

MODÈLES PRÉDICTIFS, RÉGRESSIONS LINÉAIRES, LOGISTIQUES, PLS ET ANCOVA

Cette formation s'adresse à des personnes souhaitant mettre en œuvre des méthodes de modélisation.

Différents types de modélisation seront traités :

- Régression linéaire multiple
- Modèle linéaire général type ANCOVA
- Régression logistique (réponse binaire ou multimodale)
- Régression PLS

L'objectif de cette formation est de transmettre aux participants un savoir-faire méthodologique sur ces analyses :

- Contexte et objectifs
- Conditions d'utilisation
- Mesure de la qualité des modèles
- Mise en œuvre et interprétation des résultats, ...

Durée : 28 heures (4 jours)

Lieu : Formation disponible en présentiel dans vos locaux, ou en ligne en visioconférence

Tarifs :

- Formation en ligne inter-entreprise : n/a
- Formation en ligne intra-entreprise : 6 960,00€ HT / formation
- Formation sur site intra-entreprise : 6 960,00€ HT / formation + frais de déplacement du formateur

Profils des stagiaires :

- Technicien – Employé – Cadre – Chercheur – Etudiant
- Toute personne souhaitant mettre en œuvre et interpréter les résultats d'une méthode d'analyse statistique prédictive de type régression ou de type modèle linéaire général (PLS, ANCOVA...).

Prérequis :

- De bonnes connaissances sur les outils statistiques de base : statistiques descriptives, intervalles de confiance, p-value, risque alpha, tests d'hypothèses...
- Quelques connaissances sur la corrélation et la régression linéaire

Objectifs pédagogiques

- Définir le contexte et les objectifs des différentes méthodes
- Mettre en œuvre et interpréter les résultats des méthodes suivantes :
 - Régression linéaire multiple
 - Modèle linéaire général (ANCOVA, MANOVA...)
 - Régression PLS

- Régression logistique
- Vérifier les conditions de mise en œuvre des différentes méthodes
- Connaître les grandes lignes des concepts mathématiques inhérents à ces méthodes
- Identifier et prendre en compte les problèmes de multi colinéarités sur les modèles linéaires : sélection des variables, PLS...
- Mesurer la qualité d'ajustement et la qualité de prédiction (d'estimation des coefficients) du modèle
- Détecter et analyser les points aberrants / les points influents

Contenu de la formation

- **Rappels sur les notions de corrélation**
 - Définition du coefficient de corrélation
 - Interprétation de la valeur du coefficient de corrélation
 - Les confusions : corrélation, causalité, pente...
 - Les différents coefficients de corrélation : Coefficient de Pearson - Coefficient de Spearman
- **Modélisation de type régression linéaire simple**
 - Principes et concepts mathématiques inhérents à la régression linéaire simple
 - Tests d'hypothèse de significativité du modèle
 - Qualité du modèle
 - Coefficient de détermination R^2 , R^2 ajusté, R^2 Prev
 - Utilisation du modèle : Prédiction de valeurs individuelles, Intervalles de confiance des prédictions
 - Traitement graphique des résultats
 - Principes et concepts mathématiques inhérents à la régression linéaire multiple
 - Inférence du modèle, inférence des variables (statistique de Fisher)
 - Analyse des résidus :
 - Calculs des résidus
 - Sens physique et statistique
 - Homogénéité
 - Distribution, Normalité
 - Valeurs suspectes
 - Analyses graphiques
- **Valeurs suspectes et points influents**
 - Résidus :
 - Résidus studentisés
 - Effet levier
 - Distance de Cook
 - Qualité du modèle :
 - Qualité d'ajustement, coefficient de détermination R^2 , R^2 ajusté
 - Qualité de prédiction, erreur d'estimation
 - Utilisation du modèle :

- Prédiction (prévision) de valeurs individuelles
 - Intervalles de confiance des prédictions (prévisions)
 - Illustration graphique des résultats
- **Le modèle de régression multiple**
 - Significativité des coefficients
 - Hiérarchie des coefficients
 - Problèmes liés à la multi-colinéarité
 - Mesures des colinéarités : Coefficient de corrélation - Facteur d'Inflation de la Variance (VIF)
 - Résolution des problèmes de multicollinéarité
- **Analyse des problèmes de multicollinéarité à travers la sélection des variables**
 - Détection de la colinéarité :
 - Effets nocifs de la colinéarité entre variables explicatives
 - Outils de détection : corrélation, VIF, cohérence des signes
 - Les solutions proposées :
 - Expérimentation structurée
 - Sélection de variables
 - PLS
 - Traitement de la colinéarité – Sélection de variables :
 - Sélection par optimisation. Critère du R^2 , du R^2 ajusté, de l'AIC et du BIC
 - Algorithmes de sélection « pas à pas » : Forward selection, Backward selection, Stepwise regression
- **Mise en œuvre et interprétation des résultats d'une régression PLS**
 - Contexte et objectifs
 - Présentation des différentes méthodes de régression sur des données colinéaires : PCR, régression de Ridge et PLS
 - Principes et concepts mathématiques inhérents à la PCR et la régression PLS
 - Présentation des différentes versions de la régression PLS
 - Mise en œuvre et interprétation des résultats : graphiques, coefficients du modèle...
 - Choix du nombre de composantes (validation croisée)
 - Composantes et coefficients de régression
 - Qualité d'ajustement, qualité de prédiction
 - Coefficients Q^2 et R^2
 - Importance des variables explicatives pour la prédiction : Coefficients normalisés - VIP
 - Sélection de variables
- **Mise en œuvre et interprétation des résultats d'une ANCOVA (modèles linéaires généraux)**
 - Contexte et objectifs
 - Notion d'interaction entre les variables explicatives qualitatives et quantitatives
 - Modèle à droites confondues
 - Modèle additif
 - Modèle complet
 - Mise en œuvre et interprétation des résultats des différents modèles
 - Lecture et utilisation du modèle

- Tests de significativité des différents termes (F de Fisher)
- Épuration du modèle (sélection des termes et des variables influents)
- Conditions d'utilisation de l'ANCOVA
- **Modélisation de type régression logistique**
 - Contexte et objectifs
 - Différences entre la régression linéaire et la régression logistique
 - Définition du modèle Logit
 - Mise en œuvre et interprétation des résultats
 - Découpage en classes des variables explicatives quantitatives
 - Estimation et interprétation des coefficients du modèle
 - Tests d'apport d'une variable (test de Wald, tests sur les rapports de vraisemblance)
 - Interprétation du Khi^2 de Wald
 - Odds-ratios
 - Parallèle odds ratios et risques relatifs
 - Analyse du tableau de classement :
 - Taux de réussite, taux d'échec - Vrais positifs, vrais négatifs, faux positifs, faux négatifs
 - Probabilités ajustées et utilisation du modèle à des fins de prédiction
 - Conditions d'utilisation

Organisation de la formation

Equipe pédagogique Arkesys :

- Responsable de la filière statistique du groupe Arkesys et formateur en statistique, **Thierry Anthouard** est passionné depuis toujours par le domaine de la statistique. C'est en 1992 qu'il a lancé le développement de la filière formation statistique du Groupe Arkesys. Son approche pédagogique "par l'exemple" lui permet de vulgariser la statistique et de la rendre accessible à tous les apprenants. Consultant dédié à l'accompagnement de clients grands comptes, il s'adapte à tout type de contexte et de problématique d'apprentissage.
- Formateur en statistique, **Jérôme-Philippe Garsi** bénéficie d'une expérience de 13 ans dans le domaine de la formation. Depuis son doctorat sur des questions cliniques, ses travaux visent principalement l'intérêt des populations, leur santé et leur bien-être. A l'aise quel que soit le public, il fait de la pédagogie et de la vulgarisation des connaissances scientifiques un impératif. Pour cela, il fait toujours preuve du plus grand soin de clarté aussi bien dans ses documents écrits que dans ses présentations orales.

Moyens pédagogiques et techniques :

- Accueil des apprenants dans une salle dédiée à la formation
- Explications théoriques suivies de pratiques guidées puis mises en autonomie
- Mise à disposition de documents supports en ligne suite à la formation
- Les exercices d'accompagnement peuvent être récupérés sur clef USB

Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation :

- Feuilles de présence

Addinsoft

40 rue Damrémont
75018 Paris
Email: training@xlstat.com
Tel: 0170060488

- Exercices de synthèse et d'évaluation
- Evaluation de fin de stage
- Certificat de réalisation de l'action de formation

Accessibilité aux personnes handicapées :

Les personnes en situation de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

Délais d'accessibilité à la formation avant le démarrage : 2 semaines

Formation en ligne : Le lien de connexion à la classe virtuelle sera envoyé par email la semaine qui précède la formation.

Contact : Pour toute demande d'informations complémentaires vous pouvez nous contacter par email training@xlstat.com ou par téléphone au 0170060488.