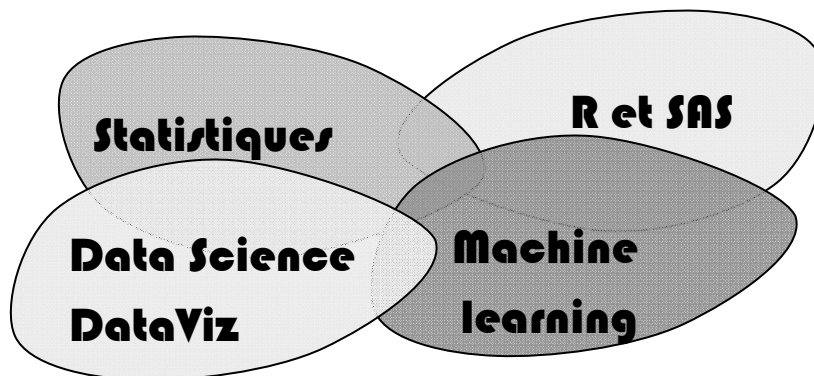




Olivier Decourt SARL



Catalogue formations 2019

Nous contacter : formation@od-datamining.com

06 64 77 95 63

Visitez notre site : <http://www.od-datamining.com/>





Sommaire

Tarifs et conditions.....	5
R : maîtriser un logiciel souple, polyvalent et gratuit.....	6
[R_BASE] Initiation à R.....	7
[SAS2R] Migrer de SAS vers R.....	8
[R_PLUS] Aller plus loin avec R	9
[R_GGLOT] Produire des graphiques avec le package ggplot2	10
[R_REPORT] Construire des rapports avec R.....	11
[R_SHINY] Construire des applications interactives avec le package shiny	12
SAS : exploiter et mettre en valeur des données volumineuses	13
[SAS_BASE] Initiation à SAS	14
[SEG] Utiliser SAS Enterprise Guide	15
[SEG_PGM] Programmer dans SAS Enterprise Guide.....	16
[SAS_OPTIM] Optimisation de code SAS	17
[SAS_MACRO] Automatisation de code SAS	18
[SAS_ODS] Diffusion de résultats SAS	19
[SAS_GRAPH1] Procédure SGLOT.....	20
[SAS_GRAPH2] Modèles graphiques et GTL.....	21
[SAS_CARTO] Cartographie avec SAS	22
Formations statistiques : DataViz, Data Science, Data Mining, machine learning	23
[VISUAL] Présenter clairement des données, construire des graphiques intelligents	24
[STAT101] Initiation à la statistique.....	25
[ANADON] Analyse des données	26
[GENMOD] Modèle linéaire généralisé.....	27
[MIXED] Modèles mixtes et hiérarchiques	28
[ML_R] Machine learning avec R	29
[REGQUALI] Régression logistique.....	30
[SCORING_SAS] Construire un score avec SAS.....	31



Formations métier : biostatistique, marketing, actuariat	32
[BAYES] Introduction à la statistique bayésienne.....	33
[GLMTARIF] Modélisation pour la tarification en assurance	34
[POWER] Calcul de puissance et de nombre de sujets nécessaires	35
[TYPOMKT] Typologie pour la segmentation client	36



Tarifs et conditions

Groupes de 1 à 8 personnes

Nous proposons un **tarif unique** en intra-entreprise, quel que soit le niveau du cours : **1 250 € HT par jour de formation**.

Ce tarif est applicable à tout groupe de huit personnes maximum. Il inclut la réalisation et la distribution de supports de cours pour chaque participant.



Notre organisme de formation est déclaré auprès de la préfecture de région, et facture la TVA (20%). Nous sommes également enregistrés dans la base DataDock, ouvrant le droit au financement de nos formations par votre OPCA.

Les cours peuvent se dérouler en anglais ou en français.

Formations sur mesure

Les contenus de cours qui sont détaillés dans les pages suivantes ne sont pas limitatifs. Il ne s'agit que des contenus **standards**.

Ceux-ci peuvent être **adaptés à vos besoins**, qu'il s'agisse de la modification d'un contenu existant ou de la création d'un nouveau cours. Ces modifications n'entraînent **aucune modification des tarifs ci-dessus**.

Les logiciels sur lesquels la formation peut s'appliquer sont indiqués à chaque cours. Si vous avez une demande sur un autre outil, n'hésitez pas à en faire la demande.

Cours en province ou à l'étranger

Les frais de déplacement et d'hébergement sur place du formateur seront facturés en sus sur présentation de justificatifs.

Supports

Les supports sont fournis à chaque participant sur des clés USB. Ces supports sont en français (ou, sur demande, en anglais) et mêlent théorie et applications.



R : maîtriser un logiciel souple, polyvalent et gratuit



[R_BASE] Initiation à R

Le logiciel R est principalement conçu pour des utilisations statistiques. Il recèle cependant de très nombreuses fonctionnalités de gestion de fichiers. Cette formation peut être couplée avec STAT101, pour prolonger l'apprentissage avec la production de statistiques descriptives sous R.

Durée : 2 jours

Pré-requis : aucun

Logiciels possibles : R

1. Présentation de R

- Télécharger le logiciel
- Packages
- Environnement de base
- R Studio
- R Commander

2. Principes du langage R

- Fonctions : principes, utilisation, personnalisation
- Types de données
- Structures de données

3. Récupération de données

- Import d'un fichier texte
- Import d'une table SAS
- Import d'un fichier Excel

4. Vérification du contenu d'une table

- Statistiques descriptives
- Tableaux de fréquence
- Graphiques en bâtons
- Histogrammes et boxplots

5. Manipulation de données

- Création de variables, fonctions
- Empilement de tables
- Fusions de tables
- Transpositions et réorganisations des données

6. Sauvegardes

- Sauvegarde de données
- Export de données



[SAS2R] Migrer de SAS vers R

Cette formation est spécialement conçue pour des utilisateurs réguliers de SAS voulant découvrir la logique de R et les différences qui les attendent dans ce nouvel outil. L'accent est mis sur les fonctionnalités courantes de SAS et les pistes pour reproduire ces traitements dans R.

Durée : 2 ou 3 jours selon les attentes et le niveau du public

Logiciels possibles : R

**Nouvelle
formation**

Pré-requis :
utilisation régulière de SAS (niveau SAS_BASE minimum)

1. Quelques points de repère dans R

- Concepts essentiels : fonction, package, objet, affectation
- Environnement R Studio
- Récupérer ses tables SAS dans R : comment sont-elles stockées
- Sauvegarde et récupération de données
- Enchaînements

2. Remplacer les procédures usuelles

- Proc Contents : fonctions colnames, class, str, glimpse
- Proc Print : fonctions print, as.tibble
- Proc Means : fonctions summary, aggregate, describe
- Proc Freq : fonctions table, prop.table, addmargins
- Proc Tabulate : package {tables}
- Proc Transpose : package {reshape2}
- Proc SQL : package {dplyr}
- Proc Sort : fonctions order, duplicated, duplicated2
- Proc Export : fonction write.table, package {xlsx}

3. Remplacer l'étape DATA

- Ajout de variables : principales fonctions
- Empilement de tables : fonction rbind
- Jointures : fonctions merge, *_join
- Retain : fonctions cumsum, rank

4. Remplacer l'ODS

- ODS HTML, ODS RTF, ODS PDF : documents Rmarkdown
- ODS RTF, ODS POWERPOINT : package {officer}
- ODS EXCEL : packages {openxlsx} et {xltabr}

5. Remplacer le langage macro

- Macro-variables : vecteurs isolés
- Macro-programmes : fonctions, boucles, vectorisation



[R_PLUS] Aller plus loin avec R

Ce stage permet de se perfectionner autour des fonctionnalités de packages du tidyverse (`{dplyr}`, `{tidyr}`, `{forcats}`, `{purrr}`), de la création de fonctions personnalisées et des `data.tables`.

Durée : 2 jours

Pré-requis :
R_BASE

Logiciels possibles : R

1. Fonctions personnalisées et automatisation

- Construire une fonction
- Boucles
- Structures conditionnelles
- Vectorisation (`sapply`, `lapply`, package `{purrr}`)

2. Gestion des données

- Package `{reshape2}` : transpositions
- Package `{dplyr}` : requêtes
- Package `{forcats}` : gestion des facteurs
- Package `{lubridate}` : gestion des dates
- Gestion des doublons

3. Le package `{data.table}`

- Filtres
- Index
- Agrégats
- Jointures



[R_GGLOT] Produire des graphiques avec le package ggplot2

Ce stage vise d'application des principes du dataviz avec le package {ggplot2} : construction de graphiques pour l'exploration statistique et la publication.

Durée : 2 jours

Logiciels possibles : R

Pré-requis :
R_PLUS, VISUAL

1. Le package {ggplot2}

- Principes
- Exporter un graphique, sauvegarde et recyclage

2. La fonction ggplot et ses compléments

- Grammaire des graphiques, les composants d'un graphique
- Données
- Esthétique
- Géométrie
- Traitement statistique
- Coordonnées
- Légendes
- Axes
- Thème

3. Eclatement (*facetting*)

- Principe
- Eclatement



[R_REPORT] Construire des rapports avec R

Ce stage permet d'explorer les différents packages de restitution automatique avec R. Il permet de produire des documents Word ou PowerPoint avec le package `{officer}`, des classeurs Excel avec `{opensxls}` et `{xltabr}`, HTML, PDF et Word avec Rmarkdown.

Durée : 2 jours

Logiciels possibles : R 3.4 et +

**Nouvelle
formation**

Pré-requis :
R_GGPLOT

1. Reporting avec Rmarkdown

- Principes
- Textes
- Tableaux
- Graphiques
- En-tête YAML

2. Tableaux croisés

- Package `{reshape2}`
- Package `{tables}`
- Package `{flextable}`

3. Documents Word et PPT (package `{officer}`)

- Créer un document avec `{officer}`
- Insérer du texte
- Insérer un tableau
- Insérer un graphique
- Utilisation de liens et de document-maître

4. Classeurs Excel (packages `{openxlsx}` et `{xltabr}`)

- Créer un classeur avec `{openxlsx}`
- Ajouter un onglet
- Insérer un `data.frame`
- Insérer un tableau
- Ajouter de la mise en forme



[R_SHINY] Construire des applications interactives avec le package shiny

Ce stage permet de s'initier au développement de mini-applications d'entreprise avec le package {shiny} : interfaces homme/machine, calculs et restitution.

Durée : 1 jour

Pré-requis :
R_PLUS

Logiciels possibles : R

1. Principes de shiny

- Fonctionnement d'une application shiny
- Transmission d'informations
- Réactivité

2. Fonctions pour les calculs

- Scripts server.R et global.R
- Fonction shinyServer
- Fonction reactive
- Fonction validate
- Fonction isolate
- Fonction observe
- Mise à jour d'objets input

3. Fonctions pour l'interface

- Fonction de mise en page
- Exemples de mise en page
- Objets input
- Objets output
- Insérer des éléments



SAS : exploiter et mettre en valeur des données volumineuses



[SAS_BASE] Initiation à SAS

Le langage SAS permet de manipuler de très gros volumes de données et de les valoriser par des statistiques, des tableaux de bord et des graphiques. Cette introduction permet de se familiariser avec l'interface (SAS classique, WPS Workbench ou SAS Enterprise Guide) et la logique du code. On pourra la prolonger par une formation au langage macro pour l'automatisation, par un approfondissement des fonctionnalités graphiques, des capacités de restitution (reporting et tableaux de bord, graphiques, cartographie) ou d'analyse statistique.

Durée : 2 jours

Logiciels possibles : SAS,
WPS

**Nouvelle
formation**

Pré-requis :
aucun

1. Présentation de SAS

- Environnements : SAS classique, SAS Enterprise Guide, WPS
- Logique du code : étape Data, procédure, ODS, etc.
- Organisation des données : bibliothèque, table, variable

2. Requêtes

- Listes et extractions : proc PRINT et clause WHERE
- Tris et doublons : proc SORT
- Restrictions : options de table KEEP, DROP et OBS

3. Production de statistiques basiques

- Comptages : proc FREQ
- Autres indicateurs : proc MEANS
- Esthétique, recodage et catégories : proc FORMAT
- Graphiques simples : proc SGPLOT

4. Modification du contenu d'une table

- Fonctionnement global de l'étape DATA
- Création de variables : formules simples et conditionnelles
- Jointures

5. Imports et exports

- Importation de formats courants : proc IMPORT
- Diffusion de résultats mis en forme : ODS



[SEG] Utiliser SAS Enterprise Guide

L'interface SAS Enterprise Guide utilisée en mode presse-bouton permet d'utiliser la puissance du langage SAS sans en connaître la syntaxe en détail. Cette formation présente la logique de fonctionnement de ce logiciel, et permet de s'approprier rapidement les traitements du quotidien : requêtes, statistiques descriptives, imports et exports.

Durée : 2 jours

Logiciels possibles : SEG
(SAS Enterprise Guide)

**Nouvelle
formation**

Pré-requis :
aucun

1. Environnement et vocabulaire

- Architecture informatique
- Projet
- Flux
- Données
- Les différentes fenêtres

2. Organisation des données

- Bibliothèque
- Table SAS
- Variables, type, nom et label
- Observations, nombre de lignes

3. Générateur de requêtes

- Requête simple
- Filtres simples et complexes
- Gestion des doublons
- Création de variables
- Agrégation statistique
- Jointure

4. Tableaux et graphiques

- Statistiques simples
- Graphiques
- Tableau croisé

5. Import et export

- Import de données
- Export de données
- Export de tableaux et résultats



[SEG_PGM] Programmer dans SAS Enterprise Guide

L'environnement SAS Enterprise Guide (SEG) propose de nombreuses facilités pour écrire du code. Cette formation les détaille pour que ceux qui codent puissent tirer parti au maximum de l'éditeur de code, des invites, des flux et des projets.

Durée : 1 jour

Logiciels possibles : SAS

**Nouvelle
formation**

Pré-requis :
SAS_BASE

1. Environnement et vocabulaire

- Architecture informatique
- Projet
- Flux
- Données
- Les différentes fenêtres

2. Fenêtre de code

- Abréviations
- Souffleur de syntaxe
- Gestion du Journal (Journal du projet, résumé du Journal)
- Réorganisation automatique du code

3. Programmes existants

- Lien externe ou incorporation du projet
- Flux AUTOEXEC
- Instructions à bannir
- Instructions à adapter

4. Applications stockées (facultatif)

- Créer une application stockée
- Utiliser une application stockée depuis Add-In Office



[SAS_OPTIM] Optimisation de code SAS

Certains traitements SAS portant sur de très grosses volumétries peuvent durer assez longtemps. Cette formation propose une série d'astuces de programmation pour diminuer le temps de traitement, ou la quantité de mémoire (disque ou RAM) consommée, selon les contraintes.

Durée : 1 jour

Logiciels possibles : SAS

**Nouvelle
formation**

Pré-requis :
SAS_BASE

1. Optimisation générale : grands principes

- Lire les données le moins souvent possible
- Lire le strict nécessaire
- Minimiser la mémoire utilisée

2. Agrégation

- Choisir entre BY et CLASS dans la procédure MEANS
- Instructions WAYS et TYPES, option CHARTYPE dans la procédure MEANS
- SQL ou procédure MEANS, quel code est le plus rapide ?

3. Jointures

- Rappels sur les jointures SQL et sur l'instruction MERGE
- Index et double SET
- Table de hachage
- Formats



[SAS_MACRO] Automatisation de code SAS

Le langage macro de SAS permet de répéter un traitement avec des paramètres ; on produit ainsi du code modulaire plus souple, plus générique et plus réutilisable.

Cette formation s'adresse à des personnes utilisant déjà régulièrement le code SAS.

Durée : 2 jours

Logiciels possibles : SAS

**Nouvelle
formation**

Pré-requis :
SAS_BASE

1. Principes du langage macro

- Manipuler du texte
- Rôle des deux compilateurs
- Véhicule principal : les macro-variables

2. Macro-variables

- Macro-variables automatiques
- Affichage : %PUT, & et variantes
- Création : %LET, CALL SYMPUTX, SELECT INTO
- Suppression : %SYMDEL
- Modification : macro-fonctions

3. Macro-programmes

- Paramètres
- Macro-conditions
- Macro-boucles
- Macro-variables locales et globales
- Stockage, appel, débogage



[SAS_ODS] Diffusion de résultats SAS

Parmi les fonctionnalités enthousiasmantes de SAS, l'ODS occupe une place de choix. Il permet de transcrire dans divers formats courants (Excel, Word, PDF, Powerpoint, HTML, etc.) n'importe quelle sortie mise en forme, que ce soit du texte, un tableau ou un graphique. Une formation spécifique aux graphiques pourra d'ailleurs être couplée avec celle-ci pour produire des tableaux de bords plus complets.

Durée : 2 jours

Logiciels possibles : SAS

**Nouvelle
formation**

Pré-requis :
SAS_BASE

1. Principes de l'ODS

- Destinations
- Objets, liste noire, liste blanche
- Syntaxe commune aux principales destinations

2. Contenu : produire de beaux tableaux

- Tableaux croisés : proc TABULATE
- Tableaux et calculs : proc REPORT
- Tableaux sur mesure : ODSOUT

3. Contenu : insérer du texte

- Titres et pieds de page
- Caractères spéciaux et ODS ESCAPECHAR
- Paragraphe simple : ODS TEXT
- Plusieurs paragraphes, listes : proc ODSTEXT et proc ODSLIST

4. Principales destinations : usage

- Classeurs Excel : ODS TAGSETS.EXCELXP, ODS EXCEL
- Documents Word : ODS RTF
- Documents Acrobat : ODS PDF
- Pages web : ODS HTML
- Présentations : ODS POWERPOINT



[SAS_GRAPH1] Procédure SGPLOT

Ce stage est destiné aux chargés d'études souhaitant découvrir toute la richesse de la procédure SGPLOT. Cette formation présente également son complément, la procédure SGPANEL, et peut se poursuivre par une formation aux modèles écrits en GTL.

Durée : 1 ou 2 jours selon les attentes et le niveau du public

Pré-requis :
SAS_BASE

Logiciels possibles : SAS

1. ODS Graphics : un principe de fonctionnement et une instruction-clé

- Filière graphique historique vs filière « ODS Graphics » Java
- Instruction ODS Graphics
- Instructions ODS : choisir l'emplacement de l'image produite

2. Graphiques sur données qualitatives

- Diagrammes en bâtons
- « Dot plots »
- Barres et courbes

3. Graphiques sur données quantitatives

- Histogrammes, courbes de densité
- Boxplots
- Nuages de points, courbes, lissages
- Droites de régression, intervalles de confiance
- Zones colorées

4. Titres, légendes, axes et textes

- Gérer des légendes
- Insérer du texte fixe ou dynamique
- Définir des axes
- Utiliser des caractères spéciaux

5. Plusieurs graphiques avec la procédure SGPANEL

- Gérer les variables catégorielles
- SGPANEL et bloc BY : différences et complémentarité



[SAS_GRAPH2] Modèles graphiques et GTL

Ce stage est destiné aux chargés d'études qui voudraient produire des modèles réutilisables de graphiques ; pour cela, à partir de SAS 9.1, le GTL ou Graph Template Language, qui sous-tend le système ODS Graphics, permet de produire de belles sorties standardisées.

Cette formation indiquera selon les versions de SAS quelles syntaxes et quelles possibilités offre le GTL, un langage qui s'enrichit très rapidement au fil des versions.

Durée : 1 ou 2 jours selon les attentes et le niveau du public

Pré-requis :
SAS_GRAPH1

Logiciels possibles : SAS

1. Modèles de graphiques : le principe du GTL

- Procédures statistiques, modèles graphiques
- ODS PATH : où se trouvent les modèles ?
- Créer un modèle à partir de la procédure SGPLOT
- Utiliser un modèle avec la procédure SGRENDER

2. Paramétrer un modèle

- Paramètres dynamiques
- Macro-variables contenant du texte
- Macro-variables contenant des nombres

3. Principaux éléments d'un modèle

- Canevas (« lattice »)
- Élément graphique
- Axes

4. Principaux éléments graphiques

- Barres, boxplots, dot plots
- Nuages de points, courbes, lissages, régressions
- Zones colorées, quadrillage (« blockplot »)
- Légendes, titres
- Caractères spéciaux, mise en forme des éléments

5. Canevas dynamique

- Le couple LAYOUT DATALATTICE / LAYOUT PROTOTYPE
- Canevas dynamique vs macro-programme



[SAS_CARTO] Cartographie avec SAS

Ce stage est destiné aux chargés d'études (géomarketing, épidémiologie, sociologie, ...) ayant à produire des cartes avec SAS.

Cette formation permet de se familiariser avec les cartes via la procédure SGPLOT – et la procédure SGMAP à partir de SAS 9.4M5.

Durée : 1 journée

Logiciels possibles : SAS

**Nouveau
contenu**

Pré-requis :
SAS_GRAPH1

1. Fonds de carte

- Fonds de carte fournis par SAS (bibliothèque MAPS et MAPSSAS)
- Fonds de carte GFK (bibliothèque MAPSGFK, SAS 9.4)
- Import de fonds de cartes externes (proc MAPIMPORT, étape Data)
- Création d'un identifiant géographique unique

2. Cartographie

- Données à représenter
- Fonction POLYGON
- Légende
- Bulles
- Textes
- Flèches
- Procédure SGMAP (SAS 9.4M5 et versions supérieures)

3. Manipulations sur les fonds de carte

- Fusion des éléments de la carte : proc GREMOVE
- Simplification d'un fond de carte : proc GREDUCE
- Plusieurs cartes côte à côte (zoom urbain, outremer) en GTL



Formations statistiques : DataViz, Data Science, Data Mining, machine learning



[VISUAL] Présenter clairement des données, construire des graphiques intelligents

Ce stage est destiné à tous ceux qui ont besoin de résumer des jeux de données par des graphiques, qu'il s'agisse de présentations, de tableaux de bord ou de recherche exploratoire.

Cette formation se concentre sur les différents types de graphiques et les moyens de leur donner une efficacité maximale. Elle pourra être couplée avec une formation aux graphiques SAS ou R. Cette formation peut être couplée avec STAT101 pour découvrir les indicateurs statistiques usuels et faire le lien avec leur diffusion efficace.

Durée : 1 ou 2 jours selon les attentes et le niveau du public

Pré-requis :
aucun

Logiciels possibles : Excel, SAS, R

1. Principes de base pour une présentation efficace

- Principes du gestaltisme
- Attributs pré-attentifs

2. Présentation des tableaux

- Styles des cellules (nombres, pourcentages, etc.)
- Bordures, fonds de couleur : comment orienter la lecture d'un tableau
- Sources et titres

3. Différents types de graphiques pour représenter...

- ... une variable qualitative
- ... une variable quantitative
- ... une série chronologique (évolution, tendance, lissage)
- ... plusieurs variables simultanément

4. Efficacité d'un graphique

- Axes, légendes
- Choix des couleurs
- Treillis et graphiques multiples



[STAT101] Initiation à la statistique

Ce stage est destiné aux personnes désireuses de découvrir les principes et les applications de la statistique. Il couvre principalement la statistique descriptive à une ou deux variables (graphiques et tableaux) et se termine sur un élargissement aux techniques plus avancées (tests, prévisions).

Ce cours est prévu sous forme appliquée ; seules les formules indispensables seront présentées.

Durée : 2 jours

Logiciels possibles : Excel, SAS, R

Pré-requis :
manipulations de
base du logiciel
d'application

1. Vocabulaire et outils

- Distinguer qualitatif et quantitatif
- Modalité, cardinalité
- Population, champ d'une étude, individu

2. Résumer (une information)

- Notion de distribution
- Indicateurs numériques : moyenne, médiane, quantiles
- Dispersion : variance, écart-type, coefficient de variation
- Tableaux de fréquence
- Graphiques : bâtons, boxplots, histogrammes, courbes de densité

3. Comparer (deux informations ou plus)

- Graphiques : nuages de points, courbes
- Droite de tendance, lissage
- Graphiques : bâtons, radars
- Tableaux croisés
- Corrélations
- Indices

4. Extrapoler

- Sondages, échantillons, pondérations, redressement
- Tests statistiques : principe
- Test du khi-2, test de Student
- Modélisation (régressions)
- Traitement de séries chronologiques : désaisonnalisation, tendance



[ANADON] Analyse des données

Ce stage est destiné aux chargés d'études qui désirent voir ou revoir les principes de l'analyse factorielle (ACM, ACP) et surtout leur utilisation à travers SAS/STAT ou R (package {FactoMineR}). On y aborde également la classification.

Durée : 2 jours

Pré-requis :
STAT101

Logiciels possibles : SAS, R

1. L'analyse en composantes principales (ACP)

- Choix du nombre d'axes factoriels
- Nuages des individus et des variables
- Cercle des corrélations
- Rotation VARIMAX et ACP

2. L'analyse des correspondances multiples (ACM)

- Choix du nombre d'axes factoriels
- Nuages des individus et des variables
- Individus et variables supplémentaires

3. Typologies

- Classification ascendante hiérarchique
- Nuées dynamiques
- Méthode mixte de Wong
- Description des classes
- Modéliser l'appartenance aux classes pour réaffecter



[GENMOD] Modèle linéaire généralisé

Les modèles présentés ici font de la régression linéaire et de la régression logistique des cas particuliers. Les Modèles Linéaires Généralisés (MLG) se proposent d'étudier les variables dont la normalité est prise en défaut (coûts, fréquences d'évènements, ...) et proposent des outils puissants.

Durée : 2 jours

Pré-requis :
STAT101

Logiciels possibles : SAS, R

1. Principes de la régression

- Vocabulaire et concepts
- La régression linéaire
- La régression logistique
- Leurs points communs

2. Modèle linéaire généralisé

- Loi de Y
- Fonction de lien
- Qualité du modèle
- Analyse de la déviance
- Analyse des résidus et autres vérifications

3. Exemples de modèles linéaires généralisés

- Régression de Poisson
- Régression binomiale négative
- Régression Gamma

4. Données répétées et corrélées

- GLMM : des modèles très flexibles
- Choix d'une structure de corrélation



[MIXED] Modèles mixtes et hiérarchiques

Cette formation permet de maîtriser des modèles qui généralisent l'analyse de variance, en particulier pour le traitement des données à mesures répétées. Un autre champ d'application est la construction de modèles hiérarchiques (ou multiniveaux) faisant intervenir des organisations emboîtées (en éducation : établissement, classe, élève).

Durée : 2 jours

Logiciels possibles : SAS, R

**Nouveau
contenu**

Pré-requis :
STAT101

1. Principes des modèles mixtes

- Effet fixe et effet aléatoire
- Données corrélées, répétées
- Différents types de modèles mixtes
- Méthodologie de construction des modèles

2. Modèle linéaire mixte

- Données corrélées non répétées
- Données répétées : pentes aléatoires
- Données répétées : structure de covariance
- Modèle hiérarchique / multiniveaux

3. Modèles mixtes généralisés (GLMM)

- Principe et théorie des modèles linéaires généralisés
- Interprétation des coefficients
- Données corrélées non répétées
- Données répétées : pentes aléatoires
- Données répétées : structure de covariance
- Modèle hiérarchique / multiniveaux



[ML_R] Machine learning avec R

Ce stage apprend à construire des scores et des modèles statistiques prédictifs avec R : statistique exploratoire supervisée, régression logistique, arbres de décision, réseaux de neurones, forêts aléatoires, ... et la comparaison de modèles pour choisir le plus optimal.

Durée : 3 jours

Pré-requis :

Logiciels possibles : R

R_BASE,
STAT101

1. Gestion des données

- Imputation
- Equilibrage
- Bases d'apprentissage, de validation et de test

2. Statistiques descriptives

- Graphiques
- Mesures de liaison
- Découpage en tranches

3. Modélisation

- Régression logistique
- Arbre de décision
- Analyse discriminante
- SVM
- Forêts aléatoires
- Réseaux de neurones (PMC)

4. Comparaison de modèles

- Indicateurs numériques
- Représentations graphiques
- Seuil optimal



[REGQUALI] Régression logistique

Destiné aux chargés d'étude s'intéressant à la modélisation d'une variable binaire, ce stage permet de construire efficacement des modèles explicatifs et prédictifs (construction de scores).

Durée : 2 jours

Pré-requis :
STAT101

Logiciels possibles : SAS, R

1. Principe de la régression logistique

- Quelle est la forme des données à utiliser ?
- Lien avec la régression linéaire
- Les différentes fonctions de lien
- Mesurer la qualité d'un modèle logistique

2. La régression logistique à but descriptif

- Tests de type 3
- Stratégies de construction de modèles cohérents
- Les coefficients
- Les odds-ratios
- La multicolinéarité

3. La régression logistique à but prédictif

- Qu'est-ce qu'un score ?
- La courbe ROC et le seuil optimal
- La courbe de lift
- Qualité d'ajustement



[SCORING_SAS] Construire un score avec SAS

Cette formation s'adresse aux chargés d'étude désirant avoir, en quelques jours, un aperçu technique et pratique des techniques usuelles de scoring sous SAS.

Durée : 2 jours

Logiciels possibles : SAS

**Nouveau
contenu**

Pré-requis :
STAT101 et
SAS_BASE

1. Gestion des données

- Imputation
- Equilibrage
- Bases d'apprentissage, de validation et de test

2. Statistiques descriptives

- Graphiques
- Mesures de liaison
- Découpage en tranches

3. Modélisation

- Régression logistique
- Arbre de décision
- Analyse discriminante

4. Comparaison de modèles

- Indicateurs numériques
- Représentations graphiques
- Seuil optimal



Formations métier : biostatistique, marketing, actuariat



[BAYES] Introduction à la statistique bayésienne

Ce stage décrit les principes et les applications en épidémiologie des mécanismes de statistique bayésienne et ses différences avec la statistique classique (fréquentiste).

Des applications seront présentées en utilisant des logiciels libres (R et Winbugs).

Durée : 1 jour

Pré-requis :
STAT101

Logiciels possibles : R, Winbugs

- 1. Introduction à l'approche bayésienne**
- 2. Lois a priori, vraisemblance, lois a posteriori, formule de Bayes**
- 3. L'analyse bayésienne**
- 4. Initiation aux méthodes MCMC (chaînes de Markov par Monte-Carlo), échantillonneur de Gibbs**
- 5. Intervalles de crédibilité**
- 6. Modèles de régressions bayésiens, DIC**
- 7. Notions de tests bayésiens, facteur de Bayes**
- 8. Comparaison des estimateurs bayésiens et fréquentistes**



[GLMTARIF] Modélisation pour la tarification en assurance

Ce stage est destiné aux chargés d'étude et actuaires qui ont à modéliser les sinistralités et coûts moyens en assurance non vie. Les applications se font sur des modèles composites (IARD) et de prime pure (santé).

Durée : 2 jours

Pré-requis :
STAT101

Logiciels possibles : SAS, R

1. Modèle Gamma pour le coût moyen

- Phase exploratoire : adéquation à la loi
- Phase exploratoire : lien log et variables quantitatives
- Construction et simplification du modèle
- Commentaire du modèle
- Cas particulier : modèle Gamma pour la prime pure

2. Modèle poissonnien pour la sinistralité

- Agrégation des données : comment et pourquoi ?
- Phase exploratoire : lien log et variables quantitatives
- Choix de la loi : Poisson ou binomiale négative ?
- Construction et simplification du modèle
- Commentaire du modèle
- Extension : modèles ZIP et ZINB

3. Modélisation des données répétées

- Approche GEE
- Approche GLMM
- Choix de la structure de corrélation
- Interprétation des résultats



[POWER] Calcul de puissance et de nombre de sujets nécessaires

Ce stage est destiné aux biostatisticiens désireux de calculer la puissance de leurs tests statistiques ou voulant quantifier le nombre de sujets à inclure dans un essai clinique pour détecter un effet de manière significative.

Durée : 1 journée

Pré-requis :
STAT101

Logiciels possibles : SAS

- 1. Rôle statistique du nombre de sujets dans les études cliniques**
- 2. Méthodes d'estimation de la taille d'étude pour les tests simples**
- 3. Application à des cas concrets, présentation de la proc POWER**
- 4. Estimation dans des situations plus complexes : régression logistique, modèle de survie**
- 5. Introduction du calcul de taille d'étude et de puissance *a posteriori* avec le modèle linéaire mixte**



[TYPOMKT] Typologie pour la segmentation client

Ce stage est destiné aux chargés d'étude marketing qui ont besoin de bâtir des groupes homogènes de clients. Les techniques abordées sont la classification hiérarchique, les K-moyennes et la segmentation par arbre.

Durée : 2 jours

Pré-requis :
STAT101

Logiciels possibles : SAS, R

1. Analyse factorielle

- Analyse exploratoire
- Création de facteurs synthétiques
- Influence du codage sur le résultat
- Sélection du nombre optimal de facteurs

2. Segmentation

- Classification Ascendante Hiérarchique, dendrogramme
- K-moyennes, CCC, formes fortes
- Méthode mixte / de Wong
- Classification explicite par arbres

3. Description des classes, reclassement

- Description univariée (caractérisation)
- Réaffectation aux classes par arbre
- Réaffectation aux classes par régression logistique
- Réaffectation géométrique (distance aux centres)