MODULE 10

MODÈLE LINÉAIRE GÉNÉRALISÉ

Présentation

"Une famille de lois de probabilité sur R^n définit une structure exponentielle courbe si elle peut être plongée de manière lisse dans une structure canonique saturée et régulière." (Régression non linéaire et applications, A. Antoniadis, J. Berruyer, R Carmona, Ed. Economica).

Dispositif Formation Statistique FPSTAT2

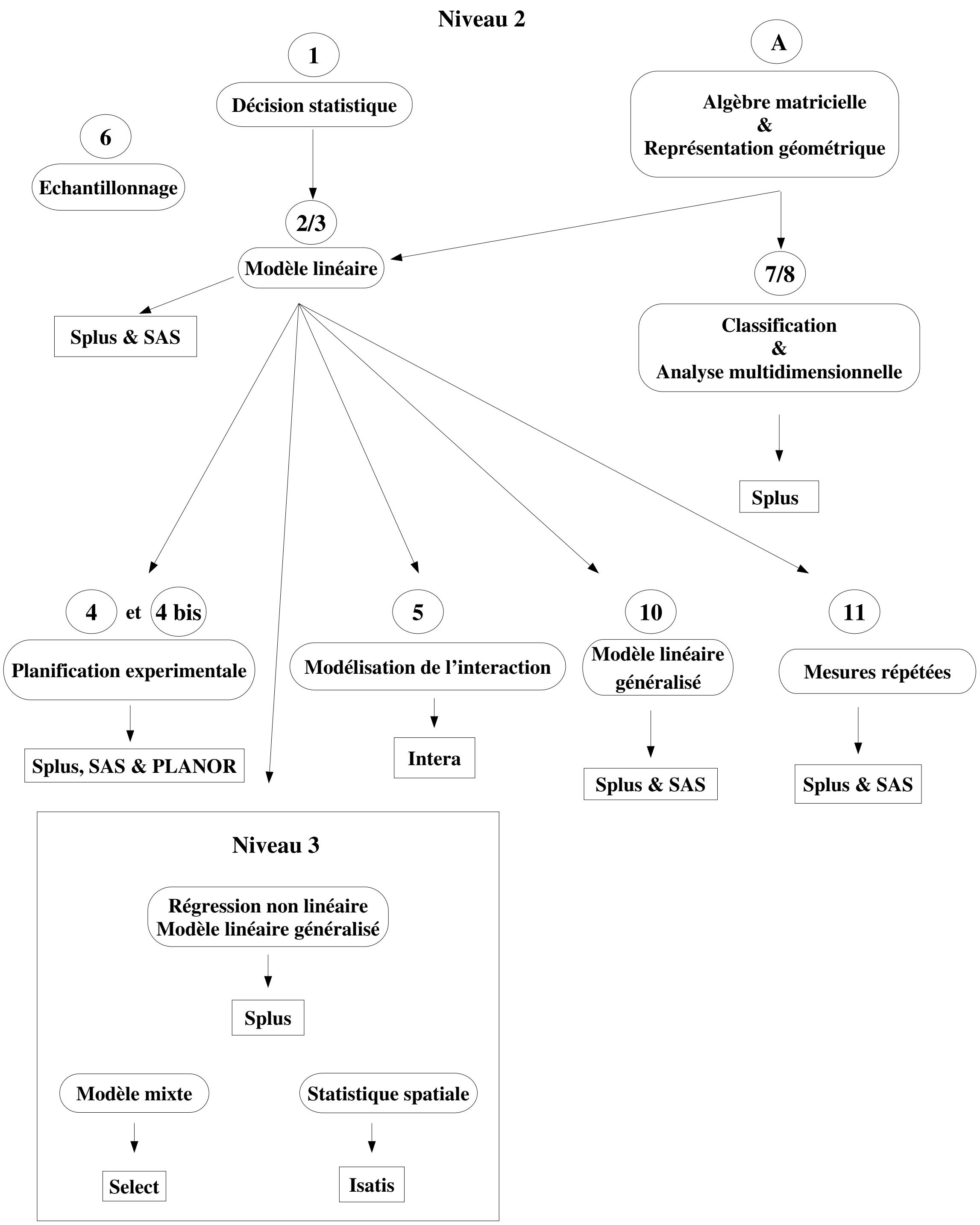
Pour répondre aux besoins de formation aux statistiques de l'INRA, le département de Biométrie et Intelligence Artificielle et le service de Formation Permanente ont construit un dispositif national de formation interne. Trois niveaux sont distingués en fonction de l'activité professionnelle des agents à former :

- le niveau 1 concerne les agents qui mettent en place les expérimentations, collectent et présentent les données,
- le niveau 2 concerne les agents qui élaborent les protocoles, traitent et analysent les résultats,
- le niveau 3 s'adresse aux agents de niveau 2 qui ont besoin de méthodes statistiques non standard.

La formation de niveau 2 est organisée en modules indépendants mais néanmoins positionnés les uns par rapport aux autres selon une hiérarchie permettant d'acquérir dans un module les connaissances prérequises des modules sur lesquels il pointe. Cette hiérarchie est décrite dans le schéma ci-joint. Le module "Modèle Linéaire généralisé" s'articule naturellement à la suite du module "Modèle Linéaire".

L'**objectif** principal du module "Modèle Linéaire généralisé" est de donner au stagiaire une idée claire des raisons qui peuvent pousser à utiliser le modèle linéaire généralisé, et des différentes composantes à choisir pour la construction d'un modèle. L'objectif est aussi de faire apparaître les difficultés éventuelles, les pièges à éviter, et de montrer une ou des stratégies d'analyse. Un des points importants de la formation est de fournir des outils graphiques de validation de modèle. La mise en œuvre avec les logiciels Splus ou SAS sera montrée "sur le papier", de même que l'interprétation des résultats.
Une pédagogie de **présentation par l'exemple** est adoptée. Dans la mesure du possible, toutes les notions présentées sont concrétisées par des exemples. Le formalisme mathématique et les notations utilisés sont simples, leur signification sous-jacente est toujours explicitée. Les démonstrations sont exclues de cette formation.

FPstat2


Niveau 2
1
Décision statistique
A
Algèbre matricielle
&
Représentation géométrique
6
Echantillonnage
2/3
Modèle linéaire
Splus & SAS
7/8
Classification
&
Analyse multidimensionnelle
Splus
4 et 4 bis
Planification experimentale
Splus, SAS & PLANOR
5
Modélisation de l'interaction
Intera
10
Modèle linéaire généralisé
Splus & SAS
11
Mesures répétées
Splus & SAS
Niveau 3
Régression non linéaire
Modèle linéaire généralisé
Splus
Modèle mixte
Select
Statistique spatiale
Isatis

Le **matériel pédagogique** est constitué de *transparents commentés*, pour chacun des chapitres, auxquels s'ajoutent un *glossaire anglais/français* des principaux termes utilisés, une *bibliographie commentée*, une fiche de vérification des *prérequis*, un *complément formateur sur l'algorithme IRLS* et le document *"Modèle linéaire généralisé, mise en œuvre des exemples de Collett avec SAS et Splus"*. Ce document est fourni dans la valise. Il est trop volumineux pour être dupliqué pour chaque stagiaire, mais il est mis à disposition des stagiaires, sur le serveur du département de Biométrie http://www-bia.inra.fr.

Les *fichiers* script.chap2.SAS, script.chap2.Splus, script.chap4.Splus, script.chap7.Splus contiennent les *enchaînements de commande Splus ou SAS* permettant de mettre en œuvre les exemples présentés dans ces chapitres.

Des exemples issus des recherches menées à l'INRA ont été utilisés pour illustrer le cours. Des exercices ont été intégrés, correspondants aux points clefs du module. Ils ne nécessitent ni calcul ni démonstration, mais un peu de réflexion, et permettent de vérifier la bonne compréhension de quelques points importants.

Un exemple est d'abord présenté, il montre la manière dont on peut utiliser la méthodologie du modèle linéaire généralisé, et justifie la nécessité d'une telle méthodologie. Des éléments de théorie sont ensuite présentés, dans le cadre de données binomiales. Des outils de validation de modèle sont proposés, ainsi qu'une stratégie de choix de modèle. Le modèle multinomial est abordé ainsi que le cas d'autres types de lois.

Le **plan** du module est le suivant :

I INTRODUCTION
Du modèle linéaire au modèle linéaire généralisé

II LE MODÈLE BINOMIAL
Présentation par l'exemple

III ÉLÉMENTS DE THÉORIE, MAXIMUM DE VRAISEMBLANCE
1. Loi binomiale
2. Cas général

IV OUTILS DE VALIDATION D'UN MODÈLE

1. Validation du lien, étude des résidus
2. Amélioration du modèle :
   - modification du prédicteur linéaire
   - changement de fonction de lien

V CHOIX DE MODÈLE

1. Non orthogonalité
2. Instabilité des paramètres

VI MODÉLISATION, ESTIMATION PONCTUELLE ET PAR INTERVALLE

1. Une adaptation de la fonction de lien
2. Surdispersion
3. Intervalle de confiance pour une fonction des paramètres

VII LE MODÈLE MULTINOMIAL

1. Catégories non ordonnées
2. Catégories ordonnées

VIII AUTRES LOIS, AUTRES MODÈLES

1. Loi de poisson
2. Loi binomiale négative
3. Loi gamma

Cette formation peut raisonnablement être dispensée sur trois jours, avec le découpage approximatif suivant : chapitres I à III le premier jour, IV à VI le deuxième, et VII et VIII le troisième.

Nous remercions ici Jacques Badia, Jean-Louis Foulley et Sylvie Huet, animateurs de la formation de niveau 3 intitulée "Formation du modèle Linéaire Généralisé" sans qui ce module n'aurait pas été réalisé, Sylvie Huet et Christelle Hennequet pour les compléments qu'elles nous ont apportés au cours de la préparation du module, Denis Tourvieille de Labrouhe, Christian Salles, Sylvie Bernard et André Madéore pour les données qu'ils nous ont fournies, ainsi que D. Collett d'avoir écrit un ouvrage qui nous a été très utile (Modelling Binary Data, D. Collett, Chapman & Hall, 1991). Nous tenons à remercier tout particulièrement Sylvie Huet qui a relu et critiqué les transparents commentés.