

Statistique

MODELES PREDICTIFS, REGRESSIONS LINEAIRES, LOGISTIQUES, PLS ET ANCOVA

Référence : 2020-STAT-N2-MO

4 jours (28h) – Présentiel ou À distance

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

Public concerné : 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...).

La formation en quelques mots

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

Thèmes principaux

- Contexte et objectifs des différentes méthodes
- Qualité d'ajustement et qualité d'estimation des coefficients du modèle (qualité de prédictions)
- Régression linéaire
- Régression pas à pas
- Modèles linéaires généraux (ANCOVA...)
- Régression logistique
- Régression PLS
- Problématique des multi colinéarités
- Conditions d'utilisation des différentes méthodes

Jeux de données

Afin de s'approcher au mieux des réalités quotidiennes des praticiens, nous suggérons de nous appuyer pour l'animation pratique de thématiques et surtout de jeux de données reflétant le quotidien des apprenants.

Cet élément est un facteur de réussite pour la formation. Elle permet aux apprenants de :

- Se "reconnaitre" dans les thèmes abordés,
- Mieux percevoir l'intérêt des notions étudiées
- S'approprier le contenu de la formation

Il sera donc pertinent que les apprenants puissent réfléchir en amont de la formation à des problématiques, jeux de données ou documents susceptibles d'être utilisés en support lors de la formation.

Outil logiciel

Cette formation n'est pas dédiée à la pratique d'un logiciel particulier. Les exercices et les illustrations se feront autour d'un des logiciels de statistique suivants : XlStat, R, Jmp, MiniTab, StatGraphics, ...

Le logiciel retenu sera choisi en accord avec celui utilisé par le plus grand nombre des stagiaires.

Prérequis

Il est indispensable que les participants aient :

- 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

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

- 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

Programme

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
 - Vifs
- 🌀 Résolution des problèmes de multi-colinéarité

Analyse des problèmes de multi colinéarité 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)
- 🌀 Epuration 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

Méthodes et moyens

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

Méthodes d'évaluation des acquis

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

Profil formateur

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

Support stagiaire

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

Accessible à tous

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