer des rapports automatisés avec Rstudio
- Être capable de manipuler efficacement des jeux de données
- Être capable d'utiliser la syntaxe en tuyaux
- Être capable de mettre en pratique des conditions et des fonctions personnelles
- Être capable d'utiliser des packages modernes et puissants pour les visualisations graphiques statiques et générer des graphiques interactifs

Méthode

Les deux premiers jours sont consacrés à une présentation théorique des différents outils. Chaque sous-paragraphe est accompagné de plusieurs applications pratiques et d'exercices en autonomie.

La dernière journée est une journée d'atelier sur données personnelles où les participants pourront appliquer directement les outils vus les deux premiers jours.

Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises
4926 € intra-entreprise

Nos formations sont exonérées de TVA.

La manipulation des données avec des outils simples et performants devient incontournable à l'heure des Big Data. La visualisation est essentielle tant pour l'exploration des données que pour la publication de résultats.

Après deux jours de présentation et de pratique, cette formation vous permet d'appliquer immédiatement ces outils sur vos propres données le troisième jour, avec l'accompagnement de notre formateur.

Programme

R-Studio

Prise en main

Produire des documents à partir des sorties R

Principe du R Markdown

Créer un document HTML, Word ou pdf avec Rstudio

La syntaxe de base de R Markdown

Manipulation des données avec R

Fonctions prédéfinies classiques

Les opérateurs

Les fonctions de manipulation

Fonctions prédéfinies des packages {dplyr} et {tidyr}

Fonctions de manipulation

Enchaînement des commandes avec {dplyr}

Fonctions de jointure

Fonctions personnelles

Visualisations graphiques avec R

Utiliser {ggplot2}

Principes généraux

Utiliser qplot()

Utiliser ggplot()

Améliorations

Utiliser {ggpubr}

Utiliser {GGally}

Utiliser {ggvis}

Atelier sur données personnelles

17. Créer des applications Shiny avec le logiciel R

Public concerné

Techniciens, ingénieurs, chercheurs ou doctorants. Des notions d'HTML et/ou de CSS sont les bienvenues mais pas indispensables. Une utilisation régulière et fluide du logiciel R est très souhaitable.

Prérequis

Avoir suivi la formation Utilisation avancée du logiciel R OU Manipulation et visualisation des données OU savoir utiliser le logiciel R pour importer des données, les visualiser, gérer les packages nécessaires ; savoir modifier des lignes de codes et/ou savoir écrire ses propres lignes de codes; savoir écrire des fonctions pour un usage personnel. Un questionnaire préalable permettra leur évaluation.

Objectifs

- Être capable de décrire le concept général des applications Shiny (entrées, sorties, réactivité)
- Être capable de construire des applications simples et plus élaborées
- Être capable de nommer trois possibilités pour la mise en forme des applications
- Savoir mettre en pratique la réactivité des exécutions
- Savoir partager son application

Méthode

Jours 1 et 2 : Après une présentation des éléments de base permettant de réaliser des applications basiques, les notions de réactivité et ses nuances sont expliquées au travers d'exemples concrets et de schémas.

Enfin, des éléments de mise en forme sont présentés et illustrés

Chaque section de la formation alterne théorie et mise en pratique et propose des exercices en autonomie.

Jour 3 : La dernière journée est consacrée à la construction d'applications par les participants pour répondre à un besoin qui leur est propre. Chaque participant travaille en autonomie, avec l'aide du formateur.

Durée : Trois jours

Interentreprises (Narbonne)

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 12 personnes au plus.

Dates : à définir avec le client.

Prix : 1625 €/personne interentreprises

4926 € intra-entreprise

Nos formations sont exonérées de TVA.

Programme

Introduction

Le fonctionnement de base d'une application Shiny

Principe

Le corps d'une application

Les entrées et les sorties : les entrées, les sorties, les fonctions `render()` et la réactivité*

Partager son application

Exercices

Personnaliser les réactions

Pourquoi la réactivité "Classique" peut ne pas convenir ?

Utiliser les fonctions `reactive()` ou `isolate()`

Utiliser les boutons d'action

Utiliser des valeurs réactives

Schéma récapitulatif

Exercice

Personnaliser l'apparence de l'application

L'organisation de la page

Ajout d'éléments statiques

Habillage de l'application

Exercices

Atelier : mise en application sur données personnelles avec R Studio

18. Atelier de traitement de données personnelles

Public concerné

Techniciens, ingénieurs, doctorants ou chercheurs souhaitant bénéficier d'un soutien personnalisé pour la mise en œuvre de leurs analyses statistiques à la suite de la ou des formations qu'ils auraient précédemment suivie(s). Les connaissances de base des outils statistiques à mettre en œuvre pendant l'atelier sont très souhaitables.

Prérequis

Avoir suivi l'une des formations dispensées par AnaStats. Maîtriser le logiciel choisi pour les analyses à réaliser. Un questionnaire préalable permettra une évaluation.

Objectifs

Mettre en pratique les enseignements sur les données des participants. Permettre aux participants d'appliquer les acquis des formations au contexte familier de leurs problématiques, avec le soutien du formateur, de manière à optimiser les analyses de données une fois de retour dans leur laboratoire.

Méthode

La présentation collective des données permet aux stagiaires de faire le point sur les méthodes qu'ils devront utiliser mais également de s'associer à d'autres participants en fonction de la proximité de leurs données et/ ou de leur problématique pour travailler collectivement et échanger.

Par la suite, les participants évoluent en totale autonomie avec leurs données.

Le formateur fera en sorte de partager son temps entre les stagiaires qui ont régulièrement besoin de conseils pour les méthodes utilisées et pour les interprétations statistiques.

Durée : Deux jours

Interentreprises :

nbre de stagiaires limité à 6

Dates : nous consulter

Intra-entreprise :

groupe de 6 personnes au plus.

Dates : à définir avec le client.

Prix : 1192 €/personne
interentreprises
3448 € intra-entreprise

Nos formations sont exonérées de TVA.

Mettre en application les enseignements des statistiques sur ses propres données garantit l'assimilation des enseignements et le gain de temps dans la mise en route du traitement de données.

Programme

Présentation des données par les stagiaires et discussions collectives relatives aux analyses.

Rappels propres au(x) logiciel(s) et aux traitements statistiques réalisés en fonction des besoins spécifiques de chacun.

Travail individuel sur les données personnelles avec l'appui du formateur qui partage son temps entre les participants. Des rappels pourront être faits régulièrement selon les besoins.

Synthèse de l'atelier de traitement de données par les participants. Cette étape permet aux stagiaires de confronter leurs avis sur les différents traitements réalisés, d'ouvrir des pistes de réflexion et de développer leur sens critique sur les outils qu'ils auront maîtrisés.

Si besoin, des rappels de statistiques seront formulés.

Chaque participant apportant à la fois sa problématique et ses données, celles-ci devront être bien choisies et préparées à l'avance de manière à permettre l'analyse statistique sans avoir à retravailler la présentation du ou des tableaux de données.

~~~~~





## Pour nous contacter

courriels : [sophie.dubois@anastats.fr](mailto:sophie.dubois@anastats.fr)  
[severine.montaudouin@anastats.fr](mailto:severine.montaudouin@anastats.fr)

**Les Vigneaux**  
**F - 37220 Rilly Sur Vienne**  
Tél. 02 47 58 70 33

**[www.anastats.fr](http://www.anastats.fr)**