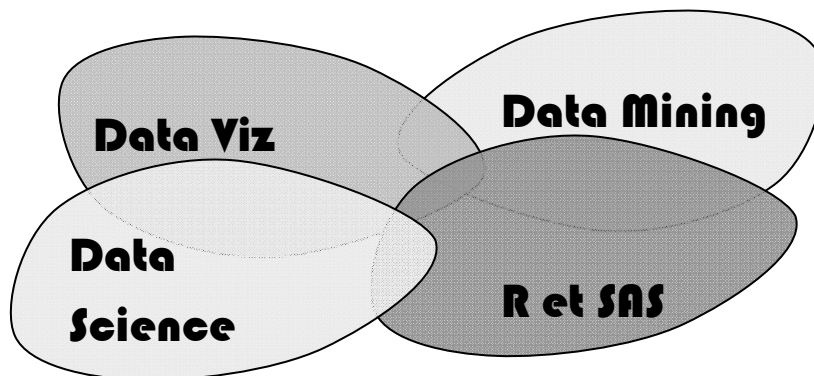




# Olivier Decourt SARL



# Catalogue formations 2018

Nous contacter par email :  
formation@od-datamining.com

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





## Sommaire

Tarifs et conditions.....	4
R : maîtriser un logiciel souple, polyvalent et gratuit.....	5
[R_BASE] Initiation à R.....	6
[SAS->R] Migrer de SAS vers R.....	7
[R_PLUS] Aller plus loin avec R .....	8
[R_REPORT] Construire des rapports avec R.....	9
[R_SHINY] Construire des applications interactives avec le package shiny .....	10
Formations au DataViz : présentation graphique des données.....	11
[VISUAL] Présenter clairement des données, construire des graphiques intelligents	12
[SAS_GRAPH1] Procédure SGPLOT .....	13
[SAS_GRAPH2] Modèles graphiques et GTL.....	14
[SAS_CARTO] Cartographie avec SAS.....	15
[R_GGLOT] Produire des graphiques avec le package ggplot2 .....	16
Formations statistiques : Data Science, Data Mining .....	17
[STAT101] Initiation à la statistique.....	18
[ANADON] Analyse des données .....	19
[GENMOD] Modèle linéaire généralisé.....	20
[MIXED] Modèles mixtes et hiérarchiques .....	21
[R_SCORING] Machine learning avec R.....	22
[REGQUALI] Régression logistique.....	23
[SCORING] Panorama et comparaison des méthodes de scoring .....	24
Formations métier : biostatistique, marketing, actuariat.....	25
[BAYES] Introduction à la statistique bayésienne.....	26
[GLMTARIF] Modélisation pour la tarification en assurance .....	27
[POWER] Calcul de puissance et de nombre de sujets nécessaires .....	28
[TYPOMKT] Typologie pour la segmentation client .....	29



## 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

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



## [SAS->R] 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**

**Nouvelle  
formation**

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

**Logiciels possibles : R**

### 1. Quelques points de repères 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

### 2. Remplacer les procédures usuelles

- Proc Means : fonctions summary, aggregate, describe
- Proc Freq : fonctions table, prop.table, addmargins
- Proc Tabulate : fonction tabular
- Proc Transpose : package reshape2
- Proc SQL : package dplyr
- Proc Sort : fonctions order, duplicated, duplicated2
- Proc Sgplot et Sgpanel : package ggplot2
- Proc Export : fonction write.table, package openxlsx

### 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



## [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\_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 openslsx et xltabr)

- Créer un classeur avec openslsx
- 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 : 2 jours**

**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



# **Formations au DataViz : présentation graphique des données**



# [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 (*dashboards*) 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



## [SAS\_GRAPH1] Procédure SGPLOT

Ce stage est destiné aux chargés d'études souhaitant découvrir la nouvelle procédure SAS produisant des graphiques à partir de la version 9.2 : 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  
(formation  
Educasoft)**

**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



# [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**

**Pré-requis :  
R\_PLUS,  
VISUAL**

**Logiciels possibles : R**

## 1. Le package ggplot2

- Principes
- Exporter un graphique, sauvegarde et recyclage

## 2. La fonction qplot

- Syntaxe de base
- Diagramme en bâtons
- Histogramme
- Boxplot
- Nuage de points
- Courbe

## 3. La fonction ggplot et ses compléments

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

## 4. Eclatement

- Principe
- Eclatement





# **Formations statistiques : Data Science, Data Mining**



## [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, SAS Enterprise Guide**

**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 de données à la française (ACM, AFC, ACP) et surtout leur utilisation à travers SAS, R (package FactoMineR) ou Spad. On y aborde également la classification.

**Durée : 2 jours**

**Pré-requis :  
STAT101**

**Logiciels possibles : SAS, R ou SPAD**

### **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

L'étude des données avec une analyse de la variance se conduit d'ordinaire sur des facteurs considérés comme fixes : c'est à dire qu'on se limite dans l'analyse et l'inférence aux valeurs qui ont été collectées au cours de la constitution des données.

Des facteurs aléatoires et un modèle mixte étendent de manière très importante la puissance des modèles d'analyse de variance, et facilitent également 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



# [R\_SCORING] 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, comparaison de modèles.**

**Durée : 3 jours**

**Pré-requis :  
R\_PLUS,  
REGQUALI**

**Logiciels possibles : R**

## **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**

- L'analyse de la déviance, étude de l'impact d'une covariable
- 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] Panorama et comparaison des méthodes de scoring

Cette formation s'adresse aux chargés d'étude désirant avoir, en quelques jours, un aperçu technique et pratique des techniques usuelles de scoring. La formation s'achève avec une comparaison des forces et des faiblesses des différentes méthodes.

**Durée : 3 jours**

**Pré-requis :  
aucun**

**Logiciels possibles : SAS, R**

## 1. Scoring avec les arbres de décision

- Principe général d'un arbre de décision
- Croissance et élagage
- Les principaux algorithmes : CHAID, CART, C4.5
- Arbres, bagging et boosting : comment rendre un arbre robuste
- Avantages et inconvénients

## 2. Scoring avec la régression logistique

- Modèle linéaire et modèle logistique
- Choix des variables, automatisation
- Coefficients et odds-ratios
- Courbe ROC, discrimination
- Avantages et inconvénients

## 3. Scoring avec l'analyse discriminante

- Approche géométrique
- Fonction linéaire discriminante
- Méthode DISQUAL : l'analyse discriminante sur données qualitatives
- Avantages et inconvénients

## 4. Scoring avec les réseaux de neurones

- Le neurone artificiel
- Apprentissage supervisé
- Lecture et interprétation des résultats
- Avantages et inconvénients

## 5. Autres méthodes de scoring

- Raisonnement basé sur la mémoire
- Machines à vecteurs supports (méthode Vapnik)
- Bagging et boosting
- Comparaison générale des méthodes de scoring





# **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**

- 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. Des applications sur les modèles composites (IARD) et de prime pure (santé) sont proposées sous SAS ou R.

**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)