

DIDACTICIEL DES TECHNIQUES DE LA STATISTIQUE

- Module Ajustement linéaire -

Michel JANVIER
E.R.E.S.
Université Montpellier II
Sciences et Techniques du Languedoc
Place Eugène Bataillon
34095 MONTPELLIER CEDEX 5
France

Eric VERDOIRE
Conservatoire National des Arts et Métiers
(C.N.A.M.)
Centre Régional du Languedoc-Roussillon
11, avenue Professeur Grasset
34000 MONTPELLIER
France

Présentation

Ce didacticiel fait partie d'un ensemble qui couvrira tout le programme de la demi-valeur du premier cycle du CNAM.

Le public visé comprend :

- * en formation professionnelle, tous les auditeurs CNAM,
- * en formation permanente, toutes les personnes utilisant les statistiques dans leur activité professionnelle,
- * en formation initiale, de nombreuses catégories d'étudiants dont le cursus comporte une initiation à la statistique : sciences économiques, médecine, pharmacie, biologie, technologie, gestion, écoles de commerce, lycées, classes de S.T.S. .

Contextes d'utilisation

Le didacticiel est susceptible d'être utilisé :

- * en remplacement de l'enseignement traditionnel (utilisation à distance),
- * comme soutien ou approfondissement des connaissances (utilisation en auto-formation à domicile ou en centre ressource),
- * pour illustrer un enseignement présentiel en exploitant les possibilités de simulation ou de traitements des données disponibles dans le didacticiel ainsi que les nombreux exemples et fonctions de données inclus dans le didacticiel (utilisation en salle de cours).

Stratégie didactique

La stratégie didactique choisie tend à faciliter l'appropriation des concepts par la présentation de nombreuses situations-problèmes. L'élaboration des situations s'appuie sur une classification préalable des relations de base intervenant dans le champ conceptuel considéré et l'analyse des classes de problèmes que l'on peut générer à partir de ces relations. La résolution des situations-problèmes

présentées permet d'introduire l'utilisation de procédures de traitement. Pour faciliter l'assimilation des concepts, le cours ne se limite pas à une présentation traditionnelle dans le symbolisme mathématique. Il est imagé par de nombreuses représentations graphiques et simulations. Des questions sont posées en langue naturelle et les réponses de l'élève, attendues sous cette forme, sont analysées par le système.

Un grapheur intégré permet la construction et l'analyse de graphiques. La progression dans le cours impose la lecture et l'utilisation de formules mathématiques. L'élève doit aussi se familiariser avec la pratique du calcul numérique à l'aide d'une calculette intégrée. Ainsi, les concepts sont présentés sous des formes symboliques aussi variées que possible.

Pour prendre en compte la diversité des voies d'accès à un problème, ainsi que les diversités cognitives des apprenants, des changements de registre ont été utilisés pour faire évoluer les conceptions. Par exemple, passage du registre géométrique à celui de l'ajustement d'un modèle à partir d'un échantillon, dans le module sur l'ajustement linéaire.

Contenu

Le module "ajustement linéaire" a pour but de présenter deux aspects fondamentaux d'une analyse de données :

- * il introduit la notion de modèle, qui conduit naturellement à la partie inférentielle, complémentaire de la partie exploratoire d'une analyse statistique,
- * par l'analyse des résidus on aborde la problématique du diagnostic et du contrôle de la qualité du modèle.

Le modèle de la régression simple met en relation, par l'intermédiaire d'une fonction affine, la variable expliquée Y et la variable explicative X .

$$Y = aX + b$$

Mais à cette fonction affine qui permet de modéliser le comportement de Y en fonction de X s'ajoute un terme d'erreur traduisant notamment le fait que la fonction retenue ne permet pas de prendre en compte tous les éléments explicatifs de Y . Ainsi, pour un ensemble de n observations relatives aux variables Y et X :

$$y_i = ax_i + b + r_i \quad i \in [1, n] \text{ et } i \in \mathbb{N}$$

La méthode des moindres carrés consiste à minimiser la somme des carrés des résidus r_i :

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - ax_i + b_i)^2$$

Elle permet de déterminer les coefficients a et b , à partir des n observations.

Une décomposition de la variance totale, en variance expliquée par le modèle et variance résiduelle, permet d'introduire le coefficient de détermination R^2 . La qualité d'un ajustement peut être, en partie jugée par la valeur de R^2 , mais il est aussi nécessaire d'étudier le graphique des résidus. Ce graphique est obtenu facilement grâce au grapheur intégré au didacticiel.

Lorsque les relations entre les variables ne sont pas affines, différentes transformations permettent de se ramener au cas affine, et on peut juger de la qualité du nouveau modèle, au moyen des instruments introduits.