STA212 - Données catégorielles

Mis à jour le 17-02-2025

Code : STA212

Unité d'enseignement de type cours

9 crédits

Volume horaire de référence (+/- 10%) : **70 heures**

Responsabilité nationale : EPN06 - Mathématique et statistique / Vincent AUDIGIER

Contact national :

EPN06 Mathématiques et statistiques

2 rue conté
Accès 35 3 ème étage porte 19
75003 Paris

Anne - Solenne Maroulle
anne-solenne.marroulle@lecnam.net

Présentation

Prérequis

Ce cours s'adresse en priorité aux auditeurs préparant le master MR123.
Il s'adresse également à des auditeurs préparant le diplôme d'actuaire.
Niveau requis : STA101 (méthodes exploratoires multidimensionnelles), STA102 (régression et analyse de la variance) et STA104 (statistique mathématique).

Objectifs pédagogiques

Permettre aux ingénieurs, cadres d'entreprises ou d'administrations, médecins, chercheurs, de construire des modèles explicatifs de variables qualitatives. Le cours s'appuie sur la pratique du logiciel R.

Programme

Contenu

Présentation des méthodes statistiques traitant des variables qualitatives
Principes généraux d'estimation d'un modèle (maximum de vraisemblance)

La régression logistique simple
Notion de variable latente
Les modèles PROBIT, LOGIT
Le modèle logistique et son interprétation
Analyse des résidus, des observations

La régression logistique multiple
Le modèle : prédicteurs quantitatifs ou qualitatifs
Sélection de variables
Résumé des tests de validité générale d'un modèle
Tables de classement, courbe ROC
Interprétation des coefficients de la régression logistique : odds ratio
La régression logistique dans le cas où Y est une variable polytomique ordonnée
Aspects pratiques de la mise en oeuvre des méthodes de régression logistique

Modélisation d'une réponse multinomiale
Estimation par maximum de vraisemblance ou par moindres carrés généralisés

Comparaison de la régression logistique avec d'autres méthodes de modélisation d'une réponse qualitative
Analyse discriminante sur variables quantitatives et qualitatives
Arbres de décision
Eléments pratiques de création d'un score

Méthodes PLS
Présentation de NIPALS, PLS1,PLS2
Applications: régression logistique PLS, analyse discriminante PLS

Modalités de validation

- Projet(s)

Description des modalités de validation

Rédaction d'un mémoire utilisant les méthodes présentées en cours

Bibliographie

Titre	Auteur(s)
Applied logistic regression 2ième édition (Wiley, 2000)	HOSMER D., LEMESHOW S.
Statistique explicative appliquée (Technip, 2003)	NAKACHE J.P., CONFAIS J.
Modèles statistiques pour données qualitatives (Technip 2005)	Editeurs DROESBEKE, LEJEUNE, SAPORTA; Auteurs: CROUX, GONZALEZ....
Categorical Data Analysis using the SAS System (SAS Institute Inc, Cary, NC, 2000)	STOKES M.E., DAVIS C.S. & KOCH G.G.
Analyse discriminante (Dunod, 2001)	BARDOS M.
Categorical data analysis 2ième édition (Wiley, 2002)	AGRESTI A.